2022

Habitat loss and fragmentation: effects on genetic diversity and differentiation

Gabriele Maria Sgarlata



Dissertation presented to obtain the **Ph.D degree in**

Integrative Biology and Biomedicine

Oeiras, October, 2022

INSTITUTO DE TECNOLOGIA QUÍMICA E BIOLÓGICA ANTÓNIO XAVIER | UNIVERSIDADE NOVA DE LISBOA

Habitat loss and fragmentation: effects on genetic diversity and differentiation

Gabriele Maria Sgarlata

Dissertation presented to obtain the Ph.D degree in Integrative Biology and Biomedicine

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Research work coordinated by: 

Oeiras, October, 2022

Photo Cover: On the way to reach the forest of Andantony, close to the city of Andapa, in north-eastern Madagascar. I thought that this picture would well represent, symbolically, the journey of my PhD. The picture was taken by Costanza Piccoli.



The work described in this thesis received financial support from Fundação para a Ciência e a Tecnologia, through the grant PD/BD/114343/2016 awarded to Gabriele Maria Sgarlata.

This work was conducted at the Instituto Gulbenkian de Ciência (IGC), Oeiras, Portugal, under the supervision of Dr. Lounès Chikhi.

Dedicated to my travelling companions, my Malagasy friends, my community, my family, and her…

Preface

In the present PhD project, I investigated the consequences of habitat loss and fragmentation (HL&F) on species genetic diversity and differentiation. On one hand, I have used **theoretical approaches** by performing extensive spatially-explicit simulations and proposing numerical models that can capture the spatio-temporal dynamic of genetic variability in the context of HL&F. On the other hand, I have used an **empirical approach** by collecting and analysing genetic data from rodent (genus *Eliurus*) and primate species (genus *Microcebus*) endemic to Madagascar. In both approaches, I have explored HL&F at various spatial resolutions: *regional-scale*, *landscape-scale* and *habitat patch-scale*.

The samples used for genetic analysis were collected throughout several years of fieldwork (2010 – 2018), to which I have contributed for a total amount of ~11 months (3 fieldwork seasons). The fieldwork consisted in surveying remote and little explored forest fragments across Northern Madagascar. During the fieldwork, I have been collecting tissue and faecal samples from endemic rodents, chameleons, geckos and lemurs. I have characterised the biodiversity of the surveyed forests throughout opportunistic observations and estimated lemur population density via line-transect distance-sampling, when sufficient number of observations were obtained. The fieldwork was a collaborative effort which involved several colleagues (assistant researchers, students and researchers) from Madagascar and elsewhere. It led me to supervise four Master students from Madagascar, inform local stakeholders and members of the State involved in the conservation of Malagasy forests on the encountered biodiversity, and contribute to update the conservation status of 13 lemur species during the last IUCN Red List workshop in 2018 (Antananarivo, Madagascar). Several of the scientific and conservation-related outcomes I obtained throughout my PhD project will not be present in this thesis. However, I wanted the reader to

know that my PhD experience goes way beyond the content of the Chapters I am going to introduce in this preface. I feel to have grown both as young scientist and as person, enriched by experiences and new cultures. And for all this, first of all, I have to thank my PhD supervisor: Dr. Lounès Chikhi.

This PhD thesis has been organised in the following way:

Chapter 1: *General Introduction on the consequences of HL&F on biodiversity*

Chapter 1 provides a 'General Introduction' on the ubiquity of HL&F across temporal and spatial scales, and on the effect of HL&F on biodiversity (ecological and genetic). I present some of the controversies around the effect of habitat fragmentation on biodiversity. I briefly introduce the main models in spatial population genetics and reviews some of their results. I then continue with an overview of the natural history of Madagascar and the ongoing debate on the origin of HL&F in this island. I conclude reviewing some of the genetic studies carried out in Madagascar, investigating the effect of *recent* versus *ancient* HL&F on species genetic diversity. Part of the 'General Introduction', in particular the section dedicated to Madagascar, has been included in a book chapter entitled '*Genetics of Malagasy Mammals: An Overview*' that my supervisor (Dr. Lounès Chikhi) and I have redacted for the new edition of 'The Natural History of Madagascar'. It is currently *in press* and will soon be available.

Chapter 2: *On some genetic consequences of habitat contraction: multi edge-effect and habitat loss*

Chapter 2 investigates the effect of HL&F on genetic diversity at the scale of a habitat patch. More precisely, we explored the genetic consequences of habitat loss, looking at the interplay between two HL&F

processes (*patch-size* and *edge-effect*). We propose a statistical approach to estimate 'edge-impacted population size' from information on habitat cover and geometry, and use this measure to predict local spatial genetic diversity in both equilibrium and non-equilibrium populations. This work shows how population genetic theory and landscape ecology approaches can be successfully integrated for quantifying the influence of multiple edge-effects and habitat contraction on spatial genetic diversity.

Chapter 3: *Isolation-by-distance and time in the context of habitat loss and fragmentation*

Chapter 3 explore the genetic consequences of HL&F at the scale of a landscape. In particular, we studied how the correlation between genetic and geographic distance ('Isolation-by-distance'; IBD) changes over time when a continuous landscape undergoes HL&F. We used a simulation-based approach to explore the behaviour of IBD under different scenarios of HL&F. More specifically, we ask *i)* what is the spatio-temporal dynamic of IBD in the case of an *instantaneous HL&F*, and how *ii)* range expansion and *iii)* variation in tempo and mode of HL&F influences the persistence of IBD patterns. We conclude by presenting a theoretical framework modelling IBD in a toroidal stepping-stone undergoing HL&F, obtained by modifying the original theorical results on IBD by Slatkin (1991) and Slatkin (1993). Our analytical model opens the possibility for future development of an inferential-approach allowing to interpret F_{st} pattern considering both space and time.

Chapter 4: *Landscape genomics of tufted-tailed rats (Eliurus spp.) in northern Madagascar*

Chapter 4 focuses on the landscape and population genomics of five forest-dwelling rodent species (*Eliurus spp.*) endemic to Madagascar. We obtained genomic data from samples collected in northern Madagascar and

investigated how *Eliurus* species associated to different habitat types (humid *versus* dry habitats) respond to present-day HL&F. In particular, we used classical population genomic approaches to i) uncover landscape-wide patterns in genomic diversity; ii) compare landscape genomic pattern across regions separated by natural barriers (e.g., rivers); and iii) provide some insights on the humid and dry forest cover changes that occurred during the late Holocene.

Chapter 5: *Genetic diversity of mouse lemurs and tufted-tailed rats in northern Madagascar: new species and insights on past forest cover changes*

Chapter 5 includes one published paper and one manuscript investigating regional-scale genetic/genomic diversity pattern across northern Madagascar for two groups of endemic small mammals, respectively: mouse lemurs (*Microcebus spp.*) and tufted-tailed rat (*Eliurus spp.*). For both mammal groups, we identified two new lineages: *Microcebus sp.* and *Eliurus sp. nova*. For the *Eliurus* genus, the new lineage corresponds with high probability to a previously undescribed species. By integrating morphological, climatic, genetic, and spatial data of the several *Eliurus* and *Microcebus* species we show that the spatial distribution of forest- dwelling species can provide valuable insights on the past changes in forest cover and vegetation type that have occurred in northern Madagascar.

Chapter 6: *General Discussion and Future Directions*

Chapter 6 overviews the results of this PhD project and integrate them in a general discussion on the genetic consequences of HL&F. More precisely, it proposes alternative interpretations of spatial genetic pattern (IBD) commonly observed in natural populations, bearing in mind that, although habitats are no longer continuous, IBD can be maintained for very

long time after HL&F. It discusses under which scenarios IBD pattern may disappear, and attempts to contribute to the ongoing controversy around the effect of habitat fragmentation *per se* on biodiversity (i.e., genetic diversity). This Chapter concludes this PhD thesis with few final remarks on future directions and possible improvements for the proposed spatial theoretical frameworks.

Gabriele Maria Sgarlata

Table of Contents

Acknowledgments

Summary

Resumo

Chaper 1: General Introduction

1.1 Current estimate of HL&F in ancient and recent times.....1

1.2 The processes underlying habitat loss and fragmentation 5

1.3 Effects of HL&F on biodiversity.....8

1.3.1 Ecological response to HL&F 8

1.3.2 Genetic response to HL&F 14

1.4 Spatial population genetics.....16

1.5 Theoretical studies on genetic response to HL&F 24

1.5.1 Patch-scale process 24

1.5.2 Landscape-scale process.....32

1.6 Madagascar: the impact of ancient and recent HL&F on biodiversity36

1.6.1 Madagascar Natural History 36

1.6.2 An overview of genetic studies of Malagasy mammals and HL&F 41

1.6.2.1 Influence of HL&F on species abundance and dispersal.....41

1.6.2.2 Reconstruct past-forest cover changes in Madagascar.....44

1.7 Conclusive Remarks 48

1.8 References.....50

Chaper 2: On the genetic consequences of habitat contraction edge effects and habitat loss

2.1 Abstract.....81

2.2 Introduction 82

2.3 Material and Methods.....89

2.3.1 Background 89

2.3.2 Edge-corrected Effective Population Size 90
2.3.3 Two-dimensional spatial model at equilibrium 91
2.3.4 A statistical approach for testing molecular edge-effect 94
2.3.5 Two-dimensional spatial model at non-equilibrium 95
2.3.6 Population genetics simulations 99
2.3.7 Simulated scenarios 101
2.3.8 Genetic diversity 102
2.3.9 Demographic analysis 103
2.4 Results 104
2.4.1 Equilibrium 104
2.4.2 Non-Equilibrium 109
2.5 Discussion 113
2.5.1 Edge-effect in equilibrium populations (habitat patches) 113
2.5.2 Edge-corrected population size: a proxy of edge-determined changes in genetic diversity 115
2.5.3 Non-Equilibrium pattern of genetic diversity during habitat contraction 118
2.5.4 Possible applications and future implementations 121
2.6 References 123
2.7 Supplementary Figures 137

Chapter 3: The effect of habitat loss and fragmentation on isolation-by-distance and time

3.1 Abstract 155
3.2 Introduction 156
3.3 Materials and Methods 163
3.3.1 Population genetics simulations 163
3.3.2 Simulated scenarios 165
3.3.3 Measuring spatial genetic patterns 166
3.3.4 A toroidal stepping-stone model under HL&F 167
3.4 Results 177

3.4.1 Instantaneous HL&F 177
3.4.2 Range expansion prior to instantaneous HL&F 180
3.4.3 Two-steps HL&F 181
3.4.4 Population divergence in space and time 183
3.5 Discussion 189
3.5.1 Overview of the study 189
3.5.2 Instantaneous HL&F 192
3.5.3 Range expansion priori to instantaneous HL&F 194
3.5.4 Two-steps HL&F 197
3.6 Conclusions 199
3.6.1 Limitations of the study and Future Directions 201
3.7 References 204
3.8 Supplementary Figures 222

Chapter 4: Landscape genomics of tufted-tailed rats (*Eliurus* spp.) in northern Madagascar reconstructing past forest cover changes from forest-dwelling species

4.1 Abstract 237
4.2 Introduction 238
4.3 Materials and Methods 242
4.3.1 Study area 242
4.3.2 Study species 247
4.3.3 Sampling and Laboratory procedures 249
4.3.4 RAD-sequencing 250
4.3.5 RAD-seq data processing and de novo assembly 251
4.3.6 RAD-seq datasets building 251
4.3.7 Clustering analyses (ADMIXTURE, PCA) 252
4.3.8 Genetic Diversity and Differentiation 253
4.3.9 Effective migration surfaces 254
4.4 Results 257
4.4.1 RAD-data processing: assembly of nuclear loci de novo 257

4.4.2 Genomic diversity within Eliurus species 259
4.4.3 Population Genomic structure within Eliurus species 262
4.4.3.1 Principal Component Analysis 262
4.4.3.2 ADMIXTURE analysis 265
4.4.3.3 Isolation-by-distance (IBD) 269
4.4.3.4 Genomic differentiation within Eliurus species 272
4.4.3.5 Estimating Effective Migration Surfaces (EEMS) 273
4.5 Discussion 276
4.5.1 Population genomic diversity of Eliurus carletoni 276
4.5.2 Population genomic diversity of Eliurus myoxinus 281
4.5.3 Population genomic diversity of Eliurus ellermani 285
4.5.4 Population genomic diversity of Eliurus minor 288
4.5.5 Population genomic diversity of Eliurus sp. nova 288
4.5.6 Insights on forest cover changes: dry versus humid vegetation 289
4.6 References 295
4.7 Supplementary Information 313

Chapter 5: Genetic diversity of mouse lemurs and tufted-tailed rats in northern Madagascar: new species and insights on past forest cover changes

Genetic and morphological diversity of mouse lemurs (Microcebus spp.) in northern Madagascar: The discovery of a putative new species?
5.1 Abstract 331
5.2 Introduction 332
5.3 Material and Methods 336
5.3.1 Study species 336
5.3.2 Study sites 337
5.3.3 Sampling and laboratory procedures 338
5.3.4 Phylogenetic analyses 339

5.3.5 Species validation 340
5.3.6 Morphological analyses 341
5.4 Results 343
5.4.1 Phylogenetic analyses 343
5.4.2 Species validation analyses 346
5.4.3 Morphological analyses 347
5.5 Discussion 354
5.5.1 Update of species distributions 354
5.5.2 A new location of Microcebus sympatry 355
5.5.3 Insights on recent forest changes from forest-dwelling species distributions 356
5.5.4 New mtDNA lineage 358
5.5.5 Conservation implications 360
5.6 Appendix: detailed methods 362
5.6.1 Phylogenetic analyses 362
5.6.2 Species validation 363
5.6.3 Morphological analysis 364
5.7 References 366
5.8 Supplementary Information 378

The genomic diversity of the Eliurus genus in northern Madagascar with a putative new species

5.9 Abstract 391
5.10 Introduction 392
5.11 Materials and Methods 396
5.11.1 Study region 396
5.11.2 Study species 400
5.11.3 Sampling and laboratory procedures 404
5.11.4 RAD-sequencing 405
5.11.5 RAD-seq data processing and de novo assembly 406
5.11.6 Mitochondrial DNA analyses 410

5.11.7 RAD-seq datasets building 411
5.11.8 Phylogenomic analysis 413
5.11.9 Clustering analyses (ADMIXTURE, PCA) 414
5.11.10 Species Validation 415
5.12 Results 418
5.12.1 Species identification from mtDNA cytb reference phylogeny 418
5.12.2 RAD-data processing: assembly of nuclear loci de novo 420
5.12.3 Phylogenomic-data analysis 420
5.12.4 Genomic structure 421
5.12.5 Eliurus species delimitation 422
5.12.6 Eliurus species distribution in northern Madagascar 424
5.13 Discussion 427
5.13.1 Eliurus mitochondrial genetic diversity 427
5.13.2 Nuclear genomic diversity of Eliurus species in northern Madagascar 428
5.13.3 Eliurus species delimitation 429
5.13.4 Eliurus species distribution in northern Madagascar 433
5.13.5 Eliurus sympatry: more common than rare 439
5.13.6 Conclusions and Future Directions 440
5.14 References 441
5.15 Supplementary Material 460

Chapter 6: General Discussion

6.1 Overview on scientific path: initial projects, fieldwork, conservation activities and trainings 480
6.2 Patch-scale processes in HL&F: edge-effect and habitat loss 491
6.3 Landscape-scale processes in HL&F: isolation-by-distance in the context of HL&F 499
6.3.1 HL&F in a torus-like space (stepping-stone model) 510
6.3.2 HL&F in continuous and finite space 512
6.3.3 HL&F as barrier to gene-flow in continuous space 517

6.4 Landscape genomics of forest-dwelling species in northern Madagascar: insights of past forest cover changes 521
6.5 Implications for conservation 530
6.6 References 532

Annex

7.1 Genetic Differentiation and Demographic History of the Northern Rufous Mouse Lemur (*Microcebus tavaratra*) Across a Fragmented Landscape 557
7.2 The unique and isolated *Phaner* population of Analafiana (Vohémar, SAVA) 559
7.3 The genetic structure of a mouse lemur living in a fragmented habitat in Northern Madagascar 561
7.4 Lemur inventory of the Analabe forest, SAVA region, and observation of an albino *Eulemur* 563
7.5 Distribution and conservation status of the Endangered Montagne d'Ambre fork-marked lemur (*Phaner electromontis*) 564
7.6 Population density estimates of northern rufuous mouse lemur (*Microcebus tavaratra*) and Arnhold's mouse lemur (*Microcebus arnholdi*) across northern Madagascar 566
7.7 Genetics of Malagasy mammals: An Overview 575
7.8 Population genomic diversity of three co-occurring lemur species across northern Madagascar 701

Acknowledgments

Through my educational path, many theses have been written. Starting from the primary school, secondary school, high school, Bachelor degree and Master degree. The present manuscript is my 'final' thesis: a PhD thesis. It has been a journey that necessarily cannot disregard my past, my history. It is impossible to deny that along these years of PhD, I am not the same person as I was at the beginning. Thus, my acknowledgements must necessarily go to those that, since my childhood, shaped my behaviour, values and vision of the world, and to those that along these 6 years have guided me to be a critical scientist and citizen.

In particular, I want to thank my father and my mother. There are too few official documents where these words have been written and publicly recorded, and I want to reinforce it here. I thank my father and my mother because they always believed and supported my dreams and aspirations. An invaluable gift that too rarely sons and daughters have access to, and for which I will never stop acknowledging them. I must thank my entire 'Sicilian' family, which gave me the human tools for living this life with curiosity, hopefully humbleness and joy. A special thank goes to my brother Franscesco, my great companion in the discovery of my/our passion for science and dreams: him as a particle physicist, me as an evolutionary biologist. Since our childhood time spent on toy microscopes and toy chemical experiments to our present-day discussions on the warm and sunny beaches of our Sicily during our family holidays.

I want also to thank Sonia Scalco and Carlo Carzan for their dedication to fight for a better world, for having taught me and my brother the values of a good citizen and for having created a place, the toy library 'Le Terre del Gioco', where we could learn the values of justice, altruism and creativity. These were definitely years that shaped my soul.

I want to thank my high school professors of the 'Liceo Scientifico Galileo Galilei': the Italian Literature teacher Prof.ssa Barletta for transmitting

me the power of literature and critical thinking, to have cultivated in me art and political discussions, the Phylosophy and History teacher Prof.ssa Passarello for introducing me to philosophy not as a dogmatic high school subject of study but as a tool of development in human society and for our intense philosophical discussions during classes, the Mathematics teacher Prof.ssa Equizzi for having showed and taught me the beauty of mathematics; the Physics teacher Prof. Stira for having taught me with his creative attitude, sometimes harsh, physics.

A special thanks goes also to the Scout Community of which I was part while still living in Palermo. To Mimmo and Chiara for made me live my dream of explorer in the Sicilian forests, to teach me that everything can be done and that we must believe that no limits exist to our will. To Palermo 11, a great Scout community that helped me in the more difficult moments, and that made me believe that we could fight for our ideals as other man and women have done through history. For having me taught that we are not alone, that we must work for making this world better than how we have found it, to teach me how to relate with MYSELF, the OTHERS, and the WOLRD. In particular, I want to thank PeppeMeli for this.

Another special thank is for those professors that have inspired me during my studies at the University of Padova. To Prof. Antonella Russo for having me closely followed during my years at the University and beyond, to have helped me to achieve my objectives and for having believed on me. To Andrea and Elisa, at that time PhD students of Prof. Giorgio Valle, for having organised stimulating and fascinating lab practicals in molecular biology which even today (> 7 years later) remained an important experience that helped me to understand and appreciate the power and creativity behind molecular biology. For those evenings spent performing experiments and those afternoons spent collecting water samples for metagenomics practicals. Another big thank goes to Prof. Rodolfo Costa, which probably won't remember of me, but that with his classes on genetics showed me the most profound approach to science, which goes beyond methods and techniques,

and rather touches the fundamental philosophy in biology. It was also thanks to Prof. Rodolfo Costa that I met Prof. Guido Barbujani, invited to teach in one of Prof. Rodolfo Costa classes on genetics. In that afternoon, Prof. Guido Barbujani presented a talk on human genetic variation, after which it became quite clear in my mind that I wanted to learn and understand population genetics. Surprisingly, that afternoon influenced several decisions on my academic path, until now.

After my Master degree in Paris and a hitchhiking trip in the beautiful lands of South-America, I decided to work on the publication of my Master thesis, developed in collaboration with my current PhD supervisor, Dr. Lounès Chikhi, and one of his previous PhD students, now Dr. Jordi Salmona. To do that, I moved to Lisbon and spent four months in Lounès research group, at the Instituto Gulbenkian de Ciencia (IGC). These four months had a tremendous and wonderful impact on my life path. I had fallen in love with IGC, Lounès research group and Portugal, creating the ideal conditions for my PhD journey. It was at about 2 months after my arrival in IGC that I was accepted to the IBB PhD program of IGC. Waiting for the start of my PhD, Lounès gave me the chance to work for 5 months as a volunteer in one of the expeditions organised in the forests of northern Madagascar. During this first field mission in Madagascar, I met those that later on, during my PhD, had become friends and my main fieldwork collaborators. I refer in particular to Emmanuel Rasolondraibe, Tantely Ralantoharijaona, Ando Rakotonanahary and Jacquis Randriamaroson. In particular, I want to thank Emmanuel for his constant support and leadership during our fieldwork, for making me know and understand Madagascar and the Malagasy culture, for teaching me all that I needed to know about fieldwork and northern Madagascar biodiversity. I want to thank Tantely, the great photographer (and not only) of our team, for the good spirit he has always brought to the field team and to our late-night appetizers in company of a good Pastisse and peanuts after a long day of fieldwork activity. I want to thank Jacquis for his constant help during fieldwork, during our long trips in motorcycle across the arduous roads of

northern Madagascar and the great hospitality has always shown to me. And I want to thank Ando, a great person and researcher, that helped me a lot under critical situations during some of our field mission. Misaotra betsaka djabe! The fieldwork would not have been the same without Barbara Le Pors, the fieldwork leader, together with Emmanuel, of our team and also my flatmate here in Lisbon. I am immensely in debt with her, an example of a person that love life and all the surprises it brings, that accompanied me during several fieldwork missions, that put me in a closer understanding of the Malagasy culture. I want to also thank Lucienne Wilmé and Patrick Weber for useful debates on wildlife conservation in Madagascar and for their profound sense of 'resistance' and determination in fighting the loss of biodiversity in Madagascar.

Finally, I want to thank all past and present members of the Population Genetics and Conservation Group here in IGC. Isa Pais for welcoming me in IGC and be always a positive vibe in the group, my great companion of simulations Tiago Maié, the knowledgeable and fair Barbara Parreira, Tania Minhos, Filipa Borges, Ines Carvalho, Hugo Laine, Adam Marques, Remi Tournabize, Margarida Cardoso, Ravi Vishwakarma for all the moments shared along these years, and with some along these months, related or not to science.

I want to spend a 'few' more words for the two persons that have welcomed me in IGC and have created the condition for this PhD experience to be as it has been: Jordi Salmona and Lounès Chikhi.

I want to thank Jordi because I found in him a friend, a scientific partner, and a person that prove how the joy of life brings towards great life adventures. I want to thank Jordi because it has represented an example to me during my PhD, to look at for his brave to always go ahead and not be afraid.

I want to thank Lounès because he is the person that during my PhD has influenced the most who I am today. I want to thank him for giving me the chance to get close to my dreams, to explore the forests of Madagascar, for leading me to a more critical and fair thinking, as evolutionary biologist as well

as citizen of the world. I want to thank him for his contagious optimism and appreciation of the diversity and quality that each person holds. I want to thank him for having always supported my choices, for celebrating all my little achievements, for guiding me but also for letting me make my own mistakes. I want to thank him for supporting my scientific freedom. I know that PhD students rarely have such opportunity, and for this my acknowledgements will probably never be enough. Why? Because these years have probably shaped me for the rest of my years, as scientist and as citizen. Thank you, Lounès…although, not even 100 parmigianas will be enough for thanking you.

My experience as PhD student of Lounès would not have been the same without the community of IGC. A place where the word 'community' is expressed closely to its true meaning, where labs, tools and material are shared, where thoughts and achievements are shared. With all the respect, I do not want to thank in this moment the people that "lead" the IGC, I want to thank the people that are themselves in IGC and make it diverse, stimulating and exciting. I want to thank every single student, post-doc, technician and PI that brings his/her/their whole spirit in this place, at any context and at any moment. I want to thank them because they truly make IGC a 'flavoured' environment where joy, "frivolousness", ingenuity, anger, hope, determination, curiosity, merge. To you all, go my biggest THANK YOU. In particular, I want to thank Roberto, Renato, Massimo, Shafi, Ines, Wilson, Mariana, Ojas, Dragan, Catarina, Gonçalo, Rafa, Alba, Abdul, Yara, Carmen, Anton, Elias, Isabel, Aleikos, Ana Aranda, Jorge, Elio, and many others that probably are not coming into my mind in this night in Davis, CA.

I want to conclude these long acknowledgments, thanking my friends Roberto, Massimo, Wilson, Renato, my various flatmates of Rua Castelo Branco Saraiva Adrien, Wilson (again), Wally, Duda, and many others that have passed by. I want to thank my recent love, Carolla…for bridging my passion for art and science, for your patience, strength and continuous inspiration.

Summary

Throughout Earth's natural history, habitats have undergone drastic changes in quality and extent, influencing the distribution of species worldwide. In the last few hundred years, human activities have destroyed natural habitats at an unprecedent rate, converting continuous habitat into fragmented and isolated patches (e.g., habitat loss and fragmentation, HL&F). Ecological and genetic studies have shown that such environmental changes have an overall negative effect on biodiversity. This is alarming notably for tropical regions, characterised by high rate of species richness and endemism, where anthropogenic habitat destruction has been particularly strong in the last few decades. Investigating the consequences of HL&F on biodiversity is therefore pivotal i) for our understanding of how species and habitats have changed over the last thousands of years, and ii) for predicting the effect of recent habitat destruction on the future of natural ecosystems.

Historical records, such as palaeoecological data, have provided useful insights on how species have responded to past habitat changes. Still, the palaeoecological approach has drawbacks that limit our understanding of fine-scale species responses. Genetic data, with their own limitations, can complement paleoenvironmental reconstructions, by inferring, for instance, past changes in population connectivity and abundance across species. However, the ability of correctly inferring changes in connectivity and population abundance requires that we understand the processes underlying spatial and temporal changes in genetic variation in the context of HL&F.

In this PhD thesis, I have investigated the genetic consequences of HL&F by undertaking two complementary approaches that address this question considering various spatial and temporal scales: i) a theoretical approach, using spatially-explicit genetic simulations and analytical models;

and ii) an empirical approach, using genetic and genomic data from endemic species occurring in the fragmented forests of northern Madagascar.

Following the **theoretical approach**, I performed spatially-explicit simulations to study the effect of habitat contraction on the spatial distribution of genetic diversity, taking into account two processes relevant at this spatial scale: patch-size and edge-effect. I propose a statistical approach to estimate 'edge-impacted population size' from habitat cover information and use this measure to predict spatial genetic diversity in both equilibrium and non-equilibrium populations. Then, I consider how landscape-scale spatial genetic variation changes after HL&F, focusing in particular on isolation by-distance (IBD) patterns among habitat patches. To this aim, I have investigated through spatial simulations the behavior of IBD under scenarios of HL&F that are common in natural populations, and which include: i) instantaneous HL&F, ii) gradual (two-steps) HL&F, and iii) instantaneous HL&F following range expansion. In addition, I propose a theoretical model of *isolation-by-distance in time* in the context of HL&F, obtained by modifying the original theorical results on isolation-by-distance (Slatkin, 1991; Slatkin, 1993) and applying them to a toroidal stepping-stone model in the context of HL&F.

The **empirical approach** consisted in the analyses of genetic and genomic data of endemic forest-dwelling primates (mouse lemurs, *Microcebus* genus) and rodents (tufted-tailed rats, *Eliurus* genus) occurring in the fragmented forests of northern Madagascar. In northern Madagascar, we identified four previously known *Eliurus* species and proposed the discovery of a new putative species (henceforth called *Eliurus* sp. nova). Population genomic analyses across the five *Eliurus* taxa showed higher genetic diversity in individuals sampled in recently connected forest sites than in isolated forest fragments. We identified evidence of recent expansion in four *Eliurus* species, associated to either dry or humid habitat, suggesting that both habitat types-associated species have undergone range expansion at some point in the recent past, probably following the climatic oscillations that occurred during the last 25,000 years. Ultimately, genetic and genomic phylogeographic

patterns recovered from both mouse lemurs and tufted-tail rats suggest that humid forests were probably covering the present-day dry regions of northern Madagascar in the recent past. In conclusion, the projects carried out during this PhD included both theoretical and empirical works. The latter required fieldwork in northern Madagascar, which gave rise to other works that are not the main subject of this PhD thesis. In particular, these studies included i) assessment of primate biodiversity across remote forests of northern Madagascar; ii) update of mouse lemur and tufted-tailed species distribution; iii) the estimation of population density for two mouse lemur species, the dry-associated *M. tavaratra* and the humid-associated *M. arnholdi*; iv) the estimation of relative population density for the Montagne d'Ambre fork-marked lemur (*Phaner electromontis*) and identification of the environmental variables that explains the species' distribution; and v) a review on the "genetics of Malagasy mammals" synthetising the use and interpretation of genetic and genomic data generated over the last decades on terrestrial Malagasy mammals. These side-projects contributed to update the IUCN conservation status of several lemur species and thus represent valuable knowledge for the conservation of biodiversity in the threatened forests of Madagascar.

Resumo

Ao longo da história natural da Terra, os habitats sofreram mudanças drásticas em qualidade e extensão, influenciando a distribuição das espécies em todo o mundo. Nas últimas centenas de anos, as atividades humanas destruíram habitats naturais a uma taxa sem precedentes, convertendo habitats contínuos em manchas fragmentadas e isoladas (por exemplo, perda e fragmentação de habitat, HL&F). Estudos ecológicos e genéticos têm demonstrado que tais mudanças ambientais têm um efeito geral negativo sobre a biodiversidade. Isto é alarmante principalmente em regiões tropicais, caracterizadas por uma alta taxa de riqueza de espécies e endemismo, onde a destruição de habitats antropogénicos tem sido particularmente forte nas últimas décadas. Investigar as consequências da HL&F na biodiversidade é, portanto, fundamental i) para a nossa compreensão de como as espécies e habitats têm vindo a mudar nos últimos milhares de anos e ii) para prever o efeito da destruição recente de habitats no futuro dos ecossistemas naturais.

Registros históricos, como dados paleoecológicos, forneceram informações úteis sobre como as espécies responderam a mudanças de habitat no passado. Ainda assim, a abordagem paleoecológica tem desvantagens que limitam a nossa compreensão das respostas das espécies em escala fina. Dados genéticos, com as suas próprias limitações, podem complementar as reconstruções paleoambientais, inferindo, por exemplo, mudanças passadas na conectividade populacional e abundância entre as espécies. No entanto, a capacidade de inferir corretamente mudanças na conectividade e abundância populacional requer que entendamos os processos subjacentes às mudanças espaciais e temporais na variação genética no contexto de HL&F.

Nesta tese de doutorado, investiguei as consequências genéticas da HL&F realizando duas abordagens complementares que abordam essa questão considerando várias escalas espaciais e temporais: i) uma **abordagem teórica**, usando simulações genéticas espacialmente explícitas

e modelos analíticos; e ii) uma **abordagem empírica**, usando dados genéticos e genómicos de espécies endémicas que ocorrem nas florestas fragmentadas no norte de Madagáscar.

Seguindo a **abordagem teórica**, realizei simulações espacialmente explícitas para estudar o efeito da contração do habitat na distribuição espacial da diversidade genética, tendo em consideração dois processos relevantes nessa escala espacial: tamanho de mancha e efeito de borda. Proponho uma abordagem estatística para estimar o 'tamanho populacional impactado na borda' a partir de informações de cobertura de habitat e uso essa medida para prever a diversidade genética espacial em populações de equilíbrio e não-equilíbrio. Em seguida, considero como a variação genética espacial muda numa escala de paisagem depois de HL&F, tendo como particular foco os padrãos de isolamento por distância (IBD) entre manchas de habitat. Para tal, e por meio de simulações espaciais, investiguei o comportamento do IBD em cenários de HL&F comuns em populações naturais, e que incluem: i) HL&F instantâneo, ii) HL&F gradual (dois passos) e iii) HL&F instantâneo seguindo a faixa de expansão. Além disso, proponho um modelo teórico de isolamento por distância no tempo no contexto de HL&F, obtido através da modificação de resultados teóricos originais sobre isolamento por distância (Slatkin, 1991; Slatkin, 1993) e aplicando-os a um toroidal modelo de trampolim no contexto da HL&F.

A **abordagem empírica** consistiu na análise de dados genéticos e genómicos de primatas endémicos da floresta (lémures camundongos, género *Microcebus*) e roedores (ratos de cauda tufada, género *Eliurus*) que ocorrem nas florestas fragmentadas do norte de Madagáscar. No norte de Madagáscar, identificamos quatro espécies de *Eliurus* previamente conhecidas e propusemos a descoberta de uma nova espécie putativa (doravante denominada *Eliurus* sp. nova). Análises genómicas populacionais nos cinco táxons de *Eliurus* mostraram maior diversidade genética em indivíduos amostrados em áreas florestais recentemente conectadas do que em fragmentos florestais isolados. Identificamos evidências de expansão

recente em quatro espécies de *Eliurus*, associadas a habitats secos ou húmidos, sugerindo que ambas as espécies associadas a estes tipos de habitats sofreram expansão de alcance em algum momento no passado recente, provavelmente seguindo as oscilações climáticas que ocorreram durante os últimos 25.000 anos. Em última análise, os padrões filogeográficos genéticos e genómicos recuperados de lêmures de camundongos e ratos de cauda em tufos sugerem que as florestas úmidas provavelmente cobriam as atuais regiões secas do norte de Madagáscar no num passado recente. Em conclusão, os projetos realizados durante este doutoramento incluíram tanto trabalhos teóricos como empíricos. Este último exigiu trabalho de campo no norte de Madagáscar, o que deu origem a outros trabalhos que não são o tema principal desta tese de doutorado. Em particular, esses estudos incluíram i) avaliação da biodiversidade de primatas em florestas remotas do norte de Madagáscar; ii) atualização da distribuição das espécies de lémures e caudas tufadas; iii) a estimativa da densidade populacional para duas espécies de lémures camundongos, o *M. tavaratra* associado à seca e o *M. arnholdi* associado à humidade; iv) a estimativa da densidade populacional relativa para o lémure-marcado em forquilha Montagne d'Ambre (*Phaner electromontis*) e identificação das variáveis ambientais que explicam a distribuição da espécie; e v) uma revisão sobre a "genética dos mamíferos malgaxes" sintetizando o uso e interpretação de dados genéticos e genómicos gerados nas últimas décadas sobre mamíferos malgaxes terrestres. Esses projetos paralelos contribuíram para atualizar o estado de conservação da IUCN de várias espécies de lémures e, portanto, representam um conhecimento valioso para a conservação da biodiversidade nas florestas ameaçadas de Madagáscar.

Chapter 1

General Introduction

1.1 Current estimate of HL&F in ancient and recent times

Throughout Earth's natural history, habitats have undergone drastic changes in quality and extent, depending on their latitude and topography (e.g., natural barriers) (Hewitt, 2000; Williams, 2003; Sandel et al., 2011; Weigelt et al., 2016). The Late Quaternary period is a period for which high resolution records have provided major insights on the environmental changes that are associated with earth climate and biodiversity change in the last 0.5 – 1 million years ago (Tierney et al., 2010; Sinninghe Damsté et al., 2011; Ivory et al., 2012; Tierney and deMenocal, 2013). Worldwide Quaternary environmental changes were associated with the growth and decay of continental ice sheets due to the Milankovitch cycles (consisting of changes in Earth's orbit, obliquity and direction of Earth's axis rotation) (Bradley, 1985; Imbrie and Imbrie, 1986), which consequently affected earth ecosystems (Denton et al., 2010). Those changes have affected the distribution of species worldwide (Graham et al., 1996; Williams et al., 1998; Lyons, 2003, 2005). For example, mammals moved either south or north, as late Quaternary climates were cooling or warming, respectively (Graham et al., 1996). These climatic oscillations contributed also to megafaunal extinctions (Koch and Barnosky, 2006). Paleorecords show that continents were populated with more than 150 genera of megafauna about 50K years ago, and by 10K years ago more than 90 of those genera got extinct (Barnosky et al., 2004). Evidence from palaeontology, climatology and archaeology support now that humans also contributed to the extinction of megafauna species (Alroy, 2001; Burney and Flannery, 2005; Prescott et al., 2012; Sandom et al., 2014; Broughton and Weitzel, 2018).

Environmental changes during the Quaternary period, however, did not occur at the same rate and with the same intensity across the world (Loarie et al., 2009; Sandel et al., 2011). For instance, while high latitudes regions were characterised by relative climatic 'stability' during the Holocene (11.7 Kya – 0.0 Kya) (Walker et al., 2012), part of Africa (north of the Equator) underwent major climatic fluctuations (Hassan, 1997; Gasse, 2000). The most noteworthy example is the so called 'African Humid Period' (AHP) (14.8 – 5.5 Kya; deMenocal et al., 2000) which represents a period significantly more humid than present-day, with major ecosystem changes such as the greening of the Sahara (Kutzbach and Street-Perrott, 1985; Gasse, 2000). Thus, these and other evidence point out to the fact that, over time, ecological communities have shifted or disappeared, but also that the properties and intensity of climatic change were heterogeneous across the globe.

More recently, human activities have dramatically modified natural environments (Barlow et al., 2016). The magnitude, variety and long-term effect of human-induced environmental changes suggest that Earth may have indeed entered a new geological epoch, often referred to as the 'Anthropocene' (Waters et al., 2016). The starting time of the 'Anthropocene' is currently debated (Lewis and Maslin, 2015), but two possible candidates are: i) the 1600, characterised by a *Orbis Spike* dip in CO_2 due to the near-cessation of farming and consequently to the re-growth of Central and South American forests, following the deaths of ~50 M Native Americans during the European colonization of the continent; and ii) 1964, characterised by a bomb spike in the radioactive isotope ^{14}C due to nuclear bomb testing.

One tangible form of anthropogenic environmental change is deforestation. Historical and current data show that human-mediated deforestation was already common in temperate regions between 18th and 20th centuries (Williams, 2003), and increased in the tropical regions over the last fifty years (Hansen et al., 2013; Song et al., 2018). Recently, the

development of satellite remote sensing technologies has provided an unprecedent quantification of the Earth's ecosystem status (White et al., 2016), and a better understanding of the impact of anthropogenic activities on natural habitats. For instance, the first global analysis of forest cover has shown that in 2000 over 70% of forest areas were within 1 km of an edge (that is in close proximity to human-modified non-forest ecosystems) (Haddad et al., 2015) (Fig. 1a, b). Another global-scale study has quantified global land changes during the period 1982-2016 showing that, contrary to the general narrative of a global forest decline, forest cover has actually increased by 7% since 1982 (Song et al., 2018). Yet, this apparent optimistic estimate is actually the result of a net loss in forest cover in the tropics (e.g., tropical dry forest: -8%; tropical moist deciduous forest: -2%) and, indeed, a net gain in tree cover in the extra-tropics (e.g., temperate forest: +33%; boreal forest: +12%) (Fig. 1c). Song et al., (2018) have also quantified the relative contribution of human activities on global land changes, finding that 60% are associated with direct human activities and 40% with indirect drivers such as climate change. What we can learn from these global analyses is that humans are definitely reshaping earth ecosystem at an extremely high rate, but that such environmental changes are not equally distributed across the globe. In fact, in the last few hundred years, human-mediated environmental changes have geographically shifted from temperate to tropical regions.

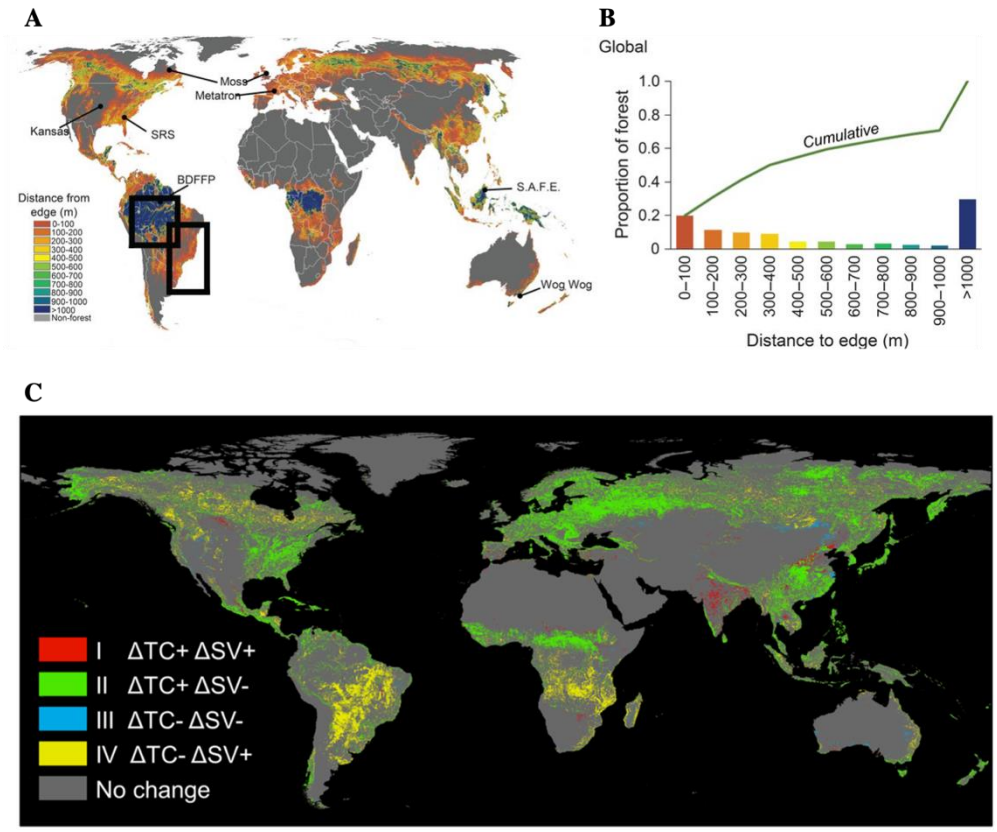


Figure 1. Global extent of habitat loss and fragmentation. A) Average distance to forest edge for forested pixels within each 1-km cell. Blue: continuous forests; Red: severely fragmented forests; B) Distribution of the world's forest at each distance to the forest edge (modified from Haddad et al., 2015). C) Land-cover change from 1982 to 2016. ΔTC: changes in tree canopy cover; ΔSV: changes in short vegetation cover (modified from Song et al., 2018).

Tropical forests are one of the many ecosystems on Earth that play a major role for the ecological 'health' of the entire planet (Malhi and Phillips, 2005), i.e. from regulating carbon storage, water supply and climate stability, to be one of the major *reservoirs* of global biodiversity (Barlow et al., 2007; Gibson et al., 2011; Baccini et al., 2017). In the last decades, we have obtained much accurate estimates of how human activities and climate change have changed such ecosystems. The current data are somewhat

alarming. Between 1990 and 2015 (25 years) only, about 10% of tropical forest cover has been lost (Keenan et al., 2015). These estimates can be even larger, depending on the vegetation type, forest quality and the region of the world. For instance, worldwide tropical moist forests (undisturbed and degraded forests) have declined by 17% between 1990 and 2019 (Vancutsem et al., 2021). Within tropical moist forests, the decline of only undisturbed forest cover (i.e., excluding degraded forests) reaches about 24% since 1990. These estimates are even more dramatic if we consider the existing variation within continents, in fact the estimates rise from 17%, averaging across the world, up to 81.5% in Ivory Coast, 73.7% in Mexico, 70.8% in Ghana, 69% in Madagascar, 67.8% in Vietnam, 67.1% in Angola, 65.8% in Nicaragua, 65.1% in Lao People's Democratic Republic (PDR) and 63.9% in India (Vancutsem et al., 2021). As noted by Vancutsem et al., (2021), the real status of forest cover threats is likely underestimated given that as much as 45.4% of the forest degradation (not actual deforestation) is predicted to undergo actual forest clearance within 7.5 years, on average. This means that the current extent of tropical humid forest destruction is certainly going to increase in the next 7-8 years, unless a concrete effort is put in place for restoring degraded forests.

1.2 The processes underlying habitat loss and fragmentation

The destruction and degradation of natural ecosystems typically leads to habitat loss and fragmentation (HL&F), a process characterized by a loss in the amount of habitat area and the splitting of habitat into smaller patches that are separated by a matrix of relatively unsuitable habitats (Curtis, 1956; Moore, 1962; Fahrig, 2003).

HL&F is often a very complex process, which can follow different spatial and temporal trajectories in different landscapes, resulting in diverse

landscape configurations. In front of such complexity, it can be useful to start by describing the 'fundamental' effects that HL&F has on a continuous landscape, namely: *i*) the decrease in the amount of habitat area; and *ii*) generation of (or increase in) number of habitat fragments (or patches) (Fahrig, 2003). These two processes occur at the spatial scale of a landscape (*landscape-scale*) (Fig. 2).

However, *landscape-scale* processes are, in turn, the result of processes occurring at the smaller spatial scale of the habitat fragments (*patch-scale*), which can be summarised as being: i) decrease in patch-size; ii) increase in edge-effect; and *iii*) increase in isolation of habitat fragments (i.e., reduction in connectivity between patches) (Fig. 2). Among the two *landscape-scale* effects and the three *patch-scale* effects mentioned above, perhaps edge-effect might be the less intuitive for the reader, this is why I will focus for a moment on describing it. *Edge-effect* represents the change in biophysical environments at the boundary between two or more habitats (Didham and Lawton, 1999). The impact of *edge-effect* on the spatial and temporal stability of a given ecological community depends on several factors, such as climate and habitat type (Didham and Lawton, 1999). For a given habitat, in the more extreme cases, *edge-effect* can cause local species extinction, over spatial scale that exceed 1 km of distance from the habitat edges, towards the interior of the habitat (Ewers and Didham, 2008). Ultimately, the degradation in habitat quality (vegetation structure) and the distance at which changes in biophysical property penetrate within a given habitat determine the amount of available habitat within a given patch. As a consequence, edge-effect reduces habitat amount, beyond the boundaries of the habitat patch.

In the HL&F literature, the term *habitat fragmentation* has been often used interchangeably for referring either to i) both the reduction of habitat amount (habitat loss) and the breaking apart of a continuous landscape into

smaller habitat patches (fragmentation *per se*), or ii) exclusively to the breaking apart of a continuous landscape into smaller habitat patches (fragmentation *per se*) (Fahrig, 2017; Fletcher et al., 2018). In this section, I have introduced the processes underlying HL&F, distinguishing those occurring at the *landscape*- and *patch-scale*, hoping to lead the reader into a simplified and clearer description of HL&F. In reality, habitat loss and habitat fragmentation processes are intertwined and co-occur, such that habitat loss rarely takes place independently from fragmentation *per se*. However, as it has been pointed out in Fletcher et al., (2018), the distinction of habitat loss, fragmentation *per se*, and *landscape*- and *patch-scale* processes may be useful to better discriminate the mechanisms by which HL&F impact biodiversity, and in turn dissolve controversies and reduce the risk of confusion (Ruffell et al., 2016; Fletcher et al., 2018).

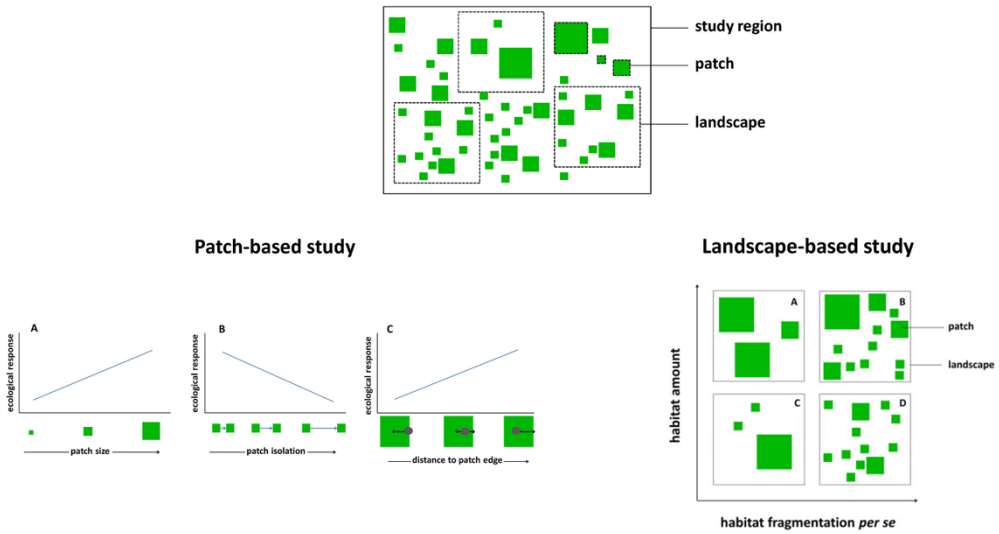


Figure 2. Habitat loss and fragmentation processes at the patch- and landscape-scale. Patch-based study: patch size, patch isolation, patch edge-effect. Landscape-based study: habitat amount, habitat fragmentation *per se*.

1.3 Effects of HL&F on biodiversity

1.3.1 Ecological response to HL&F

Scientific research investigating the impacts of HL&F on biodiversity began already in the late 1950s – early 1960s (Fahrig, 2019), with Curtis, (1956) and Moore, (1962). For instance, the paper of Moore, (1962) aimed at "illustrating some of the scientific problems of this relatively new form of applied biology [conservation] by means of an example, the conservation of the Dorset heaths". In this study, N. W. Moore investigated the present (1960) and past changes (1811) in landscape structure of the heathlands in Dorset, and asked how 'indicator' species respond to changes in land use. His findings were suggesting that increased habitat isolation was "causing slight impoverishment of the fauna". Since the 60s, many other studies have investigated the effect of HL&F on biodiversity. Most of the HL&F literature has been focusing on the effect of HL&F on ecological responses, such as species richness or species abundance. Only recently, with the advent and widespread use of genotyping approaches, researcher have started to investigate the consequences of HL&F on species genetic diversity.

The negative effect of HL&F on biodiversity is overall well established (Pereira et al., 2010; Tittensor et al., 2014; Newbold et al., 2015; Wilson et al., 2016). Global studies have shown that HL&F cause decline in local terrestrial biodiversity (Gibson et al., 2011; Pimm et al., 2014; Hanski, 2015; Newbold et al., 2015; Ceballos et al., 2015, 2017). For instance, Newbold et al., (2015) estimated that land-use and related changes have globally reduced within-site species richness by ~14% and total species abundance by ~11%. In landscapes with stronger habitat degradation, their estimates are even more dramatic: a decrease of ~77% in within-site species richness and of ~40% in total species abundance. Ceballos et al., (2017) quantified the extent of range contraction experienced between 1900 and 2015 by ~28,000

vertebrate species (about half of the known vertebrate species). They showed that 32% of the studied species have decreased in population and range size, and documented cases of local extinction for 177 species among which more than 40% have experienced over 80% in range contraction.

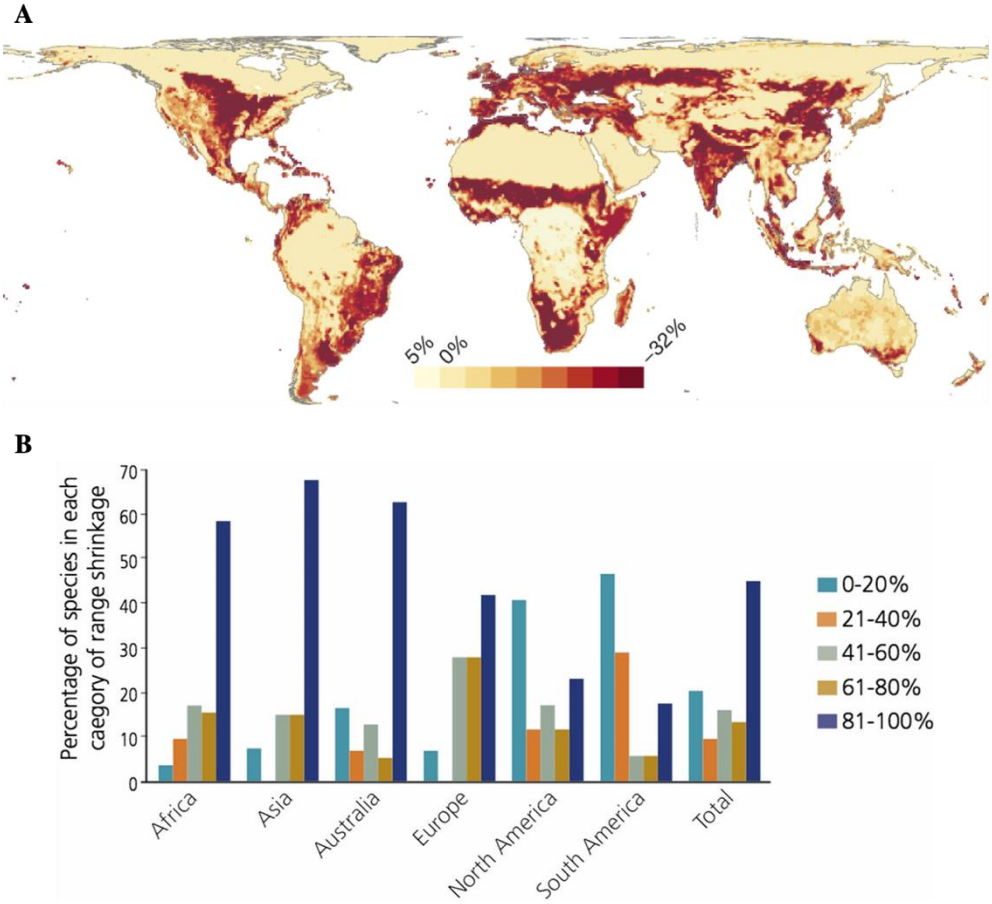


Figure 3. Habitat loss and fragmentation processes at the patch- and landscape-scale. A) Global change in local species richness from 1500 to 2000 (source: Newbold et al., 2015); B) Range contraction for terrestrial mammal species between 1900 and 2015 (source: Ceballos et al., 2017).

These two examples of global analyses provide insights on the overall effect of HL&F on biodiversity using metrics of habitat disturbance which are

discrete categories of either land-use (e.g., minimal, light or intense) or changes in species range area (e.g., class 1: 0 - 20%; class 2: 21 - 40%; etc.). Their approach is definitely useful for a *global-scale* assessment of biodiversity loss due to HL&F. However, a clear understanding of how species respond to HL&F may require a careful investigation of the mechanisms by which HL&F impact biodiversity (Fletcher et al., 2018). This is related to the idea that the relative role of the processes underlying HL&F "depend greatly on landscape structure and heterogeneity, and hence there cannot be a universal answer to which is more important" (Hanski, 2005, 2011). I refer, in particular, to the *landscape*- and *patch-scale* processes described in the previous section and, to the distinction made in the literature between the mechanisms underlying *habitat loss* and/or *habitat fragmentation per se*. In fact, although the consequences of HL&F on biodiversity have been largely investigated over the last decades, some controversies still exist (Fahrig, 2017; Fletcher et al., 2018; Fahrig et al., 2019).

There is a general agreement that *habitat loss* has detrimental effect on biodiversity (Brooks et al., 2002; Schipper et al., 2008; Fahrig, 2017; Fletcher et al., 2018; Fahrig et al., 2019). In contrast, the extent to which *habitat fragmentation per se* has negative effect on biodiversity is still controversial. The debate intensified after the publication of Fahrig, (2017), where the author reviewed 118 empirical studies and found that ecological responses to *habitat fragmentatio*n *per se* (fragmentation independent of habitat amount) were non-significant in over 70% of the studies and, when significant, usually positive. These findings were consistent with the so called 'Habitat Amount Hypothesis' (HAH; Fahrig, 2013), which states that species richness increases proportionally to the amount of habitat present in a 'local landscape', defined as the area where individuals of a given species are likely to disperse (Martin and Fahrig, 2012). This 'local landscape' can extend beyond the boundaries of the habitat patch from which ecological data have been retrieved, thus suggesting that patch-size (and other fragmentation-

related metrics) may not be a full predictor of the measured ecological response. To Fahrig, (2017) review followed the perspective paper 'Is habitat fragmentation good for biodiversity?' by Fletcher et al., (2018), where the authors argued that data from literature rather support a significant negative effect of *habitat fragmentation per se* on biodiversity. In particular, they were arguing that Fahrig, (2017) conclusions came from a narrow and possibly biased subset of the literature, which included i) studies using statistical analyses that have been shown to be limited in discriminating habitat fragmentation effects; ii) studies referring to simple linear effect of *habitat fragmentation per se* on biodiversity, thus excluding studies reporting non-linear trends; and iii) studies performed at the *landscape-scale*, thus ignoring *patch-scale* studies, many of which have been performed under controlled long-term experiments (see Haddad et al., 2015). In addition to these three points, Fletcher et al., (2018) report a series of other arguments to contradict the claim of Fahrig, (2017) on the likely general non-negative effect of *habitat fragmentatio*n *per se* on biodiversity. Among those, I would like to point out to the Fletcher et al., (2018) argument concerning the lack of *patch-scale* studies in Fahrig, (2017). In fact, although Fahrig, (2003) defines *habitat fragmentation per se* as a *landscape-scale* process, the ecological response observed in fragmented landscapes may actually be an emerging property of patch-scale changes due to HF, such as edge-effect. As an example of how the exclusion of patch-scale studies may have severely biased the conclusion of Fahrig, (2017), Fletcher et al., (2018) showed that including the results of only one experiment (Savannah River Corridor Experiment; Haddad et al., 2017) would have doubled the number of negative responses to *habitat fragmentation per se* reported in Fahrig, (2017).

This controversy revealed several limitations in HL&F research (Fletcher et al., 2018). First, *habitat loss* and *habitat fragmentation* are often correlated processes. Statistical analyses that do not explicitly take into account the collinearity of such processes can draw misleading conclusions

on the effect of HL&F (Didham et al., 2012). Second, responses to HL&F are difficult to generalize across taxa and ecosystems. For instance, a positive response to HL&F by a given species can actually go in favour of a detrimental effect of HL&F, if this given species is generalist or invasive (Pfeifer et al., 2017). Third, habitat and non-habitat categorization need to be carried out according to the species 'perspective' (Betts et al., 2014). Ultimately, HL&F processes have to be regarded as spatio-temporal processes, where different mechanisms may act depending on the spatial scale, generating ecological responses that rely on complex temporal dynamics (Hylander and Ehrlén, 2013; Ryo et al., 2019). Therefore, reconciling the scientific disagreement requires research that goes beyond statistical and correlative approaches, and focuses more accurately on the mechanisms acting across spatial-temporal scales (Fletcher et al., 2018).

Recently, few meta-analysis studies have explicitly investigated the relative effect of *landscape-scale* and *patch-scale* processes on ecological responses to HL&F, which may be equivalent to assessing the relative impact of *habitat loss* and *habitat fragmentation* on biodiversity. For instance, Watling et al., (2020) used data on 5,675 species from eight taxonomic groups (plants, fungi, gastropods, insects, amphibians, reptiles, birds and mammals) across 35 studies to test the 'Habitat Amount Hypothesis'. Their results suggest that 'habitat amount' had a stronger negative effect on ecological response (species density; i.e., number of species in a plot of fixed size) than measures of *habitat fragmentation per se* (patch size and isolation). In particular, they found that 'habitat amount' and 'patch-size' were positive in 91% and 73% of studies, respectively, and that 'patch-isolation' was negative in 82% of studies. Importantly, they showed that 'habitat amount' has a larger effect size (mean effect size = 0.606 ± 0.007) than 'patch-size' (mean effect size = 0.351 ± 0.007) and 'patch-isolation' (mean effect size = 0.287 ± 0.007) (Fig. 4a). These results provide, in my opinion, interesting insights about the controversy discussed in the previous paragraph. In fact, although the

findings of Watling et al., (2020) indeed show a stronger effect of HL on ecological response, they also show, in contrast to Fahrig, (2017) conclusions, that *patch-scale* processes (and therefore HF processes) matter and have a significant negative effect on species richness. Accordingly, Watling et al., (2020) states that HAH should be seen as a parsimonious approach, where *patch*-scale variables (patch area and isolation) are replaced "with a single measure of habitat amount" that provide "a useful approximation for describing the key landscape driver of species density" in the study area. Therefore, if HAH represents a "useful approximation" of *patch*-scale effects on species ecology, understanding processes at the *patch*-scale would i) provide accurate insights on the HL&F mechanisms impacting biodiversity; ii) clarify possible controversies; and ii) elucidate how and to which extent we can extrapolate *patch*-scale mechanisms at a larger spatial scale (*landscape*-scale).

For a few decades, several empirical studies have demonstrated the importance of mechanisms occurring at the *patch*-scale (e.g., edge-effects, patch-size) to explain ecological patterns (Hanski et al., 1995; Ries et al., 2004; Prevedello and Vieira, 2010; Hanski et al., 2013; Rybicki and Hanski, 2013; Haddad et al., 2015; Pfeifer et al., 2017; Ries et al., 2017; Haddad et al., 2017; Betts et al., 2019; Bueno and Peres, 2019). Haddad et al., (2015) have synthesized results of 76 studies across the five world's ongoing long-term fragmentation experiments, revealing consistent responses (mostly negative) of organisms and ecosystems to HF, related to 'patch-size', 'patch-isolation' and 'edge-effect' (Fig. 4b). Besides controlled experimental data, one of the most recent studies on the pervasive influence of edge-effect on ecological response (species abundance) assessed edge-determined changes in abundance among 1,673 vertebrate species in natural environments (Pfeifer et al., 2017). This study showed that 85% of vertebrate species were affected either positively (46%) or negatively (39%) by edge-effect. The species negatively impacted had 3.7 times more probability to be

listed as threatened (according to IUCN classification), whereas species with positive response were typically invasive or generalist species. This study adds up to the list of empirical studies pointing out to the importance of *patch*-scale processes on ecological response.

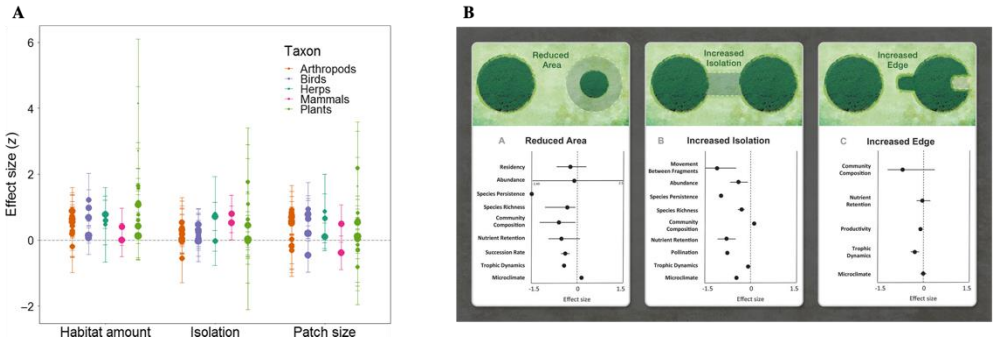


Figure 4. Ecological response to habitat loss and fragmentation. A) Mean effect size of three HL&F processes across 33 studies: one landscape-scale (habitat amount), and two patch-scale (isolation, patch-size) (source: Waitling et al., 2020); Mean effect size of three patch-scale processes across 76 studies carried out in five long-term fragmentation experiments on several ecological responses (source: Haddad et al., 2015).

1.3.2 Genetic response to HL&F

In the previous section, I have reviewed some of the studies showing the impact of HL&F (*landscape-* and *patch-scale*) on ecological response. Two metrics of ecological response are often represented in these studies: species richness and species abundance. The former refers to the number of distinct species occurring in a given area. The latter refers to the number of individuals from a given species occurring in a given area. Models in population genetics use the concept of effective population size (N_e) (or effective population density, D_e) to derive the level of 'neutral' genetic diversity that is expected at the scale of a population, where 'neutral' refers to genetic loci that are not under selection (Ellegren and Galtier, 2016). Thus, considering the evidence that HL&F decreases species abundance and

connectivity among populations, it is theoretically expected that HL&F would have an overall negative effect on species genetic diversity.

Several empirical and theoretical studies have investigated the consequences of HL&F on intraspecific genetic variation. Global meta-analyses have shown consistent evidence of an overall negative effect of anthropogenic disturbance and HL&F on species genetic diversity (Aguilar et al., 2008; Almeida-Rocha et al., 2020; Honnay and Jacquemyn, 2007; Lino et al., 2019; Miraldo et al., 2016; Monteiro et al., 2019; Rivera Ortíz et al., 2015; Schlaepfer et al., 2018; Vranckx et al., 2012). Unlike ecological studies, fewer genetic studies have explicitly investigated the relative impact of *landscape-scale* (or habitat amount) and *patch-scale* (or habitat fragmentation) on species genetic pattern. An exception is, for instance, the recent global meta-analysis of Almeida-Rocha et al., (2020) where the authors synthesized 61 studies, comparing genetic diversity in degraded and more preserved habitats/landscapes across several plant (N=31) and animal (N=38) taxa (Fig. 5). Their results showed an overall consistent negative effect of anthropogenic disturbances on the genetic diversity of terrestrial species (mean effect size: -0.45 [-0.61, -0.29]). The comparison of several HL&F metrics across spatial-scales suggests a stronger effect of *landscape*-scale than *patch*-scale processes. Among the *landscape*-scale variables, they reported 'habitat amount' (or forest cover effect size: 0.65 [-1.28, -0.02]) and 'connectivity' (effect size: -0.67 [-0.89, -0.44]). The 'connectivity' variable was defined as the level of physical isolation between habitat fragments which, according to (Fahrig, 2019), could actually be interpreted as a *patch-scale* process. Among the *patch*-scale variables, they found a slightly significant negative effect of 'local quality' (effect size: -0.08 [-0.38, -0.21]) and a stronger negative effect of 'patch size' (effect size: -0.42 [-0.65, -0.20]). Unfortunately, few genetic studies have investigated the impact of edge-effect on species genetic diversity. I am aware of two genetic studies on edge-effect, both suggesting a negative influence of edges on genetic diversity (e.g., allelic richness, A_r): i) one on dioecious tree species (e.g., A_r-edge: 3.09, A_r-

interior: 3.61) (Quevedo et al., 2013); and ii) another on mouse lemurs (A_r-edge: 7.565, A_r-interior: 8.264) (Radespiel et al., 2019). Thus, it still remains unclear to which extent data from natural populations supports the hypothesis of edge-impacted changes in genetic diversity. This reflects a general tendency in landscape genetic literature, where most studies are focused on the effect of HL&F on population genetic differentiation at the *landscape*-scale, and very few on population genetic diversity at the *patch*-scale (DiLeo and Wagner, 2016; Storfer et al., 2010).

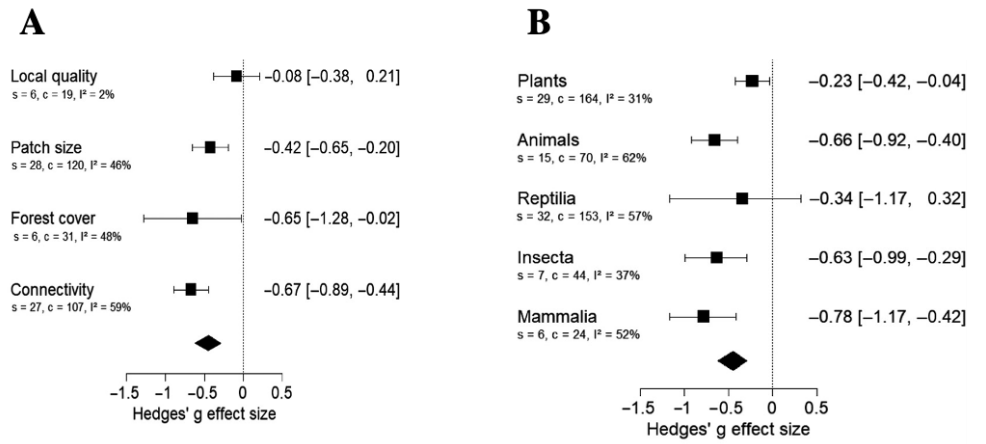


Figure 5. Mean effect size of HL&F on genetic diversity. A) effect size based on HL&F processes; B) effect size based on taxonomic group. (Source: Almeida-Rocha et al., 2020).

1.4 Spatial population genetics

Empirical studies have shown that most taxa are characterised by geographically-limited dispersal, meaning that individuals tend, on average, to reproduce at relatively short-distances from where they were born (Aguillon et al., 2017; Meirmans, 2012; Sexton et al., 2014). These observations suggest that *spatial population structure* is probably the most common form

of population structure in natural populations. If it is not the most common, at least it is sufficiently represented across taxa to deserve further investigation in order to understand the evolution of species.

The concept of geographically-limited dispersal was initially introduced by Wright, (1943), which was defined with the term '*isolation-by-distance*'. Since then, several spatial models in two-dimensions have been developed. The types of spatial genetic models could be broadly classified into three classes: i) *spatial continuous model* (Wright, 1943; Malécot, 1948); ii) *stepping-stone model* (Kimura, 1953); iii) *spatial Λ-Fleming-Viot process* (Etheridge, 2008; Barton et al., 2010).

The *spatial continuum model*, originally derived by Wright, (1943) and Malécot, (1948) with similar assumptions, consider that i) parental individuals are distributed in a two-dimensional space according to a Poisson random field with constant expected population density ρ; ii) the number of offspring for each individual is independent and sampled from a Poisson distribution with mean 1; iii) each offspring disperses from the location of the parent according to a symmetric Gaussian distribution (Fig. 6a). Under these initial assumptions, Wright, (1943) and Malécot, (1948) postulated that the offspring would also be distributed according to a Poisson random field with constant expected population density ρ. Then, using the infinitely many alleles mutation model, they derived the expressions for the probability of identity-by-descent of two genes (i.e., the probability that no mutation occurred before those genes had a common ancestor) sampled at a given distance r in the two-dimensional space, called the 'Wright-Malécot formula'. However, Felsenstein, (1975) noticed that under the initial conditions i), ii) and iii), the population will evolve and eventually form 'clumps' of individuals, which will then extinguish. Felsenstein, (1975) called the clumping-extinction problem as 'the pain in the torus'. In later works, Barton and Wilson, (1995) and Barton et al., (2002) showed that extending the Wright-Malécot formula by including local-density regulation (i.e., demes) can overcome the 'clumping-extinction'

problem identified by Felsenstein, (1975) (Eq. 3.13 in Barton and Wilson, 1995; and Eq. 14 in Barton et al., 2002). In particular, Barton et al., (2002) defined the probability of coalescence of two genes separated by a geographical distance r at time t in the past as:

$$\psi(r,t) = \gamma\, G(r,2t) \qquad (Eq.\ 1)$$

where γ is a ratio between the variance in the number of offspring for an individual and the deme population size ($1/2\rho_e$) and $G(r,2t)$ is the two-dimensional Brownian probability density function describing the movement in space and backward in time of the ancestor lineages of the sampled genes. Put in another way, $G(r,2t)$ is also equivalent to the probability of travelling a distance r in time $2t$ and can be expressed as:

$$G(r,2t) = \frac{1}{4\pi\sigma^2 t}\, e^{-\frac{r^2}{4\sigma^2 t}} \qquad (Eq.\ 2)$$

where σ^2 is the variance of the distance moved in one unit of time (generations); and t is the coalescence time (generations). In other words, the term γ describes the coalescence probability of two lineages located in the same deme, whereas $G(r,2t)$ defines the probability that two genes sampled at distance r will be in the same deme at time $2t$ in the past. Barton et al., (2002) noted that the Brownian approximation leads to almost identical results as the Wright-Malécot formula, while allowing strict local-density regulation. From this equation, it is then possible to derive the probability of identity-by-descent $f(r,t)$ of two genes separated by distance r, which is the probability that these two genes share a common ancestor t generations in the past, and that no mutations have accumulated since then:

$$f(r,t) = \gamma\, \frac{1}{4\pi\sigma^2 t}\, e^{-\frac{r^2}{4\sigma^2 t}}\, e^{-2\mu t} = \psi(r,t)e^{-2\mu t} \qquad (Eq.\ 3)$$

where $e^{-2\mu t}$ is the probability of not mutating since their most common ancestor and μ is the mutation rate per site, per generation, under an infinite-sites model. Then, the cumulative probability density function for $f(r,t)$ between $t = 0$ and $t = \infty$ is obtained with few approximations as:

$$F(r=0) \approx \frac{\log\left(\frac{\sigma}{K\sqrt{2\mu}}\right)}{4\pi\rho_e\sigma^2 + \log\left(\frac{\sigma}{K\sqrt{2\mu}}\right)} \qquad (Eq.\ 4)$$

and

$$F(r=r) \approx \frac{K_0\left(\frac{r}{\sigma}\sqrt{2\mu}\right)}{4\pi\rho_e\sigma^2 + \log\left(\frac{\sigma}{K\sqrt{2\mu}}\right)} \qquad (Eq.\ 5)$$

where K represents a 'local scale' parameter, which is equal to σ in the case of a Gaussian dispersal. These two equations are equivalent to the Wright-Malécot formula, thus they are referred as Wright-Malécot approximation (Barton et al., 2013).

The *stepping-stone model* was introduced by Kimura, (1953) and extensively studied since then (Kimura and Weiss, 1964; Maruyama, 1971, 1970; Nagylaki, 1974; Strobeck, 1987; Slatkin and Barton, 1989; Slatkin, 1991, 1993; Hey, 1991; Wilkinson-Herbots, 1998) (Fig. 6b). There exist different types of stepping-stone models, for instance: infinite, linear, circular and toroidal. Stepping-stone models are characterised by a strict local density regulation, namely populations consist of an array of colonies (discrete demes). It is assumed that, at each discrete generation, each deme exchanges a fraction m of individuals with the surrounding demes (for a

nearest-neighbour two-dimensional stepping-stone model, each deme has 4 neighbouring demes) and a proportion $1 - m$ remains in the deme. Ultimately, each deme exchanges a proportion m_∞ of migrants with a fictitious external deme that has fixed allele frequency ($\bar{p}$) equal to the average allele frequency in the population, which is equivalent to modelling reversible mutations (Kimura and Weiss, 1964). Quantities of interest in this model are, for instance, the level of genetic similarity or the mean coalescence time between two demes at a given distance. In particular, Kimura and Weiss, (1964) in *Eq. 2.22* have shown that the correlation of allele frequencies between two demes at Euclidean distance k '$corr(k)$' is proportional to:

$$corr(k) \sim \frac{e^{-\sqrt{\frac{4\mu}{m}}k}}{\sqrt{k}} \qquad (Eq.\ 6)$$

where μ is the per site mutation rate and m is the deme dispersal rate. Slatkin, (1991) has derived the expression for the mean coalescence time in a 2-dimensional toroidal stepping-stone model for two genes sampled *i* and *j* steps apart, in the x and y direction respectively (Eq. *A9* in Slatkin, 1991). It is considered that the coalescence process between the two sampled genes can be divided in two phases: *i)* the average amount of time required for two alleles sampled at a given distance to be in the same deme (S_{ij}); and *ii)* the average time to coalesce once the two alleles occupy the same deme ($\bar{t}_0$). Under a general model of dispersal, the mean coalescence time of two genes sampled at *i* and *j* steps apart can then be expressed as:

$$\bar{t}_{ij} = \bar{t}_0 + S_{ij} \qquad (Eq.\ 7)$$

Under the assumptions of the model, the average coalescence time between two genes that occupy the same deme is independent from dispersal rate and it is closely related to the total population size in the two-dimensional

stepping-stone model (*Eq. 10* and *Eq. 5* in Strobeck, 1987; Hey, 1991, respectively). That is:

$$\bar{t}_0 = 2N_e d_x d_y \qquad (Eq.\ 8)$$

where N_e is the deme population size and d_x or d_y are, respectively, the number of demes in dimension x and y. According to Slatkin, (1991) (*Eq. A9*) and Slatkin, (2005) (*Eq. 15*), the second phase of the coalescence process can be obtained from:

$$S_{ij} = \sum_{l=0}^{d_x/2} \sum_{k=0}^{d_y/2} \frac{f_{kl}\left[1 - \cos\left(2\pi ik/d_y\right)\cos(2\pi jl/d_x)\right]}{\delta(k)\delta(l)(1 - f_{kl})} \qquad (Eq.\ 9)$$

where $f_{kl} = \left[1 - m\left(1 - \cos\left(2\pi k/d_y\right)\right)\right]^2 [1 - m(1 - \cos(2\pi l/d_x))]^2$, or $f_{kl} = 0$ if $k = 0$ and $l = 0$. The coefficients $\delta(k) = 1$ or $\delta(l) = 1$ if $k = 0$ or $l = 0$, respectively, and $\delta(k) = 1/2$ or $\delta(l) = 1/2$ if $k > 0$ or $l > 0$.

The expressions for coalescence time are useful since they are closely related to measure of genetic differentiation ($F_{st(i,j)}$) and coefficients of identity-by-descent (Slatkin, 1991). Details on the average coalescence time and their application will be further discussed in Chapter 3 of this PhD thesis.

The *spatial Λ-Fleming-Viot process* is the last of the three classes of spatial genetic models. It has been recently introduced by Etheridge, (2008) and rigorously described by Barton et al., (2010a). The *spatial Λ-Fleming-Viot process* is a spatial continuum model which aims to solve the problem identified by Felsenstein, (1975) on the spatial continuum models proposed by Wright, (1943) and Malécot, (1948), namely 'the pain in the torus'. The *spatial Λ-Fleming-Viot process* will not be further discussed in this PhD thesis, but I will spend few words to give a broad overview on what it consists and

why it has been considered a promising model to solve the limitations encountered with the most known spatial models. As mentioned above, the spatial models of Wright, (1943) and Malécot, (1948) assume that, as the initial parental populations, also the offspring are distributed in space according to a Poisson random field. This assumption turned out to not be accurate (Felsenstein, 1975). In the *spatial Λ-Fleming-Viot process*, this difficulty is overcome by defining reproduction not as an individual process (parent -> child) but instead as a random sequence of events which prescribe the spatial region in which reproduction takes place (Fig. 6c). In particular, without entering into further details, it is considered that individuals living within a region A have a probability *u* to die and be replaced by a number of offspring given by a Poisson distribution with mean ρ *u* $|A|$, where ρ is the population density and $|A|$ is the surface area in the region A (Barton and Etheridge, 2019). Several variants of the *spatial Λ-Fleming-Viot process* have been extended for incorporating other processes such as selection (Etheridge et al., 2020), recombination (Barton et al., 2013) and large-scale extinction-recolonisation events (Barton et al., 2010b).

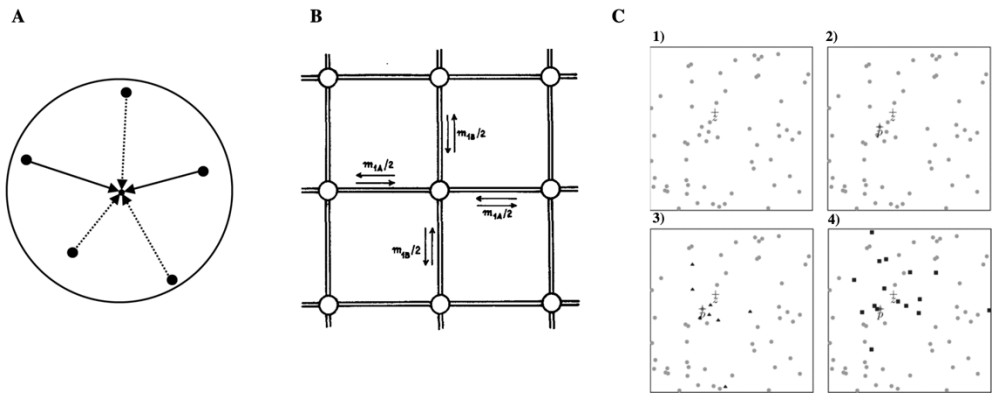


Figure 6. Models in spatial population genetics. A) Wright's neighborhood in a spatial continuum model: large dots represent neighbors, the solid arrows actual mating between them, the dotted arrows potential mating, and the smaller dot a zygote. The lengths of the arrows represent dispersal distances (source: Ishida, 2009); B) Two-dimensional stepping-stone model: $m_{1A}/2$ or $m_{1B}/2$ is the migration rate between adjacent demes in direction A or

B, respectively (source: Kimura and Weiss, 1964); C) spatial Lambda-Fleming-Viot model under the extreme case of one single parent generating all the offspring around the extinction/recolonization event in z, forward in time: a two-dimensional population is composed by individuals uniformly distributed according to a Poisson random process 1) centre of the event is chosen at z, uniformly at random, 2) the parent p at position x is randomly chosen with weight $v(z,x)$, 3) individuals die (triangles) with probability $u(z,w)$, where w is the location of the individual; 4) the number of offspring (squares) are sampled from a Poisson distribution with density $p\,u(z,w)|A|$ are scattered around z (source: Barton et al., 2010).

1.5 Theoretical studies on genetic response to HL&F

The previous overview on spatial models in population genetics show that spatial genetic patterns can be interpreted mainly in terms of effective population density or size (D_e or N_e), effective dispersal rate (σ^2) which corresponds to the mean squared axial parent-offspring distance, and balance between mutation and dispersal ($\sigma/\sqrt{2\mu}$). This is particularly relevant if we consider that geographically-limited dispersal (or isolation-by-distance) is common in natural populations, and thus the theory of spatial population genetics developed in the last decades can help us to better understand, and possibly predict, the consequences of HL&F across taxa.

There is a relatively large body of theoretical studies investigating spatial genetic pattern in the context of HL&F. Some of these studies were not explicitly asking how HL&F influences genetic response, but their findings have contributed to a better understanding and predictability of population genetic patterns.

Following the definition of HL&F as a two spatial-scale process (*landscape-scale* and *patch-scale*; see Section 1.2; Fletcher et al., 2018; Fahrig, 2019), we could perhaps equally distinguish theoretical populations genetic studies at the *patch-scale* from those at the *landscape-scale*. For sake of clarity, I remind the reader that among *patch-scale* processes, we should consider: patch-size, local habitat quality, edge-effect and patch-isolation. Whereas, at the *landscape-scale*, the HL&F literature typically refer to two processes: habitat amount and habitat fragmentation per se (roughly, the number of habitat patches in a landscape). The following two sections will provide a general overview on the current state of the art on the influence of *patch-scale* and *landscape-scale* processes on genetic variation.

1.5.1 Patch-scale process

Maruyama, (1972), for instance, studied the rate of decay of genetic variation over time in a finite two-dimensional population (2D torus-like continuous model, equivalent to an isolated and finite habitat patch), showing that the rate of decay is proportional to dispersal rate, namely that populations with lower dispersal rate would experience slower loss in genetic variability than populations with higher dispersal rate. In particular, Maruyama, (1972) found in *Eq. 4-2* and *Eq. 4-3* that the rate of decay in genetic variability ($1 - \lambda_t$) between time $t + 1$ and t is approximately equal to:

$$\frac{D\sigma^2}{2N} \text{ if } D\sigma^2 < 1 \qquad \text{and} \qquad \frac{1}{2N} \text{ if } D\sigma^2 > 1, \qquad (\textit{Eq. 10})$$

where D is population density, σ^2 is the variance in dispersal distance and N is the total number of diploid individuals in the whole population. In Maruyama, (1972), λ_t is the dominant eigenvalue of the decay in genetic variability $H(t)$ across the entire population, which is defined as $H(t+1)/H(t)$, where $H(t)$ refers to the average probability that two genes randomly chosen from the entire population are of different type at generation t.

In other words, Maruyama, (1972) provides predictions on how 'patch-size' (N), 'local habitat quality' (D) and species dispersal rate (σ^2) influence genetic diversity. Under the reasonable assumption that marginal subpopulations (or edge subpopulations) experience lower dispersal rate, given that movement of individuals is impaired at the habitat boundaries, the results of Maruyama (1972) point out also to an expected lower genetic diversity in edge than central subpopulations. In fact, it is shown from *Eq. 2-6* and *Eq. 4-2* in Maruyama (1972) that, under the condition $D\sigma^2 < 1$:

$$1 - \lambda_t = \frac{\|H_0(t)\|}{2N\|H(t)\|} \qquad (\textit{Eq. 11})$$

where $H_0(t)$ is the frequency of heterozygotes (or observed heterozygosity). Therefore, observed heterozygosity at time t can be expressed as:

$$H_0(t) = D\sigma^2 H(t) \qquad (Eq.\ 12)$$

which decreases with decreasing values of dispersal rate. A later study by Tajima, (1990) used other types of structured model, stepping-stone (spatially-explicit) and island (non-spatial) models, to examine the expected level of DNA polymorphism within subpopulations in relation to dispersal rate. In agreement with Maruyama (1972), Tajima (1990) showed that the amount of DNA polymorphism was lower at smaller dispersal rate (marginal subpopulations) than at higher dispersal rate (central subpopulations). Note that, in these two studies, quantitative predictions on the lower genetic variability of edge subpopulations were deduced from more general results on the influence of total population size, subpopulation density and dispersal rate on genetic diversity. Following studies were explicitly asking how genetic variability would change within the same population, if genes were sampled from subpopulations at the boundaries (edge) or at the centre of a two-dimensional habitat (Herbots, 1994; Hey, 1991; Wilkins and Wakeley, 2002; Wilkins, 2004). For instance, Wilkins, (2004) derived approximations for the distribution of coalescence times for a pair of genes sampled at any location in a rectangular habitat (*Eq. A3* and *Eq. A4*). In particular, it is shown that the probability density function (PDF) of coalescing at time t for a pair of genes sampled from the same location ($r = 0$) could be expressed as:

$$f(r = 0, t) = \sum_{j=1}^{t} \frac{(-1)^{j-1}}{Nbh^j} s_j\,, \qquad (Eq.\ 13)$$

with

$$s_j \approx s_{j-1} + \frac{j+1}{t} e^{-\sum_{k=1}^{j-3} k^2/t} Log^j(t + 3/2 - j) \qquad (Eq.\ 14)$$

and $Nbh = 4D_e\pi\sigma^2$ (neighbourhood size). Whereas, the PDF of coalescing at time t for a pair of genes sampled at a geographical distance $r > 0$ would be equivalent to (*Eq. A17*):

$$f(r,t) = \frac{e^{-r^2/t}}{Nbh}\frac{1}{t}\left(1 - \frac{1}{Nbh}\Gamma\left(0, \frac{r^2}{t-\gamma^*} - \frac{r^2}{t}, \frac{r^2}{\gamma^*} - \frac{r^2}{t}\right)\right) \qquad (Eq.\ 15)$$

where $\gamma^*$is 0.562 and $\Gamma(0, z_1, z_2)$ is the incomplete gamma function, defined as:

$$\Gamma(0, z_1, z_2) = \int_{z_1}^{z_2} \frac{e^{-\zeta}}{\zeta} d\zeta \qquad (Eq.\ 16)$$

Wilkins, (2004) uses these expressions in an approach analogous to the 'method of images' to approximate the effect of edges (habitat boundaries) on the distribution of coalescence times. It is considered that for a pair of genes sampled in the same subpopulation at a distance z from a reflecting boundary, the coalescence probability is described by two coalescence processes, one for two genes sampled in the same subpopulation and one for two genes sampled at distance $2z$. Further extensions and cases are discussed in Wilkins, (2004).

Using the expressions on the probability of coalescence and probability of identity-by-descent derived by Wilkins, (2004), one could then estimate the expected level of genetic diversity, within ($r = 0$) or between ($r > 0$) subpopulations:

$$H = \left(1 - \int_0^\infty f(r,t)\, dt\right)\bar{H} \qquad \textit{(Eq. 17)}$$

where $\bar{H}$ is the average heterozygosity across the whole population, and $\int_0^\infty f(r,t)\, dt$ is the cumulative density function (CDF) for all the possible coalescing events for two random genes sampled at distance r. Although, as noted by Wilkinson (2004), these expressions are relatively simple under a rectangular 2D habitat, they become computationally expensive when habitats have complex geometries.

Recently, Ringbauer et al., (2018) have introduced a framework that model the effect of a linear barrier in a two-dimensional continuous habitat on the distribution of coalescence times, extending previous theoretical results by Nagylaki, (1988) and Barton, (2008). As shown in relation to Wilkins, (2004) results, the probability of identity-by-descent can be derived from the distribution of coalescence times, which can then be used to estimate pairwise homozygosity among any pair of individuals ($r = 0$ or $r > 0$) in the habitat (both on the same side or different sides of the barrier). In particular, based on the analytical solution of Grebenkov et al., (2014), Ringbauer et al., (2018) showed in *Eq. 5* that the probability of two genes at position $p_1 = (x_0, y_0)$ and $p_2 = (x, y)$ of coalescing at time t in the past, assuming $x_0 > 0$ (i.e., on the right side of the barrier) can be expressed as:

$$\psi(p_1, p_2, t) \approx \begin{cases} G_{ss}(x, x_0, 2t)\, G(y, y_0, 2t)\ 1/2D_e \ \ if\ x > 0; \\ G_{ds}(x, x_0, 2t) G(y, y_0, 2t)\ 1/2D_e \ \ if\ x < 0; \end{cases} \qquad \textit{(Eq. 18)}$$

where $G(y, y_0, 2t)$ is the probability density function of moving from y_0 to y in time t in the absence of a barrier, $G_{ss}(x, x_0, 2t)$ and $G_{ds}(x, x_0, 2t)$ are the probability density functions of moving from x_0 to x in time t, if x_0 and x are on the same- or different- sides of the barrier, respectively. The movement probabilities are expressed as:

$$G_0(y, y_0, t) = \frac{1}{\sqrt{2\pi\sigma^2 t}}\, e^{-\frac{(y-y_0)^2}{2\sigma^2 t}} \qquad \textit{(Eq. 19)}$$

$$G_{ss}(x, x_0, t) = \frac{e^{-\frac{(x-x_0)^2}{2\sigma^2 t}} + e^{-\frac{(x+x_0)^2}{2\sigma^2 t}}}{\sqrt{2\pi\sigma^2 t}} - \frac{2K}{\sigma^2} e^{\left(\frac{4K(x_0+x+2Kt)}{\sigma^2}\right)} erfc\left(\frac{x_0 + x + 4Kt}{\sqrt{2\sigma^2 t}}\right) \qquad \textit{(Eq. 20)}$$

$$G_{ds}(x, x_0, t) = \frac{2K}{\sigma^2}\, e^{\left(\frac{4K(x_0 - x + 2Kt)}{\sigma^2}\right)} erfc\left(\frac{x_0 - x + 4Kt}{\sqrt{2\sigma^2 t}}\right) \qquad \textit{(Eq. 21)}$$

Then, since the probability of identity-by-descent $F(p_1, p_2)$ can be expressed as:

$$F(p_1, p_2) = \int_0^\infty \psi(p_1, p_2, t) \exp(-2\mu t)\, dt \qquad \textit{(Eq. 22)}$$

the expected level of pairwise homozygosity $\mathbb{E}\left(h_{p_1,p_2}\right)$ can be derived as:

$$\mathbb{E}\left(h_{p_1,p_2}\right) = F(p_1, p_2) + \left(1 - F(p_1, p_2)\right)(1 - \bar{H}) \qquad \textit{(Eq. 23)}$$

where $\bar{H}$ is the average heterozygosity across the whole two-dimensional habitat. In a few words, these expressions allow to accurately estimate the impact of soft or hard-edges (barrier) on expected genetic diversity.

Leblois et al., (2006) performed spatially-explicit simulations to study the effect of habitat contraction on genetic diversity. Their results were confirming the predictions of Maruyama (1972), namely that continuous 2D populations undergoing contraction were decreasing genetic diversity slower

(rate = $D\sigma^2/2N$ if $D\sigma^2 < 1$; Maruyama, 1972) than a Wright-Fisher (panmictic) population of equivalent size ($1/2N$ if $D\sigma^2 > 1$; Maruyama, 1972). In addition, they observed that under low dispersal rate the genetic diversity (A_r and H_e) estimated from local samples were roughly independent from the total population size, whereas samples collected across the whole habitat would restore a positive correlation between genetic diversity and total population size (≈ patch-size). Moreover, they also noted that samples collected across the whole habitat (scaled-sampling) were showing lower H_o than samples collected exclusively from the centre (local-sampling). They interpreted this observation as the result of the lower H_o in individuals sampled at the boundaries of the simulated scenario, which would then decrease the average H_o for the scaled-sampling.

Statistical approaches combining information on 'local habitat quality' and 'patch-connectivity' (or 'patch-isolation') have been developed to understand how these two variables influence within-patch genetic diversity (see Wagner and Fortin, 2013, 2015) for a review). These statistical approaches can be classified into i) *neighbourhood-level* analysis, which relates the genetic diversity of a focal patch to its connectivity with neighbouring patches, depending on the geographical distance from the focal patch and on the dispersal rate of the focal species (Keyghobadi et al., 2005; Balkenhol et al., 2009; James et al., 2011); and ii) *node-based* analysis, which relates local landscape factors (or estimate of local population size) at the focal patch to genetic variation (Wagner and Fortin, 2013).

An application of *neighbourhood-level* analysis can be found in Keyghobadi et al., (2005), where the authors investigated whether the within-habitat patches genetic diversity (H_e) of alpine meadow-dwelling butterfly (*Parnassius smintheus*) in the Kananaskis region of Alberta (Canada) was correlated to the contemporary level of connectivity of each sampled patch. Based on connectivity measures used in metapopulation ecology (Hanski, 1994), they proposed a modified connectivity measure:

$$C_i = \sum_{j \neq i} \exp\left(-\alpha_f d_{f_{ij}} - \alpha_m d_{m_{ij}}\right) A_j^{(\xi_p - \xi_{em})} \qquad \textit{(Eq. 24)}$$

where $d_{m_{ij}}$ and $d_{f_{ij}}$ are the distances between patches i and j through habitat (meadow) and non-habitat (forest), respectively; α_m and α_f are the rate of movements in the two types of vegetations; A_j is the size of patch j; ξ_p is a parameter that scales local population size in a patch to the patch-size itself, and ξ_{em} scales the probability of emigration to patch-size. The summation indicates that, for each patch, the connectivity index considers the contribution of all patches in the study area. Their results showed significant correlation i) between within-patch expected heterozygosity (H_e) and patch connectivity (Fig. 7a), and ii) between pairwise genetic differentiation (G_{st}) and pairwise patch connectivity (Fig. 7b). In agreement with previous theoretical studies showing that genetic diversity changes slower over time than genetic differentiation (Varvio et al., 1986), they also found that within-patch H_e and between-patches G_{st} were better correlating with connectivity measures estimated from past (1952) and present-day (1993) landscape structure, respectively.

Applications of node-based analysis include, for instance, the study of Taylor and Hoffman, (2014). They investigated genetic diversity pattern in the white-footed mouse (*Peromyscus leucopus*), sampled along a transect from southern Ohio to northern Michigan (USA), and assessed which landscape features would explain the observed spatial distribution in genetic diversity. In particular, landscape features, such as the area of forest habitat coverage, were measured within a radius of either 0.5, 5, or 50 km from the sampling site. They found that allelic richness (A_r) and H_e were correlated with several measures of habitat availability at a scale of 0.5 km from the focal sampling site ($r^2_{H_e} = 0.593$; $r^2_{A_r} = 0.698$) (Fig. 7c). The correlation between genetic diversity and habitat quality was disappearing when using larger spatial

scales (5 km or 50 km), pointing out to the importance of identifying the correct spatial scale when exploring the effect of habitat availability on genetic diversity.

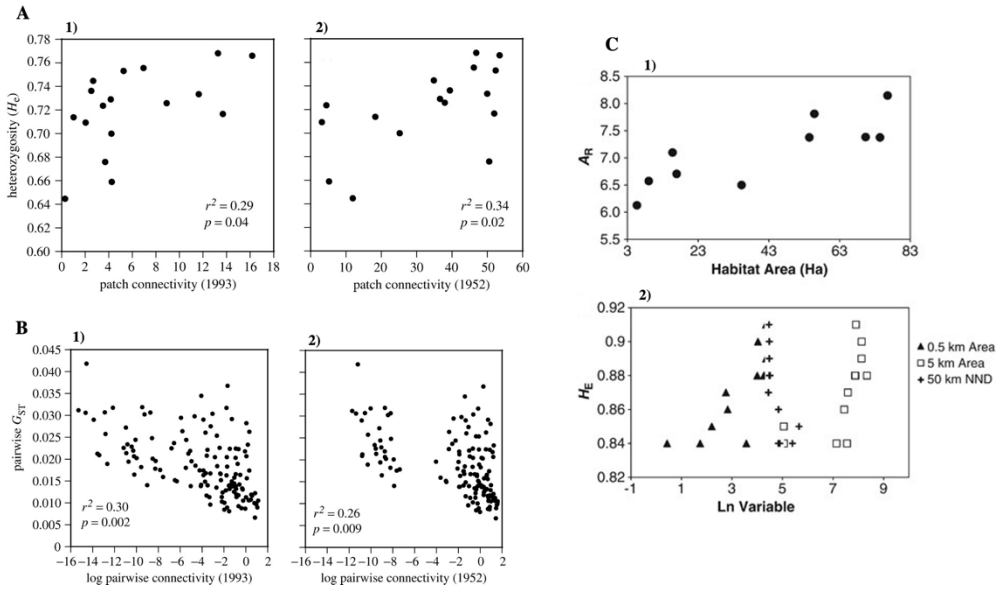


Figure 7. Examples of statistical approaches quantifying the effect of patch-scale processes on local genetic diversity. Correlation of A) H_e within habitat patches and patch connectivity, or B) G_{ST} between habitat patches pairs and pairwise patch connectivity in *Parnassius smintheus* for connectivity measured from 1) 1993 and 2) 1953 landscape habitat cover data (modified from Keyghobady et al., 2005); C) Correlation between 1) A_r and total habitat area calculated over a radius of 0.5 Km, and 2) H_e and total habitat area (variable) calculated within a radius of 0.5 Km, 5 Km and 50 Km, in *Peromyscus leucopus* (modified from Taylor and Hoffman, 2014).

1.5.2 Landscape-scale process

Simulation-based studies have been carried out to assess the relative impact of 'habitat loss' and 'habitat fragmentation *per se*' on *landscape-scale* genetic pattern (Bruggeman et al., 2010; Graves et al., 2012; Cushman et al., 2012; Mona et al., 2014; Jackson and Fahrig, 2016). For instance, Jackson

and Fahrig, (2016) investigated, through individual-based spatially-explicit simulations of landscapes with different degree of HL&F, the relative contribution of habitat amount and habitat fragmentation *per se* on population genetic structure. They simulated 25 landscapes (101 x 101 cell grid) that vary independently in habitat amount (proportion of covered area) and habitat fragmentation (clumpiness), considering 'virtual' species varying in dispersal distance (average dispersal distance = 1 – 6 cell steps) and behaviour (random *versus* informed dispersal), while fixing population density at 2 individuals per cell. For each landscape configuration, they ran 100 simulations for 1,000 generations and sampled 200 individuals across continuous areas at the end of the simulation. They measured the relative influence of habitat amount and habitat fragmentation on isolation-by-distance (IBD) (Mantel r test of individual geographic *versus* genetic distance) using the percent sum of squares (%SS), which gives the proportion of variation in the data explained by a variable (habitat amount, habitat fragmentation or habitat amount X habitat fragmentation). They found that habitat amount is a much better predictor of genetic structure than habitat fragmentation *per se*, the former measured as proportion of habitat and the latter measured as the degree to which the focal habitat is aggregated or clumped given its total area (average %SS for habitat amount = 81.1%; average %SS for habitat fragmentation = 4.9%). They suggest that habitat amount may often be the best metric for predicting population genetic structure. However, they also noticed that habitat fragmentation *per se* does have an influence when habitat amount is scarce (below 40% of the total number of cells) and dispersal distances are small (D < 2 steps). A similar study was carried out years earlier by Cushman et al., (2012). As Jackson and Farhig, (2016), Cushman et al., (2012) simulated 625 landscapes, varying in habitat amount (proportion of covered area) and fragmentation (clumpiness), sampled genetic data (10 loci) from 500 randomly sampled individuals at the end of 500 generations and measured IBD computing Mantel r test of individual genetic *versus* landscape effective distances.

Cushman et al., (2012) reached opposite conclusions, namely that IBD is mainly shaped by habitat fragmentation *per se* and less by habitat amount. Jackson and Fahrig, (2016) imputed the discordance between the two studies to the fact that Cushman et al., (2012) have used fragmentation metrics that are strongly correlated with habitat amount, therefore confounding the conclusions of the study.

Jackson and Fahrig, (2016) and Cushman et al., (2012) are two examples where the genetic consequences of HL&F have been investigated by relating genetic response with two types of independent variable: habitat amount and clumpiness. In parallel, over the last two decades new approaches have been developed to identify environmental features (biotic and abiotic) that influence dispersal, and may thus represent either barriers to gene flow or corridors at the scale of the landscape (Hendry, 2004; Nosil et al., 2008; Sexton et al., 2014). For instance, Cushman et al., (2006) proposed a methodology defined as “Causal Modelling” (CM). The idea is to build a resistance map based on landscape features and identify the least-cost path connecting every pair of individuals in the genetic dataset. Genetic differentiation and least-cost path distances are then correlated across all samples. The features showing significant relationship with genetic distance are then considered to be those having higher impact on dispersal. CM analysis considers the least-cost path between sample pairs, and by doing so it ignores the multiple dispersal pathways that individuals may have travelled over time. McRae, (2006) proposed a solution to the problem of multiple dispersal pathways by developing a model called “Isolation-By-Resistance” that incorporates all possible paths and improves predictions of genetic differentiation. It also allows to estimate a surface of connectivity among samples (called “current map”), determining the areas that facilitate or restrict gene flow (Anantharaman et al., 2020). In these two approaches is however impossible to disentangle the effect of geographically-limited dispersal from the effect of environmental features on gene flow. Thus, some authors have

gone further and proposed the concept of “Isolation-By-Environment or Isolation-By-Ecology”, allowing to assess the relative contribution of both an environmental variable and the geographical distance in limiting species gene flow (Wang and Summers, 2010; Wang, 2012; Bradburd et al., 2013). One of the advantages of these approaches is that they go beyond the simplistic (and human-centred) distinction of habitat versus non-habitat by providing a framework where to test hypothesis on which landscape features promote (or impair) gene flow in a given species.

“Isolation-By-Resistance” and “Isolation-By-Environment or By-Ecology” models, in the way in which they are currently implemented, assume that genetic patterns have reached ‘equilibrium’, meaning that they will not change over time and that they reflect the effect of present landscape features on gene-flow. Yet, present day genetic pattern may actually be significantly influenced by past landscape structure, as Keyghobadi et al., (2005) have shown (see Section 1.5.1). This is because it may take some time (lag-time) for a measure of genetic variation to show significant changes after a given demographic event (Landguth et al., 2010; Mona et al., 2014; Epps and Keyghobadi, 2015). For example, Landguth et al., (2010) showed that F_{st} between populations separated by a barrier can take few hundred generations since the appearance of a barrier before reaching the new equilibrium value. Similarly, Mona et al., (2014) simulated a two-dimensional grid of demes (50 x 50) subdivided in 100 habitat patches (5 x 5), among which gene flow was impaired (habitat fragmentation), and found that, in their simulated scenario, habitat fragmentation required over 100 generations to significantly influence patterns of population structure (F_{st}). These and other studies point out to the importance of considering space and time in population genetic models in order to disentangle the relative effect of time, geographically-limited dispersal and landscape features in shaping genetic pattern (Bradburd and Ralph, 2019). Recently, some studies have proposed new ways of retrieving fine-scale information from genomic data, identifying

how dispersal rate and population size vary in space and in recent times (e.g., Ralph and Coop, 2013; Ringbauer et al., 2017; Al-Asadi et al., 2019). For instance, Ringbauer et al. (2017), following Ralph and Coop, (2013), developed a method to accurately infer both D_e and σ, and recent changes in D_e in a two-dimensional population (spatial continuum with local-density regulation; see Barton et al., 2002). This approach uses the distribution of the length of tracts in the genome that are shared between pairs of individuals (identity-by-descent blocks) with known geographical location. Al-Asadi et al., (2019) extended the framework of Petkova et al., (2016), which estimates variation in effective dispersal from real data by modelling the study area as a two-dimensional stepping-stone model, and presented a method that infer changes in population size and dispersal rate across different time periods. However, these approaches often require either dense genome-wide data or whole-genome sequences (together with a linkage map), which are still difficult to obtain for most non-model organisms.

1.6 Madagascar: the impact of ancient and recent HL&F on biodiversity

1.6.1 Madagascar Natural History

Madagascar is an Indian Ocean island (~592 km^2) between the tropics of Cancer and Capricorn, located at c.a. 400 – 500 km from the African mainland. Based on its proximity to African mainland, it may be natural to assume that Madagascar detached at last from African mainland. However, data suggest otherwise. Between 160 and 180 Mya, the southern supercontinent of Gondwana divided into a western part, including Africa and South-America, and an eastern part composed of Madagascar, the

Seychelles, the Indian subcontinent, Antarctica, and Australia (Ali and Aitchison, 2008; Eagles and König, 2008). Only about 80-90 Mya, Madagascar separated from India and remained at its current location since then (Ali and Krause, 2011). Although being isolated from any other landmasses since at least 80 Myr, the origin of present-day Malagasy mammals, freshwater fish, amphibians and reptiles date back to after the separation from Africa and India (Simpson, 1940; Yoder et al., 1996; Vences et al., 2001, 2004; Yoder et al., 2003; Poux et al., 2005). For instance, current estimates of colonization time suggest that mammals might have arrived in Madagascar from Africa through at least two colonization events: for lemurs between 60 and 55 Mya (Herrera and Dávalos, 2016), and for carnivorans and rodents between 27 and 19 Mya (Yoder et al., 2003; Poux et al., 2005; Steppan and Schenk, 2017; Zhou et al., 2017). Colonization time estimates of tenrecs have a wide confidence interval suggesting that they may have arrived in Madagascar at the same period of either lemurs or carnivora/rodents (Douady et al., 2002; Poux et al., 2005, 2008; Everson et al., 2016). Several hypotheses have been proposed to explain the 'discordance' between the time of Madagascar isolation and the colonisation time of Malagasy mammal (McCall, 1997; Ali and Huber, 2010; Masters et al., 2021 for a synthesis). In a recent study, Masters et al., (2021) have integrated phylogenetic, ecophysiological, oceonography and paleontology data with new data from the PAMELA (Passive Margins Exploration Laboratories), based in the Mozambique Channel, to review hypotheses about the arrival of terrestrial vertebrates in Madagascar. The common narrative explaining the arrival of terrestrial vertebrates in Madagascar is represented by the hypothesis of transoceanic dispersal, either via rafting of floating vegetation or swimming. Additional *ad hoc* hypotheses were proposed to make the transoceanic dispersal scenario more realistic, for instance, that hibernation would have helped mammals to reach Madagascar, or that eastward Paleogene surface currents would have facilitated the floating rafts to reach Madagascar in 3-4 weeks. In conclusion, they proposed and analysed an

alternative hypothesis based on a scenario of geodispersal via short-lived land bridges between Africa and Madagascar. The geological data point to the existence of three short-lived land bridges: i) 66-60 Mya; ii) 36-30 Mya; and iii) 12-05 Mya.

Madagascar is known for its extraordinary biological diversity, which include approximately 80–90% of endemic species (Goodman and Benstead, 2005; Thompson et al., 2011; Ganzhorn et al., 2014). The exceptional biological diversity of Madagascar has been broadly explained by the nearly complete isolation of the island, since at least 80 Myr, and the wide diversity of climate and ecosystems (Dewar and Richard, 2007; Moat and Smith, 2007). The climatic diversity of Madagascar can be explained by i) the presence of three mountainous regions above 2,000 m of altitude, dividing the east and western part of the island; ii) its geographical location in the Indian Ocean; iii) the monsoon precipitation modulated by the Inter-Tropical Convergence Zone (ITCZ); and iv) the changes in monsoon strength (Douglass and Zinke, 2015). Phylogeographic analyses have been carried out on several Malagasy taxa in order to understand the origins and factors shaping the distribution of the exceptional biological diversity of Madagascar. Vences et al., (2009) has summarized the main diversification mechanisms in five models: i) 'eco-geographic constraint' (Martin, 1972); ii) 'watershed' (Wilmé et al., 2006); (iii) 'western rainforest refugia', an example of the general 'rainforest refugia' (Moritz et al., 2000); iv) 'riverine barrier' (Ganzhorn et al., 2006); and v) 'montane refugia' e.g., (Wollenberg et al., 2008).

Quaternary climatic oscillations have been proposed as mechanisms shaping the geographic distributions for a part of Malagasy taxa in the last several hundred thousand years (Wilmé et al., 2006, 2012; Mercier and Wilmé, 2013). The climatic oscillations consisted in phases of dry and cool climate that would alternate with wetter and hotter periods. Accordingly, forest ecosystems would also change, in extent and altitude, thus influencing the

distribution of forest-dwelling and non-forest- dwelling species. This is particularly relevant if we consider that most Malagasy endemic species (e.g., ~90% land mammals, ~78 % for endemic bird species) are strictly associated to forested habitats (Raherilalao and Wilmé, 2008). Following the model proposed by Wilmé et al., (2006), dry and cool periods would have reduced forests extent, resulting in a forest cover limited to areas where water was still persisting (along rivers or at the headwaters of rivers, or mountainous massifs). These dry and cool periods were particularly challenging for many forest-dwelling species, which in some cases got extinct, and in other cases become isolated in refugia (e.g., riparian forests along rivers and/or mountainous forests). Conversely, wetter and hotter periods would have promoted the expansion of forested habitats and associated species.

The impact of past climatic oscillations on Malagasy biomes is still debated. The extensive records of subfossils recovered from Madagascar suggest that Malagasy biodiversity was larger in the past than what it is today (Crowley, 2010). Subfossil data show also that, in the case of lemurs, large-body mass species might have been more prone to extinction, since the body-mass of the 17 described extinct subfossil lemur taxa ranged between ~11 kg to ~160 kg (Karanth et al., 2005), whereas the largest present-day lemur species have an average weight of 6.8 kg (Powzyk and Mowry, 2003). The timing of the Malagasy 'megafaunal extinction' (~1,000 years ago; Crowley, 2010) appears to coincide with the timing of human arrival in Madagascar (~2,000 years ago; Douglass et al., 2019), which has led several authors to hypothesize that humans were the major cause of this biodiversity loss (Burney et al., 1997; MacPhee and Burney, 1991; Crowley et al., 2010; Crowley, 2010; Virah-Sawmy et al., 2010). Other studies, instead, have proposed a joint effect of humans and natural climate change on the extinction of most Malagasy megafauna (Burney and MacPhee, 1988; MacPhee and Burney, 1991; Burney, 1993; Godfrey and Jungers, 2003; Burns et al., 2016; Crowley et al., 2017; Hixon et al., 2021b, 2021a). According to this

hypothesis, past natural climate drought would have led to the aridification of Madagascar ecosystems, driving, concomitantly with human activities (i.e., forest clearance, vegetation burning, hunting and the introduction of potential megaherbivore competitors) the extinction of the Malagasy megafauna.

Natural climate change and human activities are also central in the debate about the origin of grasslands in Madagascar (Irwin et al., 2010; Virah-Sawmy et al., 2010), which currently cover ~70% of the island's surface (e.g., Moat and Smith, 2007; Thompson et al., 2011; MEF, 2014). Earlier studies hypothesized that Madagascar was entirely covered by continuous forest ("forest hypothesis"), until the arrival of humans who promoted the expansion of grasslands by introducing fire and zebu cattle (Perrier de la Bâthie, 1921; Humbert, 1927; Lavauden, 1931; Koechlin, 1972). This hypothesis was mainly supported by the apparent lack of biodiversity and endemic species in the grassland ecosystems compared to the forested habitat (Koechlin, 1972; Lowry et al., 1997; Willis et al., 2008). The alternative hypothesis states that grasslands existed before human arrival ("grassland hypothesis"), originated as a response to cold and dry climates, and that human contributed only partially to the expansion of grasslands (Burney, 1987a, 1987b; Burney et al., 1997; Willis et al., 2008; Bond et al., 2008). In support of such hypothesis, Bond et al., (2008) showed that Malagasy grasslands have ~87% of the C4 grasses found in southern Africa, with a wide range of niche specialization. Vorontsova et al., (2016) found that, among the 145 grass species sampled across 60 sites in Madagascar, ~48% are endemic to the island and ~29% are narrow endemic species (within ecoregions). Fisher and Robertson, (2002) compared ant species richness in adjacent montane forest and grassland in Central Madagascar. They found more species in grassland (N = 31) than in forest (N = 26) sampling sites, and that only 5 species were shared between the two habitats, suggesting a high rate of ant species specialised to grassland. In support of this hypothesis, some palaeontologists have remarked that many of the Malagasy extinct animals appeared to be

adapted to savannah than to a forested habitat (MacPhee et al., 1985). In few words, these studies showed that grasses in Madagascar have a natural origin and a likely long local evolutionary history.

1.6.2 An overview of genetic studies of Malagasy mammals and HL&F

Most Malagasy endemic species (e.g., ~90% land mammals, ~78 % for endemic bird species) are strictly associated to forested habitats (Raherilalao and Wilmé, 2008). Genetic data from forest-dwelling species could thus be used to infer past demographic changes, and their reconstructed history could serve as a proxy to assess the past changes in forest cover. Along with this approach, it would be ideal then to compare dates of demographic changes in forest-dwelling species with other (and independent) types of data on the paleo-environmental changes that have occurred in Madagascar (e.g., paleorecords). As I will detail in this section, genetic studies have been carried out on Malagasy mammals: i) to investigate the influence of present-day HL&F on species abundance and dispersal; and ii) to reconstruct past forest cover (HL&F) and how species responded to it.

1.6.2.1 Influence of HL&F on species abundance and dispersal

A large proportion of genetic studies in Madagascar were interested in identifying the potential effects of HL&F on species genetic diversity in order to pinpoint populations that could possibly need to be protected (e.g., Olivieri et al., 2008; Ramanamahefa et al., 2009a, 2009b; Quéméré et al., 2010b). Earlier studies were mostly comparing genetic pattern across areas with different degree of HL&F. The first of such studies is Olivieri et al., (2008), which investigated the influence of geography and HL&F on the genetic diversity of three mouse lemurs (*Microcebus*

ravelobensis, *M*. *bongolavensis*, and *M*. *danfossi*) from north-western Madagascar. *M*. *ravelobensis* was sampled from four sites within continuous forest (Ankarafantsika forest, about 1000 km^2) and from four sites within forest fragments (areas: 4-37 km^2). Analyses of hidden genetic structure identified two main clusters. The first mainly consisted of individuals from the continuous forest and the two closest forest fragments (~20 km and ~1 km from Ankarafantsika). The other cluster corresponded to the two remaining forest fragments located around 80 km from the continuous forest. In a further analysis, the authors also found that the first cluster could be further subdivided into three sub-clusters corresponding each to the continuous forest, and the other two forest fragments. Finally, they also identified an isolation by distance pattern across fragmented and continuous populations, hence suggesting a recent fragmentation of forest habitats in this region. *Microcebus bongolavensis* was sampled in three forest fragments, with one location >50 km away from the other two. Despite the large geographic distance, the absence of a continuous forest and the lack of connectivity among locations, only one genetic cluster could be identified. In *M*. *danfossi*, individuals were sampled across seven forest fragments (1.5 - 37 km^2). The authors found higher levels of genetic differentiation in this species than in *M*. *ravelobensis* but no correlation between genetic and geographic distances. Despite the similarly large geographic distances between some *M*. *danfossi* samples compared to *M*. *ravelobensis* and the lack of a large continuous forest habitat, *M*. *danfossi* individuals grouped mainly in one genetic cluster with the exception of one forest fragment that formed a second cluster.

Other genetic studies shifted from the comparisons of fragmented *versus* continuous areas to the quantification of the relative importance of various landscape features on species genetic diversity. Quéméré et al., (2010a) used Causal Modelling, in addition to other methods, to study the effect of geographical distance, roads, river, and forest canopy on the dispersal of the medium-size body lemur *Propithecus tattersalli* in the fragmented Loky-Manambato region. They identified the Manankolana River

crossing the region as the main barrier to dispersal in this species. In addition, a significant effect of geographical distance was detected, implicitly suggesting that the degree of habitat fragmentation in the region might not disrupt gene flow significantly in this species. A later study by Aleixo-Pais et al., (2019) on the northern rufous mouse lemur (*Microcebus tavaratra*) sampled in the same region, found that the Manankolana River was not the main factor correlated to genetic differentiation. They rather identified the two most humid and southern forests as clustering apart from all other samples. Three other studies based on mtDNA have been carried out in the Loky Manambato region, the first two on Nesomyinae rodents, *Eliurus carletoni* and *E. tanala* (now *E. ellermani*; Jansa et al., 2019) (Rakotoarisoa et al., 2013a, b), and the second on *M. tavaratra* (Sgarlata et al., 2018). Both studies found that the Manakolana River did not have such a strong effect on the genetic structure of these species. Hence, the results of landscape genetic studies carried out in the Loky-Manambato region on *Microcebus*, *Eliurus*, and *Propithecus* identified different effects of landscape configuration on genetic differentiation and diversity of these species. These results point out that i) response to HL&F can differ across species; and ii) that multi-species studies are necessary to identify major landscape features influencing genetic variation.

In the moist evergreen forests of eastern Madagascar, Baden et al., (2019) have used the 'Isolation-By-Resistance' approach, to identify areas that facilitate or restrict gene flow in the black-and-white ruffed lemur (*Varecia variegata*). They tested the effect of rivers, topography, roads, habitat cover, and proximity to human activity on genetic differentiation (measured by F_{st}) to identify the landscape variables that significantly affected *V. variegata* dispersal in a fragmented region (~64,000 km^2) of eastern Madagascar. The authors identified a significant effect of geographic distance on gene flow (explaining 28.2% of the genetic structure). However, they also found other more important effects in their model comparisons, in particular, habitat cover

and distance to nearest village had the highest explanatory power (~52%), as compared to habitat cover only (~25%), distance to nearest village only (~22%), and geographic distance only (~1%). Considering the high rate of deforestation in eastern Madagascar during the last ~60 years and the large effect of anthropogenic-related variables on the genetic differentiation of *V*. *variegata*, they proposed that rapid land cover modification by humans may have driven changes in the genetic structure of this species.

1.6.2.2 Reconstruct past-forest cover changes in Madagascar

Whereas the genetic studies mentioned in the previous section looked at the landscape features that affect 'present-day' species dispersal, it is becoming increasingly important, in Madagascar and elsewhere, to also understand past habitat connectivity as an approach to infer how species have responded to ancient HL&F.

In the **Central Highlands** of Madagascar, Yoder et al., (2016) used spatial genomic patterns of several *Microcebus* spp. to test three hypotheses on past landscape changes 1) "forest hypothesis", 90% of the island was covered by primary forest before human arrival; 2) "grassland hypothesis", grasslands in the Central Highlands were present long before human arrival; 3) "mosaic hypothesis", the Central Highlands were characterized by a mixture of wooded-grassland and closed-canopy forest. They studied five *Microcebus* spp. occurring in several habitat types across the island, namely, *M*. *mittermeieri* and *M*. *rufus* (moist evergreen forests of the east), *M*. *myoxinus* and *M*. *berthae* (dry deciduous forests of the west), and *M*. *lehilahytsara* (moist evergreen forests of the Central Highlands). They built a 'geogenetic' map for each studied species, by drawing arrows pointing in the direction of hypothesized admixture using the method of Bradburd et al., (2016). They found a stronger geogenetic relationship between *M*. *rufus* and *M*. *berthae*, occurring respectively in the moist evergreen forests of the east

and dry deciduous forests of the west, than between species of the same ecoregion (*M*. *myoxinus* and *M*. *berthae*). The surprising geogenetic proximity and the relatively recent divergence time between *M*. *rufus* and *M*. *berthae* (55.5 Ka, 95% CI: 4.56, 153 Ka) suggested that forest-dwelling species such as *Microcebus* could disperse across the Central Highlands throughout the Pleistocene until ~55 Ky ago. In addition, the geogenetic map of *M*. *lehilahytsara* showed that subpopulations of this species occupied two distinct geo-genetic clusters, interpreted as an indication that the Central Highlands might have experienced cyclic events of fragmentation and reconnection. Overall, their genomic data pointed towards a "mosaic hypothesis" of pre-human arrival vegetation in which Central Highlands were characterized by a mosaic of forests and wooded grassland. This is in agreement with other studies (i.e., Quéméré et al., 2012) that have increasingly shown that the "forest hypothesis" is unlikely to have been correct and that natural climate change have driven patterns of forest fragmentation even long before the arrival of humans (Burney, 1987a, 1987b; Vorontsova et al., 2016).

In **North-Western** Madagascar, another spatial approach was carried out by Schneider et al., (2010), who used spatially explicit simulations to study the demographic history of *Microcebus murinus*. They sequenced the mitochondrial *D-loop* (455 bp) of individuals sampled in both continuous and fragmented forests, and performed a series of demographic analyses through spatial simulations to reconstruct the history of the species. The results of the demographic analyses supported a model with two waves of spatial expansion, one before and one after the Last Glacial Maximum (LGM, ~25,000 years ago). The authors have also shown that correlation between genetic and geographic distance (Isolation-By-Distance, IBD) computed using fragmented populations separated by open habitat was significantly different from the one obtained using only populations sampled in continuous habitat. In particular, they found that, for the same geographical distance, genetic differentiation was significantly increased across fragmented

populations compared to continuous populations. They postulated that these fragments might have been connected in the recent past (hence explaining the IBD pattern) but that recent genetic drift had increased genetic differentiation (hence explaining the IBD differences between connected and fragmented populations). They thus noted that 1) genetic data might keep memory of a pre-fragmentation past; and that 2) differences in IBD might hold information on how and when fragmentation occurred.

In the same ecoregion, Olivieri et al., (2008) identified, dated, and quantified a major bottleneck in three *Microcebus* species, which suggested that the three species had a much larger population in the recent past, possibly as a result of significant HL&F. The best supported estimate suggested a major demographic event in the last 500 or 900 years. In a similar study carried out in the same region, Craul et al., (2009) also found a strong and recent bottleneck signal for *Lepilemur edwardsi*, which most likely started during the last 1000 years. Considering that the assumed value of species generation time can influence the dating of past demographic events, we should note that, even with small changes in the assumed values, the age of the major demographic event would still be reasonably recent, likely within the last 1000 to 3000 years.

In **Northern** Madagascar, genetic studies suggested a rather different scenario, with evidence of more ancient HL&F compared to the northwest. In a study carried out in Loky-Manambato region, Quéméré et al., (2012) analysed genetic data of the medium-size lemur *Propithecus tattersalli* using the method of Beaumont, (1999) and Storz et al., (2002), as it was done in Olivieri et al., (2008), and detected a bottleneck signal with a strong support for an ancient HL&F (likely pre-human), between 4,000 and 10,000 years ago. Additionally, by analysing satellite pictures, they found that forested habitats had decreased by only 2% in the previous 50 to 60 years in the Loky-Manambato region. In a more recent study, Salmona et al., (2017) extended

the work of Quéméré et al., (2012) by applying several methods of demographic inference to *P*. *tattersalli* and to *P*. *perrieri*, the latter occurring in the Andrafiamena massif, neighbour to Loky-Manambato region. The distribution of *P*. *perrieri* is currently smaller and more fragmented than that of *P*. *tattersalli*. Following the same methods as Quéméré et al., (2012), Salmona et al., (2017) found a major signal for a bottleneck that started at least 3000 years ago and possibly even before 10,000 years ago, depending on the generation time assumed. The results were thus consistent with Quéméré et al., (2012), strengthening the conclusions of the previous study. However, Salmona et al., (2017) were uncertain whether the timing was the same in the two species given the uncertainty of the estimated dates. Studies on other mammal species occurring in the Loky-Manambato region pointed out to a similar demographic history. For instance, Rakotoarisoa et al., (2013a) reconstructed the demographic history of *Eliurus carletoni*, distributed in the low-altitude forests of the Loky-Manambato region, using mtDNA (*cytb* and *D-loop*) sequences from 215 individuals. Their demographic analyses found a clear signal for an ancient stationary population that was subjected to a recent decline between 18,750 and 7,500 years ago, possibly as a consequence of Quaternary climatic changes that influenced habitat connectivity (HL&F). These results were confirmed in another study (Rakotoarisoa et al., 2013b) where the authors also studied *E*. *tanala* (now *E*. *ellermani*; Jansa et al., 2019), distributed in the same region but in humid forests of higher elevations, and found that *E*. *ellermani* exhibited a different history with no clear signal of population size change and perhaps a slight continuous increase. The demographic history recovered from *E*. *ellermani* and *E. carletoni* is consistent with paleoenvironmental reconstructions from the Central Highlands and northern Madagascar, both suggesting an increase in humid vegetation between ~15 and ~10 Kya (Gasse and Van Campo, 1998; Teixeira et al., 2021). These results thus appeared to confirm the idea that long-term habitat change during the Holocene led to an expansion of the

distribution of *E*. *ellermani* and a reduction and fragmentation of the distribution of *E*. *carletoni*.

The genetic studies based in **Northern** Madagascar concluded that the fragmented habitat of the Loky-Manambato was likely of natural and ancient origin, making this region rather different from other areas in the island, where deforestation has clearly been very high in the last decades (Vieilledent et al., 2018).

1.7 Conclusive Remarks

This introductory chapter aimed at showing that habitat loss and fragmentation (HL&F) is a heterogeneous process varying across various spatial and temporal scales, from the ancient climatic oscillations of the Late Quaternary to the recent increase in habitat degradation due to human activities. I introduced the concept of HL&F, its history and the underlying processes. Perhaps unexpectedly, there are still ongoing controversies around the impact of habitat fragmentation on biodiversity. I have presented some of the main findings on the impact of HL&F on ecology (e.g., species abundance and richness) and genetic pattern (e.g., genetic diversity and differentiation). Since geographically-limited dispersal is common in wild species, generating a spatial pattern of genetic diversity known as 'Isolation-By-Distance' (IBD), I briefly introduced concepts and models in spatial population genetic theory. These theoretical results have represented, in different occasions along this PhD project, the fundaments for better understanding and predicting the genetic consequences of HL&F in space and time. I concluded this introduction by providing a brief overview on the natural history of Madagascar, a biodiversity hotspot that, as many other places in the world, has undergone natural HL&F during hundreds of thousands of years and more recently has been severely threatened by anthropogenic HL&F. I have synthesised the current debate on the origin of

present-day HL&F in Madagascar, which see on two opposite sides supporters for either a natural or an anthropogenic origin of non-forested habitats (grasslands). Finally, I have summarised part of the genetic studies carried out on Malagasy mammals which i) investigate the effect of present-day landscape configuration on genetic variation, and/or ii) infer species demographic history as a proxy for possible ancient changes in habitat connectivity (HL&F) or for how species have responded to ancient HL&F (when prior palaeoecological data do exist).

The empirical genetic studies, carried out in Madagascar and elsewhere, on the consequences of HL&F on species genetic diversity are not always clear, because they often identify ***patterns*** in the form of genetic structure or genetic diversity but do not necessarily aim at identifying the ***processes*** responsible for the observed genetic patterns. This is particularly relevant if we are interested in knowing how recent or ancient HL&F was in a given landscape or the extent at which present genetic pattern mirrors either past or present landscape structure. In fact, a clear interpretation of such pattern would require theoretical frameworks that jointly consider the effect of space and time, and can therefore provide some prediction on the time required to observe significant changes in genetic pattern since HL&F. For instance, referring to some of the studies mentioned in the previous section, it is tempting and probably justified to suggest that the fragmented forests currently occupied by *Microcebus*, *Eliurus*, and *Propithecus* in the north, or *M. ravelobensis* in the northwest, were continuous at some point in the 'ancient' or 'recent' past, but without proper spatial and temporal modelling, as well as palaeoecological data, it is difficult to clarify what is meant by 'ancient' or 'recent'. In the present PhD project, I have attempted to fill this gap by modifying some of the classical spatial population genetic models and applying them in the context of HL&F. Moreover, I have analysed genetic and genomic data from rodent (*Eliurus spp.*) and primate (*Microcebus spp.*)

species endemic to Madagascar to provide some insights on past HL&F in northern Madagascar and how species may have responded to that.

1.8 References

Aguilar, R., Quesada, M., Ashworth, L., Herrerias-Diego, Y., Lobo, J., 2008. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. Molecular Ecology 17, 5177–5188.

Aguillon, S.M., Fitzpatrick, J.W., Bowman, R., Schoech, S.J., Clark, A.G., Coop, G., Chen, N., 2017. Deconstructing isolation-by-distance: The genomic consequences of limited dispersal. PLOS Genetics 13, e1006911.

Al-Asadi, H., Petkova, D., Stephens, M., Novembre, J., 2019. Estimating recent migration and population-size surfaces. PLOS Genetics 15, e1007908.

Aleixo-Pais, I., Salmona, J., Sgarlata, G.M., Rakotonanahary, A., Sousa, A.P., Parreira, B., Kun-Rodrigues, C., Ralantoharijaona, T., Jan, F., Rasolondraibe, E., Minhós, T., Zaonarivelo, J.R., Andriaholinirina, N.V., Chikhi, L., 2019. The genetic structure of a mouse lemur living in a fragmented habitat in Northern Madagascar. Conserv Genet 20, 229–243.

Ali, J.R., Aitchison, J.C., 2008. Gondwana to Asia: Plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166–35 Ma). Earth-Science Reviews 88, 145–166.

Ali, J.R., Huber, M., 2010. Mammalian biodiversity on Madagascar controlled by ocean currents. Nature 463, 653–656.

Ali, J.R., Krause, D.W., 2011. Late Cretaceous bioconnections between Indo-Madagascar and Antarctica: refutation of the Gunnerus Ridge causeway

hypothesis. Journal of Biogeography 38, 1855–1872.

Almeida-Rocha, J.M., Soares, L.A.S.S., Andrade, E.R., Gaiotto, F.A., Cazetta, E., 2020. The impact of anthropogenic disturbances on the genetic diversity of terrestrial species: A global meta-analysis. Molecular Ecology 29, 4812–4822.

Alroy, J., 2001. A Multispecies Overkill Simulation of the End-Pleistocene Megafaunal Mass Extinction. Science 292, 1893–1896.

Anantharaman, R., Hall, K., Shah, V.B., Edelman, A., 2020. Circuitscape in Julia: High Performance Connectivity Modelling to Support Conservation Decisions. Proceedings of the JuliaCon Conferences 1, 58.

Baccini, A., Walker, W., Carvalho, L., Farina, M., Sulla-Menashe, D., Houghton, R.A., 2017. Tropical forests are a net carbon source based on aboveground measurements of gain and loss. Science 358, 230–234.

Baden, A.L., Mancini, A.N., Federman, S., Holmes, S.M., Johnson, S.E., Kamilar, J., Louis, E.E., Bradley, B.J., 2019. Anthropogenic pressures drive population genetic structuring across a Critically Endangered lemur species range. Scientific Reports 9, 16276.

Balkenhol, N., Waits, L.P., Dezzani, R.J., 2009. Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. Ecography 32, 818–830.

Barlow, J., Gardner, T.A., Araujo, I.S., Ávila-Pires, T.C., Bonaldo, A.B., Costa, J.E., Esposito, M.C., Ferreira, L.V., Hawes, J., Hernandez, M.I.M., Hoogmoed, M.S., Leite, R.N., Lo-Man-Hung, N.F., Malcolm, J.R., Martins, M.B., Mestre, L. a. M., Miranda-Santos, R., Nunes-Gutjahr, A.L., Overal, W.L., Parry, L., Peters, S.L., Ribeiro-Junior, M.A., Silva, M.N.F. da, Motta, C. da S., Peres, C.A., 2007. Quantifying the biodiversity value of tropical primary,

secondary, and plantation forests. PNAS 104, 18555–18560.

Barlow, J., Lennox, G.D., Ferreira, J., Berenguer, E., Lees, A.C., Nally, R.M., Thomson, J.R., Ferraz, S.F. de B., Louzada, J., Oliveira, V.H.F., Parry, L., Ribeiro de Castro Solar, R., Vieira, I.C.G., Aragão, L.E.O.C., Begotti, R.A., Braga, R.F., Cardoso, T.M., Jr, R.C. de O., Souza Jr, C.M., Moura, N.G., Nunes, S.S., Siqueira, J.V., Pardini, R., Silveira, J.M., Vaz-de-Mello, F.Z., Veiga, R.C.S., Venturieri, A., Gardner, T.A., 2016. Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. Nature 535, 144–147.

Barnosky, A.D., Koch, P.L., Feranec, R.S., Wing, S.L., Shabel, A.B., 2004. Assessing the Causes of Late Pleistocene Extinctions on the Continents. Science 306, 70–75.

Barton, Etheridge, A., Véber, A., 2010a. A New Model for Evolution in a Spatial Continuum. Electronic Journal of Probability 15, 162–216.

Barton, Kelleher, J., Etheridge, A.M., 2010b. A new model for extinction and recolonization in two dimensions: quantifying phylogeography. Evolution 64, 2701–2715.

Barton, N., Etheridge, A., 2019. Mathematical Models in Population Genetics, in: Handbook of Statistical Genomics. John Wiley & Sons, Ltd, pp. 115–20.

Barton, N.H., 2008. The effect of a barrier to gene flow on patterns of geographic variation. Genetics Research 90, 139–149.

Barton, N.H., Depaulis, F., Etheridge, A.M., 2002. Neutral Evolution in Spatially Continuous Populations. Theoretical Population Biology 61, 31–48.

Barton, N.H., Etheridge, A.M., Kelleher, J., Véber, A., 2013. Inference in two dimensions: Allele frequencies versus lengths of shared sequence blocks. Theoretical Population Biology, Coalescent Theory 87, 105–119.

Barton, N.H., Etheridge, A.M., Véber, A., 2010. A new model for evolution in a spatial continuum. Electronic Journal of Probability 15, 162–216.

Barton, N.H., Wilson, I., 1995. Genealogies and geography. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 349, 49–59.

Beaumont, M.A., 1999. Detecting population expansion and decline using microsatellites. Genetics 153, 2013–2029.

Betts, M.G., Fahrig, L., Hadley, A.S., Halstead, K.E., Bowman, J., Robinson, W.D., Wiens, J.A., Lindenmayer, D.B., 2014. A species-centered approach for uncovering generalities in organism responses to habitat loss and fragmentation. Ecography 37, 517–527.

Betts, M.G., Wolf, C., Pfeifer, M., Banks-Leite, C., Arroyo-Rodríguez, V., Ribeiro, D.B., Barlow, J., Eigenbrod, F., Faria, D., Fletcher, R.J., Hadley, A.S., Hawes, J.E., Holt, R.D., Klingbeil, B., Kormann, U., Lens, L., Levi, T., Medina-Rangel, G.F., Melles, S.L., Mezger, D., Morante-Filho, J.C., Orme, C.D.L., Peres, C.A., Phalan, B.T., Pidgeon, A., Possingham, H., Ripple, W.J., Slade, E.M., Somarriba, E., Tobias, J.A., Tylianakis, J.M., Urbina-Cardona, J.N., Valente, J.J., Watling, J.I., Wells, K., Wearn, O.R., Wood, E., Young, R., Ewers, R.M., 2019. Extinction filters mediate the global effects of habitat fragmentation on animals. Science 366, 1236–1239.

Bond, W.J., Silander Jr, J.A., Ranaivonasy, J., Ratsirarson, J., 2008. The antiquity of Madagascar's grasslands and the rise of C4 grassy biomes. Journal of Biogeography 35, 1743–1758.

Bradburd, G.S., Ralph, P.L., 2019. Spatial Population Genetics: It's About Time. Annual Review of Ecology, Evolution, and Systematics 50, 427–449.

Bradburd, G.S., Ralph, P.L., Coop, G.M., 2016. A Spatial Framework for

Understanding Population Structure and Admixture. PLOS Genetics 12, e1005703.

Bradburd, G.S., Ralph, P.L., Coop, G.M., 2013. Disentangling the Effects of Geographic and Ecological Isolation on Genetic Differentiation. Evolution 67, 3258–3273.

Bradley, R.S., 1985. Quaternary Paleoclimatology: Methods of Paleoclimatic Reconstruction. Allen & Unwin, Boston.

Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., Fonseca, G.A.B.D., Rylands, A.B., Konstant, W.R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G., Hilton-Taylor, C., 2002. Habitat Loss and Extinction in the Hotspots of Biodiversity. Conservation Biology 16, 909–923.

Broughton, J.M., Weitzel, E.M., 2018. Population reconstructions for humans and megafauna suggest mixed causes for North American Pleistocene extinctions. Nat Commun 9, 1–12.

Bruggeman, D.J., Wiegand, T., Fernández, N., 2010. The relative effects of habitat loss and fragmentation on population genetic variation in the red-cockaded woodpecker (Picoides borealis). Mol Ecol 19, 3679–3691.

Bueno, A.S., Peres, C.A., 2019. Patch-scale biodiversity retention in fragmented landscapes: Reconciling the habitat amount hypothesis with the island biogeography theory. Journal of Biogeography 46, 621–632.

Burney, D., James, H., Grady, F., Rafamantanantsoa, J.-G., Ramilisonina, Wright, H., Cowart, J., 1997. Environmental change, extinction and human activity: evidence from caves in NW Madagascar. Journal of Biogeography 24, 755–767.

Burney, D.A., 1993. Late Holocene Environmental Changes in Arid Southwestern Madagascar. Quaternary Research 40, 98–106.

Burney, D.A., 1987a. Late Quaternary stratigraphic charcoal records from Madagascar. Quaternary Research 28, 274–280.

Burney, D.A., 1987b. Late Holocene Vegetational Change in Central Madagascar. Quaternary Research 28, 130–143.

Burney, D.A., Flannery, T.F., 2005. Fifty millennia of catastrophic extinctions after human contact. Trends in Ecology & Evolution 20, 395–401.

Burney, D.A., MacPhee, R.D.E., 1988. Mysterious island: what killed Madagascar's large na- tive animals? Nat. Hist. 97, 46–55.

Burns, S.J., Godfrey, L.R., Faina, P., McGee, D., Hardt, B., Ranivoharimanana, L., Randrianasy, J., 2016. Rapid human-induced landscape transformation in Madagascar at the end of the first millennium of the Common Era. Quaternary Science Reviews 134, 92–99.

Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M., Palmer, T.M., 2015. Accelerated modern human–induced species losses: Entering the sixth mass extinction. Science Advances 1, e1400253.

Ceballos, G., Ehrlich, P.R., Dirzo, R., 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. PNAS 114, E6089–E6096.

Craul, M., Chikhi, L., Sousa, V., Olivieri, G.L., Rabesandratana, A., Zimmermann, E., Radespiel, U., 2009. Influence of forest fragmentation on an endangered large-bodied lemur in northwestern Madagascar. Biological Conservation 142, 2862–2871.

Crowley, B., Godfrey, L., Burney, D., 2010. New support for a human hand in the collapse of Madagascar's megafaunal community: Using C-14 dates to track species persistence and population decline during the Holocene.

Crowley, B.E., 2010. A refined chronology of prehistoric Madagascar and the demise of the megafauna. Quaternary Science Reviews, Special Theme: Case Studies of Neodymium Isotopes in Paleoceanography 29, 2591–2603.

Crowley, B.E., Godfrey, L.R., Bankoff, R.J., Perry, G.H., Culleton, B.J., Kennett, D.J., Sutherland, M.R., Samonds, K.E., Burney, D.A., 2017. Island-wide aridity did not trigger recent megafaunal extinctions in Madagascar. Ecography 40, 901–912.

Curtis, J.T., 1956. The modification of mid-latitude grasslands and forests by man., in: Man's Role in Changing the Face of the Earth. University of Chicago Press., Chicago, IL, pp. 721–736.

Cushman, S.A., McKelvey, K.S., Hayden, J., Schwartz, M.K., 2006. Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. Am. Nat. 168, 486–499.

Cushman, S.A., Shirk, A., Landguth, E.L., 2012. Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex landscapes. Landscape Ecol 27, 369–380.

deMenocal, P., Ortiz, J., Guilderson, T., Adkins, J., Sarnthein, M., Baker, L., Yarusinsky, M., 2000. Abrupt onset and termination of the African Humid Period:: rapid climate responses to gradual insolation forcing. Quaternary Science Reviews 19, 347–361.

Denton, G.H., Anderson, R.F., Toggweiler, J.R., Edwards, R.L., Schaefer, J.M., Putnam, A.E., 2010. The Last Glacial Termination. Science 328, 1652–1656.

Dewar, R.E., Richard, A.F., 2007. Evolution in the hypervariable environment of Madagascar. PNAS 104, 13723–13727.

Didham, R.K., Kapos, V., Ewers, R.M., 2012. Rethinking the conceptual

foundations of habitat fragmentation research. Oikos 121, 161–170.

Didham, R.K., Lawton, J.H., 1999. Edge Structure Determines the Magnitude of Changes in Microclimate and Vegetation Structure in Tropical Forest Fragments1. Biotropica 31, 17–30.

DiLeo, M.F., Wagner, H.H., 2016. A Landscape Ecologist's Agenda for Landscape Genetics. Curr Landscape Ecol Rep 1, 115–126.

Douady, C.J., Catzeflis, F., Kao, D.J., Springer, M.S., Stanhope, M.J., 2002. Molecular Evidence for the Monophyly of Tenrecidae (Mammalia) and the Timing of the Colonization of Madagascar by Malagasy Tenrecs. Molecular Phylogenetics and Evolution 22, 357–363.

Douglass, K., Hixon, S., Wright, H.T., Godfrey, L.R., Crowley, B.E., Manjakahery, B., Rasolondrainy, T., Crossland, Z., Radimilahy, C., 2019. A critical review of radiocarbon dates clarifies the human settlement of Madagascar. Quaternary Science Reviews 221, 105878.

Douglass, K., Zinke, J., 2015. Forging Ahead By Land and By Sea: Archaeology and Paleoclimate Reconstruction in Madagascar. The African Archaeological Review 32, 267–299.

Eagles, G., König, M., 2008. A model of plate kinematics in Gondwana breakup. Geophysical Journal International 173, 703–717.

Ellegren, H., Galtier, N., 2016. Determinants of genetic diversity. Nat Rev Genet 17, 422–433. https://doi.org/10.1038/nrg.2016.58

Epps, C.W., Keyghobadi, N., 2015. Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. Molecular Ecology 24, 6021–6040.

Etheridge, A., Veber, A., Yu, F., 2020. Rescaling limits of the spatial Lambda-

Fleming-Viot process with selection. Electron. J. Probab. 25.

Etheridge, A.M., 2008. Drift, draft and structure: some mathematical models of evolution. Banach Center Publications 80, 121–144.

Everson, K.M., Soarimalala, V., Goodman, S.M., Olson, L.E., 2016. Multiple Loci and Complete Taxonomic Sampling Resolve the Phylogeny and Biogeographic History of Tenrecs (Mammalia: Tenrecidae) and Reveal Higher Speciation Rates in Madagascar's Humid Forests. Syst. Biol. 65, 890–909.

Ewers, R.M., Didham, R.K., 2008. Pervasive impact of large-scale edge effects on a beetle community. PNAS 105, 5426–5429.

Fahrig, L., 2019. Habitat fragmentation: A long and tangled tale. Global Ecology and Biogeography 28, 33–41.

Fahrig, L., 2017. Ecological Responses to Habitat Fragmentation Per Se. Annual Review of Ecology, Evolution, and Systematics 48, 1–23.

Fahrig, L., 2003. Effects of Habitat Fragmentation on Biodiversity. Annu. Rev. Ecol. Evol. Syst. 34, 487–515.

Fahrig, L., Arroyo-Rodríguez, V., Bennett, J.R., Boucher-Lalonde, V., Cazetta, E., Currie, D.J., Eigenbrod, F., Ford, A.T., Harrison, S.P., Jaeger, J.A.G., Koper, N., Martin, A.E., Martin, J.-L., Metzger, J.P., Morrison, P., Rhodes, J.R., Saunders, D.A., Simberloff, D., Smith, A.C., Tischendorf, L., Vellend, M., Watling, J.I., 2019. Is habitat fragmentation bad for biodiversity? Biological Conservation 230, 179–186.

Felsenstein, J., 1975. A Pain in the Torus: Some Difficulties with Models of Isolation by Distance. The American Naturalist 109, 359–368.

Fisher, B.L., Robertson, H.G., 2002. Comparison and Origin of Forest and

Grassland Ant Assemblages in the High Plateau of Madagascar (Hymenoptera: Formicidae)1. Biotropica 34, 155–167.

Fletcher, R.J., Didham, R.K., Banks-Leite, C., Barlow, J., Ewers, R.M., Rosindell, J., Holt, R.D., Gonzalez, A., Pardini, R., Damschen, E.I., Melo, F.P.L., Ries, L., Prevedello, J.A., Tscharntke, T., Laurance, W.F., Lovejoy, T., Haddad, N.M., 2018. Is habitat fragmentation good for biodiversity? Biological Conservation 226, 9–15.

Ganzhorn, J.U., Goodman, S.M., Nash, S., Thalmann, U., 2006. Lemur Biogeography, in: Lehman, S.M., Fleagle, J.G. (Eds.), Primate Biogeography: Progress and Prospects, Developments in Primatology: Progress and Prospects. Springer US, Boston, MA, pp. 229–254.

Ganzhorn, J.U., Wilmé, L., Mercier, J., 2014. Explaining Madagascar's biodiversity, in: Conservation and Environmental Management in Madagascar. Routledge, New York, pp. 17–43.

Gasse, F., 2000. Hydrological changes in the African tropics since the Last Glacial Maximum. Quaternary Science Reviews 19, 189–211.

Gasse, F., Van Campo, E., 1998. A 40,000-yr Pollen and Diatom Record from Lake Tritrivakely, Madagascar, in the Southern Tropics. Quaternary Research 49, 299–311.

Gibson, L., Lee, T.M., Koh, L.P., Brook, B.W., Gardner, T.A., Barlow, J., Peres, C.A., Bradshaw, C.J.A., Laurance, W.F., Lovejoy, T.E., Sodhi, N.S., 2011. Primary forests are irreplaceable for sustaining tropical biodiversity. Nature 478, 378–381.

Godfrey, L.R., Jungers, W.L., 2003. The extinct sloth lemurs of Madagascar. Evolutionary Anthropology: Issues, News, and Reviews 12, 252–263.

Goodman, S.M., Benstead, J.P., 2005. Updated estimates of biotic diversity

and endemism for Madagascar. Oryx 39, 73–77.

Graham, R.W., Lundelius, E.L., Graham, M.A., Schroeder, E.K., Toomey, R.S., Anderson, E., Barnosky, A.D., Burns, J.A., Churcher, C.S., Grayson, D.K., Guthrie, R.D., Harington, C.R., Jefferson, G.T., Martin, L.D., McDonald, H.G., Morlan, R.E., Semken, H.A., Webb, S.D., Werdelin, L., Wilson, M.C., 1996. Spatial Response of Mammals to Late Quaternary Environmental Fluctuations. Science 272, 1601–1606.

Graves, T.A., Wasserman, T.N., Ribeiro, M.C., Landguth, E.L., Spear, S.F., Balkenhol, N., Higgins, C.B., Fortin, M.-J., Cushman, S.A., Waits, L.P., 2012. The influence of landscape characteristics and home-range size on the quantification of landscape-genetics relationships. Landscape Ecology. 27: 253-266. 253–266.

Grebenkov, D.S., Nguyen, D.V., Li, J.-R., 2014. Exploring diffusion across permeable barriers at high gradients. I. Narrow pulse approximation. J Magn Reson 248, 153–163.

Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D., Cook, W.M., Damschen, E.I., Ewers, R.M., Foster, B.L., Jenkins, C.N., King, A.J., Laurance, W.F., Levey, D.J., Margules, C.R., Melbourne, B.A., Nicholls, A.O., Orrock, J.L., Song, D.-X., Townshend, J.R., 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. Science Advances 1, e1500052.

Haddad, N.M., Gonzalez, A., Brudvig, L.A., Burt, M.A., Levey, D.J., Damschen, E.I., 2017. Experimental evidence does not support the Habitat Amount Hypothesis. Ecography 40, 48–55.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G.,

2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342, 850–853.

Hanski, I., 2015. Habitat fragmentation and species richness. Journal of Biogeography 42, 989–993.

Hanski, I., 2011. Habitat Loss, the Dynamics of Biodiversity, and a Perspective on Conservation. AMBIO 40, 248–255.

Hanski, I., 2005. The shrinking world: Ecological consequences of habitat loss, EXCELLENCE IN ECOLOGY. International Ecology Institute, Oldendorf (Luhe).

Hanski, I., 1994. A Practical Model of Metapopulation Dynamics. Journal of Animal Ecology 63, 151–162.

Hanski, I., Pakkala, T., Kuussaari, M., Lei, G., 1995. Metapopulation Persistence of an Endangered Butterfly in a Fragmented Landscape. Oikos 72, 21–28.

Hanski, I., Zurita, G.A., Bellocq, M.I., Rybicki, J., 2013. Species–fragmented area relationship. PNAS 110, 12715–12720.

Hassan, F.A., 1997. Holocene Palaeoclimates of Africa. The African Archaeological Review 14, 213–230.

Hendry, A.P., 2004. Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. Evol Ecol Res 6, 1219–1236.

Herbots, H.M.J.D., 1994. Stochastic Models in Population Genetics: Genealogy and Genetic Differentiation in Structured Populations (Thesis).

Herrera, J.P., Dávalos, L.M., 2016. Phylogeny and Divergence Times of

Lemurs Inferred with Recent and Ancient Fossils in the Tree. Syst Biol 65, 772–791.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907–913.

Hey, J., 1991. A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. Theoretical Population Biology 39, 30–48.

Hixon, S.W., Curtis, J.H., Brenner, M., Douglass, K.G., Domic, A.I., Culleton, B.J., Ivory, S.J., Kennett, D.J., 2021a. Drought Coincided with, but Does Not Explain, Late Holocene Megafauna Extinctions in SW Madagascar. Climate 9, 138.

Hixon, S.W., Douglass, K.G., Crowley, B.E., Rakotozafy, L.M.A., Clark, G., Anderson, A., Haberle, S., Ranaivoarisoa, J.F., Buckley, M., Fidiarisoa, S., Mbola, B., Kennett, D.J., 2021b. Late Holocene spread of pastoralism coincides with endemic megafaunal extinction on Madagascar. Proceedings of the Royal Society B: Biological Sciences 288, 20211204.

Honnay, O., Jacquemyn, H., 2007. Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. Conserv Biol 21, 823–831.

Humbert, H., 1927. La destruction d'une flore insulaire par le feu. Principaux aspects de la végétation à Madagascar. Mem. Acad. Malgache 5, 1–78.

Hylander, K., Ehrlén, J., 2013. The mechanisms causing extinction debts. Trends in Ecology & Evolution 28, 341–346.

Imbrie, J., Imbrie, K.P., 1986. Ice Ages: Solving the Mystery, Revised edition. ed. Harvard University Press, Cambridge, Mass.

Irwin, M.T., Wright, P.C., Birkinshaw, C., Fisher, B.L., Gardner, C.J., Glos, J., Goodman, S.M., Loiselle, P., Rabeson, P., Raharison, J.-L., Raherilalao, M.J., Rakotondravony, D., Raselimanana, A., Ratsimbazafy, J., Sparks, J.S., Wilmé, L., Ganzhorn, J.U., 2010. Patterns of species change in anthropogenically disturbed forests of Madagascar. Biological Conservation, Conserving complexity: Global change and community-scale interactions 143, 2351–2362.

Ivory, S.J., Lézine, A.-M., Vincens, A., Cohen, A.S., 2012. Effect of aridity and rainfall seasonality on vegetation in the southern tropics of East Africa during the Pleistocene/Holocene transition. Quaternary Research 77, 77–86.

Jackson, N.D., Fahrig, L., 2016. Habitat amount, not habitat configuration, best predicts population genetic structure in fragmented landscapes. Landscape Ecol 31, 951–968.

James, P.M.A., Coltman, D.W., Murray, B.W., Hamelin, R.C., Sperling, F.A.H., 2011. Spatial genetic structure of a symbiotic beetle-fungal system: toward multi-taxa integrated landscape genetics. PLoS One 6, e25359.

Jansa, S.A., Carleton, M.D., Soarimalala, V., Rakotomalala, Z., Goodman, S.M., 2019. A review of the Eliurus tanala complex (Rodentia, Muroidea, Nesomyidae), with description of a new species from dry forests of western Madagascar. (Bulletin of the American Museum of Natural History, no. 430).

Karanth, K.P., Delefosse, T., Rakotosamimanana, B., Parsons, T.J., Yoder, A.D., 2005. Ancient DNA from giant extinct lemurs confirms single origin of Malagasy primates. PNAS 102, 5090–5095.

Keenan, R.J., Reams, G.A., Achard, F., de Freitas, J.V., Grainger, A., Lindquist, E., 2015. Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. Forest Ecology and Management, Changes in Global Forest Resources from 1990 to 2015 352,

9–20.

Keyghobadi, N., Roland, J., Matter, S.F., Strobeck, C., 2005. Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. Proceedings of the Royal Society B: Biological Sciences 272, 553–560.

Kimura, M., 1953. “Stepping Stone” model of population. Ann. Rept. Nat. Inst. Genetics Japan 3, 62–63.

Kimura, M., Weiss, G.H., 1964. The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. Genetics 49, 561–576.

Koch, P.L., Barnosky, A.D., 2006. Late Quaternary Extinctions: State of the Debate. Annual Review of Ecology, Evolution, and Systematics 37, 215–250.

Koechlin, J., 1972. Flora and Vegetation of Madagascar, in: Battistini, R., Richard-Vindard, G. (Eds.), Biogeography and Ecology in Madagascar, Monographiae Biologicae. Springer Netherlands, Dordrecht, pp. 145–190.

Kutzbach, J.E., Street-Perrott, F.A., 1985. Milankovitch forcing of fluctuations in the level of tropical lakes from 18 to 0 kyr BP. Nature 317, 130–134.

Landguth, E.L., Cushman, S.A., Schwartz, M.K., McKelvey, K.S., Murphy, M., Luikart, G., 2010. Quantifying the lag time to detect barriers in landscape genetics. Mol. Ecol. 19, 4179–4191.

Lavauden, L., 1931. Le Déboisement et la Végétation de Madagascar. Journal d’agriculture traditionnelle et de botanique appliquée 11, 817–824.

Leblois, R., Estoup, A., Streiff, R., 2006. Genetics of recent habitat contraction and reduction in population size: does isolation by distance matter? Mol Ecol 15, 3601–3615.

Lewis, S.L., Maslin, M.A., 2015. Defining the Anthropocene. Nature 519, 171–180.

Lino, A., Fonseca, C., Rojas, D., Fischer, E., Ramos Pereira, M.J., 2019. A meta-analysis of the effects of habitat loss and fragmentation on genetic diversity in mammals. Mammalian Biology 94, 69–76.

Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B., Ackerly, D.D., 2009. The velocity of climate change. Nature 462, 1052–1055.

Lowry, P., Schatz, G., Phillipson, P., 1997. The classification of natural and anthropogenic vegetation in Madagascar. pp. 93–123.

Lyons, S.K., 2005. A Quantitative Model for Assessing Community Dynamics of Pleistocene Mammals. The American Naturalist 165, E168–E185.

Lyons, S.K., 2003. A Quantitative Assessment of the Range Shifts of Pleistocene Mammals. Journal of Mammalogy 84, 385–402.

MacPhee, R.D.E., Burney, D.A., 1991. Dating of modified femora of extinct dwarf Hippopotamus from Southern Madagascar: Implications for constraining human colonization and vertebrate extinction events. Journal of Archaeological Science 18, 695–706.

MacPhee, R.D.E., Burney, D.A., Wells, N.A., 1985. Early Holocene chronology and environment of Ampasambazimba, A Malagasy subfossil lemur site. Int J Primatol 6, 463–489.

Malécot, G., 1948. Les mathématiques de l'hérédité. Barnéoud frères.

Malhi, Y., Phillips, O. (Eds.), 2005. Tropical Forests and Global Atmospheric Change. Oxford University Press, Oxford ; New York.

Martin, A.E., Fahrig, L., 2012. Measuring and selecting scales of effect for

landscape predictors in species–habitat models. Ecological Applications 22, 2277–2292.

Martin, R.D., 1972. Adaptive radiation and behaviour of the Malagasy lemurs. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 264, 295–352.

Maruyama, T., 1972. Rate of Decrease of Genetic Variability in a Two-Dimensional Continuous Population of Finite Size. Genetics 70, 639–651.

Maruyama, T., 1971. Analysis of population structure. II. Two-dimensional stepping stone models of finite length and other geographically structured populations. Ann. Hum. Genet. 35, 179–196.

Maruyama, T., 1970. Stepping stone models of finite length. Advances in Applied Probability 2, 229–258.

Masters, J.C., Génin, F., Zhang, Y., Pellen, R., Huck, T., Mazza, P.P.A., Rabineau, M., Doucouré, M., Aslanian, D., 2021. Biogeographic mechanisms involved in the colonization of Madagascar by African vertebrates: Rifting, rafting and runways. Journal of Biogeography 48, 492–510.

McCall, R.A., 1997. Implications of recent geological investigations of the Mozambique Channel for the mammalian colonization of Madagascar. Proc Biol Sci 264, 663–665.

McRae, B.H., 2006. Isolation by Resistance. Evolution 60, 1551–1561.

MEF, (Ministry of Environment and Forests), 2014. Fifth National Report to the Conven- tion on Biological Diversity – Madagascar.

Meirmans, P.G., 2012. The trouble with isolation by distance. Molecular Ecology 21, 2839–2846.

Mercier, J.-L., Wilmé, L., 2013. The Eco-Geo-Clim model: explaining

Madagascar's endemism. Madagascar Conservation & Development 8, 63–68.

Miraldo, A., Li, S., Borregaard, M.K., Flórez-Rodríguez, A., Gopalakrishnan, S., Rizvanovic, M., Wang, Z., Rahbek, C., Marske, K.A., Nogués-Bravo, D., 2016. An Anthropocene map of genetic diversity. Science 353, 1532–1535.

Moat, J., Smith, P., 2007. Atlas of the Vegetation of Madagascar, Bilingual edition. ed. Royal Botanic Gardens, Kew, Kew.

Mona, S., Ray, N., Arenas, M., Excoffier, L., 2014. Genetic consequences of habitat fragmentation during a range expansion. Heredity 112, 291–299.

Monteiro, W.P., Veiga, J.C., Silva, A.R., Carvalho, C. da S., Lanes, É.C.M., Rico, Y., Jaffé, R., 2019. Everything you always wanted to know about gene flow in tropical landscapes (but were afraid to ask). PeerJ 7, e6446.

Moore, N.W., 1962. The Heaths of Dorset and their Conservation. Journal of Ecology 50, 369–391.

Moritz, C., Patton, J.L., Schneider, C.J., Smith, T.B., 2000. Diversification of Rainforest Faunas: An Integrated Molecular Approach. Annual Review of Ecology and Systematics 31, 533–563.

Nagylaki, T., 1988. The influence of spatial inhomogeneities on neutral models of geographical variation: I. Formulation. Theoretical Population Biology 33, 291–310.

Nagylaki, T., 1974. The Decay of Genetic Variability in Geographically Structured Populations*. Proc Natl Acad Sci U S A 71, 2932–2936.

Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Lysenko, I., Senior, R.A., Börger, L., Bennett, D.J., Choimes, A., Collen, B., Day, J., De Palma, A., Díaz, S., Echeverria-Londoño, S., Edgar, M.J., Feldman, A., Garon, M., Harrison,

M.L.K., Alhusseini, T., Ingram, D.J., Itescu, Y., Kattge, J., Kemp, V., Kirkpatrick, L., Kleyer, M., Correia, D.L.P., Martin, C.D., Meiri, S., Novosolov, M., Pan, Y., Phillips, H.R.P., Purves, D.W., Robinson, A., Simpson, J., Tuck, S.L., Weiher, E., White, H.J., Ewers, R.M., Mace, G.M., Scharlemann, J.P.W., Purvis, A., 2015. Global effects of land use on local terrestrial biodiversity. Nature 520, 45–50.

Nosil, P., Egan, S.P., Funk, D.J., 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: “isolation by adaptation” and multiple roles for divergent selection. Evolution 62, 316–336.

Olivieri, G.L., Sousa, V., Chikhi, L., Radespiel, U., 2008. From genetic diversity and structure to conservation: Genetic signature of recent population declines in three mouse lemur species (Microcebus spp.). Biological Conservation 141, 1257–1271.

Pereira, H.M., Leadley, P.W., Proença, V., Alkemade, R., Scharlemann, J.P.W., Fernandez-Manjarrés, J.F., Araújo, M.B., Balvanera, P., Biggs, R., Cheung, W.W.L., Chini, L., Cooper, H.D., Gilman, E.L., Guénette, S., Hurtt, G.C., Huntington, H.P., Mace, G.M., Oberdorff, T., Revenga, C., Rodrigues, P., Scholes, R.J., Sumaila, U.R., Walpole, M., 2010. Scenarios for Global Biodiversity in the 21st Century. Science 330, 1496–1501.

Perrier de la Bâthie, H., 1921. La végétation Malgache. Ann. Mus. Colon. Marseille 9, 1–266.

Petkova, D., Novembre, J., Stephens, M., 2016. Visualizing spatial population structure with estimated effective migration surfaces. Nature Genetics 48, 94–100.

Pfeifer, M., Lefebvre, V., Peres, C.A., Banks-Leite, C., Wearn, O.R., Marsh, C.J., Butchart, S.H.M., Arroyo-Rodríguez, V., Barlow, J., Cerezo, A., Cisneros, L., D’Cruze, N., Faria, D., Hadley, A., Harris, S.M., Klingbeil, B.T.,

Kormann, U., Lens, L., Medina-Rangel, G.F., Morante-Filho, J.C., Olivier, P., Peters, S.L., Pidgeon, A., Ribeiro, D.B., Scherber, C., Schneider-Maunoury, L., Struebig, M., Urbina-Cardona, N., Watling, J.I., Willig, M.R., Wood, E.M., Ewers, R.M., 2017. Creation of forest edges has a global impact on forest vertebrates. Nature 551, 187–191.

Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M., Sexton, J.O., 2014. The biodiversity of species and their rates of extinction, distribution, and protection. Science 344.

Poux, C., Madsen, O., Glos, J., de Jong, W.W., Vences, M., 2008. Molecular phylogeny and divergence times of Malagasy tenrecs: Influence of data partitioning and taxon sampling on dating analyses. BMC Evolutionary Biology 8, 102.

Poux, C., Madsen, O., Marquard, E., Vieites, D.R., de Jong, W.W., Vences, M., 2005. Asynchronous colonization of Madagascar by the four endemic clades of primates, tenrecs, carnivores, and rodents as inferred from nuclear genes. Syst. Biol. 54, 719–730.

Powzyk, J.A., Mowry, C.B., 2003. Dietary and Feeding Differences Between Sympatric Propithecus diadema diadema and Indri indri. International Journal of Primatology 24, 1143–1162.

Prescott, G.W., Williams, D.R., Balmford, A., Green, R.E., Manica, A., 2012. Quantitative global analysis of the role of climate and people in explaining late Quaternary megafaunal extinctions. PNAS 109, 4527–4531.

Prevedello, J.A., Vieira, M.V., 2010. Does the type of matrix matter? A quantitative review of the evidence. Biodivers Conserv 19, 1205–1223.

Quéméré, E., Amelot, X., Pierson, J., Crouau-Roy, B., Chikhi, L., 2012. Genetic data suggest a natural prehuman origin of open habitats in northern

Madagascar and question the deforestation narrative in this region. PNAS 109, 13028–13033.

Quéméré, E., Crouau-Roy, B., Rabarivola, C., Jr, E.E.L., Chikhi, L., 2010a. Landscape genetics of an endangered lemur (Propithecus tattersalli) within its entire fragmented range. Molecular Ecology 19, 1606–1621.

Quéméré, E., Louis, E.E., Ribéron, A., Chikhi, L., Crouau-Roy, B., 2010b. Non-invasive conservation genetics of the critically endangered golden-crowned sifaka (Propithecus tattersalli): high diversity and significant genetic differentiation over a small range. Conserv Genet 11, 675–687.

Quevedo, A.A., Schleuning, M., Hensen, I., Saavedra, F., Durka, W., 2013. Forest fragmentation and edge effects on the genetic structure of Clusia sphaerocarpa and C. lechleri (Clusiaceae) in tropical montane forests. Journal of Tropical Ecology 29, 321–329.

Radespiel, U., Schulte, J., Burke, R.J., Lehman, S.M., 2019. Molecular edge effects in the Endangered golden-brown mouse lemur Microcebus ravelobensis. Oryx 53, 716–726.

Raherilalao, M.J., Wilmé, L., 2008. L'avifaune des forêts sèches malgaches., in: Les Forêts Sèches de Madagascar. pp. 76–105.

Rakotoarisoa, J.E., Raheriarisena, M., Goodman, S.M., 2013. Late Quaternary climatic vegetational shifts in an ecological transition zone of northern Madagascar: insights from genetic analyses of two endemic rodent species. Journal of Evolutionary Biology 26, 1019–1034.

Rakotoarisoa, Raheriarisena, M., Goodman, S.M., 2013. A Phylogeographic Study of the Endemic Rodent Eliurus carletoni (Rodentia: Nesomyinae) in an Ecological Transition Zone of Northern Madagascar. J Hered 104, 23–35.

Ralph, P., Coop, G., 2013. The Geography of Recent Genetic Ancestry

across Europe. PLOS Biology 11, e1001555.

Ramanamahefa, R., McGuire, S.M., Louis, E.E., Brenneman, R.A., 2009a. Population genetic parameter estimates for six populations of crowned lemurs, Eulemur coronatus (Gray, 1842), from northern Madagascar. Lemur News 14, 21–26.

Ramanamahefa, R., McGuire, S.M., Louis, E.E., Brenneman, R.A., 2009b. Population genetic parameter estimates for five populations of San- ford's lemur, Eulemur sanfordi (Archbold, 1932), from northern Madagascar. Lemur News 14, 26–31.

Ries, L., Fletcher, R.J., Battin, J., Sisk, T.D., 2004. Ecological Responses to Habitat Edges: Mechanisms, Models, and Variability Explained. Annu. Rev. Ecol. Evol. Syst. 35, 491–522.

Ries, L., Murphy, S.M., Wimp, G.M., Fletcher, R.J., 2017. Closing Persistent Gaps in Knowledge About Edge Ecology. Curr Landscape Ecol Rep 2, 30–41.

Ringbauer, H., Coop, G., Barton, N.H., 2017. Inferring Recent Demography from Isolation by Distance of Long Shared Sequence Blocks. Genetics 205, 1335–1351.

Ringbauer, H., Kolesnikov, A., Field, D.L., Barton, N.H., 2018. Estimating Barriers to Gene Flow from Distorted Isolation-by-Distance Patterns. Genetics 208, 1231–1245.

Rivera Ortíz, F.A., Aguilar, R., Arizmendi, M.D.C., Quesada, M., Oyama, K., 2015. Habitat fragmentation and genetic variability of tetrapod populations.

Ruffell, J., Banks-Leite, C., Didham, R.K., 2016. Accounting for the causal basis of collinearity when measuring the effects of habitat loss versus habitat fragmentation. Oikos 125, 117–125.

Rybicki, J., Hanski, I., 2013. Species–area relationships and extinctions caused by habitat loss and fragmentation. Ecology Letters 16, 27–38.

Ryo, M., Aguilar-Trigueros, C.A., Pinek, L., Muller, L.A.H., Rillig, M.C., 2019. Basic Principles of Temporal Dynamics. Trends in Ecology & Evolution 34, 723–733.

Salmona, J., Heller, R., Quéméré, E., Chikhi, L., 2017. Climate change and human colonization triggered habitat loss and fragmentation in Madagascar. Mol. Ecol. 26, 5203–5222.

Sandel, B., Arge, L., Dalsgaard, B., Davies, R.G., Gaston, K.J., Sutherland, W.J., Svenning, J.-C., 2011. The Influence of Late Quaternary Climate-Change Velocity on Species Endemism. Science 334, 660–664.

Sandom, C., Faurby, S., Sandel, B., Svenning, J.-C., 2014. Global late Quaternary megafauna extinctions linked to humans, not climate change. Proceedings of the Royal Society B: Biological Sciences 281, 20133254.

Schipper, J., Chanson, J.S., Chiozza, F., Cox, N.A., Hoffmann, M., Katariya, V., Lamoreux, J., Rodrigues, A.S.L., Stuart, S.N., Temple, H.J., et al., 2008. The Status of the World's Land and Marine Mammals: Diversity, Threat, and Knowledge. Science 322, 225–230.

Schlaepfer, D.R., Braschler, B., Rusterholz, H.-P., Baur, B., 2018. Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. Ecosphere 9, e02488.

Schneider, N., Chikhi, L., Currat, M., Radespiel, U., 2010. Signals of recent spatial expansions in the grey mouse lemur (Microcebus murinus). BMC Evolutionary Biology 10, 105.

Sexton, J.P., Hangartner, S.B., Hoffmann, A.A., 2014. Genetic Isolation by Environment or Distance: Which Pattern of Gene Flow Is Most Common?

Evolution 68, 1–15.

Sgarlata, G.M., Salmona, J., Aleixo-Pais, I., Rakotonanahary, A., Sousa, A.P., Kun-Rodrigues, C., Ralantoharijaona, T., Jan, F., Zaranaina, R., Rasolondraibe, E., Zaonarivelo, J.R., Andriaholinirina, N.V., Chikhi, L., 2018. Genetic Differentiation and Demographic History of the Northern Rufous Mouse Lemur (Microcebus tavaratra) Across a Fragmented Landscape in Northern Madagascar. Int J Primatol 39, 65–89.

Simpson, G.G., 1940. Mammals and land bridges. Journal of the Washington Academy of Sciences. 30, 137–163.

Sinninghe Damsté, J.S., Verschuren, D., Ossebaar, J., Blokker, J., van Houten, R., van der Meer, M.T.J., Plessen, B., Schouten, S., 2011. A 25,000-year record of climate-induced changes in lowland vegetation of eastern equatorial Africa revealed by the stable carbon-isotopic composition of fossil plant leaf waxes. Earth and Planetary Science Letters 302, 236–246.

Slatkin, M., 2005. Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. Molecular Ecology 14, 67–73.

Slatkin, M., 1993. Isolation by Distance in Equilibrium and Non-Equilibrium Populations. Evolution 47, 264–279.

Slatkin, M., 1991. Inbreeding coefficients and coalescence times. Genetics Research 58, 167–175.

Slatkin, M., Barton, N.H., 1989. A COMPARISON OF THREE INDIRECT METHODS FOR ESTIMATING AVERAGE LEVELS OF GENE FLOW. Evolution 43, 1349–1368.

Song, X.-P., Hansen, M.C., Stehman, S.V., Potapov, P.V., Tyukavina, A., Vermote, E.F., Townshend, J.R., 2018. Global land change from 1982 to

2016. Nature 560, 639–643.

Steppan, S.J., Schenk, J.J., 2017. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. PLOS ONE 12, e0183070.

Storfer, A., Murphy, M.A., Spear, S.F., Holderegger, R., Waits, L.P., 2010. Landscape genetics: where are we now? Molecular Ecology 19, 3496–3514.

Storz, J.F., Beaumont, M.A., Alberts, S.C., 2002. Genetic evidence for long-term population decline in a savannah-dwelling primate: inferences from a hierarchical bayesian model. Mol Biol Evol 19, 1981–1990.

Strobeck, C., 1987. Average Number of Nucleotide Differences in a Sample from a Single Subpopulation: A Test for Population Subdivision. Genetics 117, 149–153.

Tajima, F., 1990. Relationship between migration and DNA polymorphism in a local population. Genetics 126, 231–234.

Taylor, Z.S., Hoffman, S.M.G., 2014. Landscape models for nuclear genetic diversity and genetic structure in white-footed mice (Peromyscus leucopus). Heredity 112, 588–595.

Teixeira, H., Montade, V., Salmona, J., Metzger, J., Bremond, L., Kasper, T., Daut, G., Rouland, S., Ranarilalatiana, S., Rakotondravony, R., Chikhi, L., Behling, H., Radespiel, U., 2021. Past environmental changes affected lemur population dynamics prior to human impact in Madagascar. Commun Biol 4, 1–10.

Thompson, C., van Bogaert, O., Hughes, R., Lippuner, M., Ramahaleo, T., Rafiadanantsoa, S., 2011. Treasure Island: New biodiversity on Madagascar (1999–2010). WWF Report.

Tierney, J.E., deMenocal, P.B., 2013. Abrupt Shifts in Horn of Africa

Hydroclimate Since the Last Glacial Maximum. Science 342, 843–846.

Tierney, J.E., Russell, J.M., Huang, Y., 2010. A molecular perspective on Late Quaternary climate and vegetation change in the Lake Tanganyika basin, East Africa. Quaternary Science Reviews 29, 787–800.

Tittensor, D.P., Walpole, M., Hill, S.L.L., Boyce, D.G., Britten, G.L., Burgess, N.D., Butchart, et al., 2014. A mid-term analysis of progress toward international biodiversity targets. Science 346, 241–244.

Vancutsem, C., Achard, F., Pekel, J.-F., Vieilledent, G., Carboni, S., Simonetti, D., Gallego, J., Aragão, L.E.O.C., Nasi, R., 2021. Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. Science Advances 7, eabe1603.

Varvio, S.-L., Chakraborty, R., Nei, M., 1986. Genetic variation in subdivided populations and conservation genetics. Heredity 57, 189–198.

Vences, M., Freyhof, J., Sonnenberg, R., Kosuch, J., Veith, M., 2001. Reconciling fossils and molecules: Cenozoic divergence of cichlid fishes and the biogeography of Madagascar. Journal of Biogeography 28, 1091–1099.

Vences, M., Kosuch, J., Rödel, M.-O., Lötters, S., Channing, A., Glaw, F., Böhme, W., 2004. Phylogeography of Ptychadena mascareniensis suggests transoceanic dispersal in a widespread African-Malagasy frog lineage. Journal of Biogeography 31, 593–601.

Vences, M., Wollenberg, K.C., Vieites, D.R., Lees, D.C., 2009. Madagascar as a model region of species diversification. Trends in Ecology & Evolution 24, 456–465.

Vieilledent, G., Grinand, C., Rakotomalala, F.A., Ranaivosoa, R., Rakotoarijaona, J.-R., Allnutt, T.F., Achard, F., 2018. Combining global tree cover loss data with historical national forest cover maps to look at six

decades of deforestation and forest fragmentation in Madagascar. Biological Conservation 222, 189–197.

Virah-Sawmy, M., Willis, K.J., Gillson, L., 2010. Evidence for drought and forest declines during the recent megafaunal extinctions in Madagascar. Journal of Biogeography 37, 506–519.

Vorontsova, M.S., Besnard, G., Forest, F., Malakasi, P., Moat, J., Clayton, W.D., Ficinski, P., Savva, G.M., Nanjarisoa, O.P., Razanatsoa, J., Randriatsara, F.O., Kimeu, J.M., Luke, W.R.Q., Kayombo, C., Linder, H.P., 2016. Madagascar's grasses and grasslands: anthropogenic or natural? Proceedings of the Royal Society B: Biological Sciences 283, 20152262.

Vranckx, G., Jacquemyn, H., Muys, B., Honnay, O., 2012. Meta-analysis of susceptibility of woody plants to loss of genetic diversity through habitat fragmentation. Conserv Biol 26, 228–237.

Wagner, H.H., Fortin, M.-J., 2015. Basics of Spatial Data Analysis: Linking Landscape and Genetic Data for Landscape Genetic Studies, in: Landscape Genetics. John Wiley & Sons, Ltd, pp. 77–98.

Wagner, H.H., Fortin, M.-J., 2013. A conceptual framework for the spatial analysis of landscape genetic data. Conserv Genet 14, 253–261.

Walker, M.J.C., Berkelhammer, M., Björck, S., Cwynar, L.C., Fisher, D.A., Long, A.J., Lowe, J.J., Newnham, R.M., Rasmussen, S.O., Weiss, H., 2012. Formal subdivision of the Holocene Series/Epoch: a Discussion Paper by a Working Group of INTIMATE (Integration of ice-core, marine and terrestrial records) and the Subcommission on Quaternary Stratigraphy (International Commission on Stratigraphy). Journal of Quaternary Science 27, 649–659.

Wang, I.J., 2012. Environmental and topographic variables shape genetic structure and effective population sizes in the endangered Yosemite toad.

Diversity and Distributions 18, 1033–1041.

Wang, I.J., Summers, K., 2010. Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. Molecular Ecology 19, 447–458.

Waters, C.N., Zalasiewicz, J., Summerhayes, C., Barnosky, A.D., Poirier, C., Gałuszka, A., Cearreta, A., et al., 2016. The Anthropocene is functionally and stratigraphically distinct from the Holocene. Science 351, aad2622.

Watling, J.I., Arroyo-Rodríguez, V., Pfeifer, M., Baeten, L., Banks-Leite, C., Cisneros, L.M., Fang, R., Hamel-Leigue, A.C., Lachat, T., Leal, I.R., Lens, L., Possingham, H.P., Raheem, D.C., Ribeiro, D.B., Slade, E.M., Urbina-Cardona, J.N., Wood, E.M., Fahrig, L., 2020. Support for the habitat amount hypothesis from a global synthesis of species density studies. Ecology Letters 23, 674–681.

Weigelt, P., Steinbauer, M.J., Cabral, J.S., Kreft, H., 2016. Late Quaternary climate change shapes island biodiversity. Nature 532, 99–102.

White, J.C., Coops, N.C., Wulder, M.A., Vastaranta, M., Hilker, T., Tompalski, P., 2016. Remote Sensing Technologies for Enhancing Forest Inventories: A Review. Canadian Journal of Remote Sensing 42, 619–641.

Wilkins, J.F., 2004. A Separation-of-Timescales Approach to the Coalescent in a Continuous Population. Genetics 168, 2227–2244.

Wilkins, J.F., Wakeley, J., 2002. The Coalescent in a Continuous, Finite, Linear Population. Genetics 161, 873–888.

Wilkinson-Herbots, H.M., 1998. Genealogy and subpopulation differentiation under various models of population structure. J Math Biol 37, 535–585.

Williams, J.W., Summers, R.L., Webb III, T., 1998. Applying plant functional

types to construct biome maps from eastern North American pollen data: comparisons with model results. Quaternary Science Reviews 17, 607–627.

Williams, M., 2003. Deforesting the Earth: From Prehistory to Global Crisis. University of Chicago Press.

Willis, K.J., Gillson, L., Virah-Sawmy, M., 2008. Nature or nurture: the ambiguity of C4 grasslands in Madagascar. Journal of Biogeography 35, 1741–1742.

Wilmé, L., Goodman, S.M., Ganzhorn, J.U., 2006. Biogeographic Evolution of Madagascar's Microendemic Biota. Science 312, 1063–1065.

Wilmé, L., Ravokatra, M., Dolch, R., Schuurman, D., Mathieu, É., Schuetz, H., Waeber, P., 2012. Toponyms for centers of endemism in Madagascar.

Wilson, M.C., Chen, X.-Y., Corlett, R.T., Didham, R.K., Ding, P., Holt, R.D., Holyoak, M., Hu, G., Hughes, A.C., Jiang, L., Laurance, W.F., Liu, J., Pimm, S.L., Robinson, S.K., Russo, S.E., Si, X., Wilcove, D.S., Wu, J., Yu, M., 2016. Habitat fragmentation and biodiversity conservation: key findings and future challenges. Landscape Ecol 31, 219–227.

Wollenberg, K.C., Vieites, D.R., Meijden, A.V.D., Glaw, F., Cannatella, D.C., Vences, M., 2008. Patterns of Endemism and Species Richness in Malagasy Cophyline Frogs Support a Key Role of Mountainous Areas for Speciation. Evolution 62, 1890–1907.

Wright, S., 1943. Isolation by Distance. Genetics 28, 114–138.

Yoder, A.D., Burns, M.M., Zehr, S., Delefosse, T., Veron, G., Goodman, S.M., Flynn, J.J., 2003. Single origin of Malagasy Carnivora from an African ancestor. Nature 421, 734–737.

Yoder, A.D., Campbell, C.R., Blanco, M.B., Reis, M. dos, Ganzhorn, J.U.,

Goodman, S.M., Hunnicutt, K.E., Larsen, P.A., Kappeler, P.M., Rasoloarison, R.M., Ralison, J.M., Swofford, D.L., Weisrock, D.W., 2016. Geogenetic patterns in mouse lemurs (genus Microcebus) reveal the ghosts of Madagascar's forests past. PNAS 113, 8049–8056.

Yoder, A.D., Vilgalys, R., Ruvolo, M., 1996. Molecular evolutionary dynamics of cytochrome b in strepsirrhine primates: the phylogenetic significance of third-position transversions. Mol Biol Evol 13, 1339–1350.

Zhou, Y., Wang, S.-R., Ma, J.-Z., 2017. Comprehensive species set revealing the phylogeny and biogeography of Feliformia (Mammalia, Carnivora) based on mitochondrial DNA. PLOS ONE 12, e0174902.

Chapter 2

2 On the genetic consequences of habitat contraction: edge effects and habitat loss

Gabriele Maria Sgarlata[1], Tiago Maié[1], Tiago de Zoeten[1], Rita Rasteiro[1], Lounès Chikhi[1, 2]

1. Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, 2780-156 Oeiras, Portugal;
2. Laboratoire Evolution & Diversité Biologique (EDB UMR 5174), CNRS, IRD, UPS, Université de Toulouse Midi-Pyrénées, Toulouse, France;

AUTHOR CONTRIBUTIONS: GMS, LC and TM designed the project. LC, RR and TM designed the genetic simulation framework, TM and RR developed the genetic simulation framework, GMS performed the spatial genetic simulations and data analyses, and developed the numerical framework. TZ contributed to data analysis. GMS drafted the manuscript and LC revised it.

Highlights

- Habitat edges negatively influence local genetic diversity regardless of deme population size, dispersal and mutation rate, decreasing heterozygosity between 5% and 33%, and allelic richness between 17% and 33%.
- Species with high dispersal rate are more susceptible to edge-impacted changes in genetic diversity
- We propose a statistical measure called *'edge-corrected population size*' (N_{edge}) that quantifies the additive impact of multiple edges, common in natural habitat fragments, on genetic diversity
- We use a separation-of-timescales approach to model non-equilibrium changes in genetic diversity due to increase in habitat edges and reduction in habitat amount following habitat contraction

- The *'edge-corrected population size*' statistic and the equilibrium or non-equilibrium models of genetic diversity could be used to test the effect of habitat edges on the genetic diversity of natural populations.

2.1 Abstract

Natural climate change and recent anthropogenic activities have largely contributed to habitat loss and fragmentation across the world, leading to 70% of worldwide remaining forests to be within 1 km of forest's edges (Haddad et al., 2015). Ecological studies have shown that edge-effect influences ecological communities, species richness and abundance across many taxa, contributing to worldwide decline in biodiversity. Since edge-effect reduces species abundance and connectivity, it is also expected to negatively influence species genetic variation. In fact, previous theoretical studies had showed that populations closer to the edges of a finite stepping-stone model tends to have shorter coalescence times, and therefore, lower genetic diversity, than central populations. However, predicting the impact of edge effect on local genetic diversity remains challenging in realistic and more complex habitat fragments, where the additive effect of multiple edges (= multiple barriers) is expected to take place. In the present study we explore the genetic consequence of habitat loss at the scale of a habitat fragment (*patch*-scale), looking at the interplay between *patch-size* and *edge-effect* on spatial genetic diversity. We propose a statistical approach to estimate 'edge-impacted population size' from habitat cover information and use this measure to predict spatial genetic diversity in both equilibrium and non-equilibrium populations. We address these questions using spatially-explicit simulations and propose a spatially-explicit analytical framework able to predict spatio-temporal changes in genetic diversity due to edge-effect and habitat loss. An understanding of the interplay between *patch-size* and *edge-effect*, in space and time, would allow us to both predict (future) and infer (past) the genetic consequences of habitat disturbance on natural populations.

2.2 Introduction

Throughout natural history, landscapes have undergone through changes in configuration, abiotic characteristics and level of connectivity (Williams, 2003; Hewitt, 2000; Weigelt et al., 2016). Those changes have affected the distribution of species and ecosystems worldwide, at different rate and extent (Graham et al. 1996; Williams et al., 1998; Lyons 2003, 2005). More recently, human activities had dramatically modified and destroyed natural environments (Barlow et al., 2016; Song et al., 2018), with the consequent decline in global biodiversity (Pimm et al., 2014; Newbold et al., 2015; Ceballos et al., 2015, 2017). Habitat loss and fragmentation (HL&F) has been recognized as one of the major threats for biodiversity (Groombridge, 1992; Pereira et al., 2010; Tittensor et al., 2014). Haddad et al., (2015) estimated that world remnant forests are greatly affected by tree cover loss and isolation, leading to about 70% of worldwide forests being at 1 km from an edge. Vancutsem et al., (2020) have shown that 17% of the tropical moist forests have disappeared between 1990 and 2019, and predictive models suggest that by 2050 undisturbed forests will disappear entirely in large tropical humid regions.

HL&F is a process characterized by loss of habitat amount and the breaking apart of habitat in smaller patches (Curtis 1956; Moore 1962; Fahrig, 2019). The responses of biodiversity to HL&F are different at the scale of the *landscape* (multiple habitat fragments) and at the scale of the *patch* (habitat fragment) (Pardini et al., 2010; Didham et al., 2012; Fletcher et al., 2018; Fahrig, 2019). At the *landscape*-scale there are two processes that may occur: *i*) removal of habitat (habitat loss); and *ii*) an increase in the number of habitat patches within a landscape (habitat fragmentation *per se*). At this scale the variables that are quantified for describing the mechanisms underlying a biological response (ecological or genetic) to HL&F are 'habitat amount' and 'habitat configuration' (a measure of fragmentation; e.g., number of patches). At the *patch*-scale only habitat removal can occur, since by definition a habitat patch cannot be described as 'fragmented' (broken apart), as it represents the smallest unit of study in a fragmented landscape (Fahrig 2003; Fahrig 2017). Therefore, the variables that matter at the *patch*-scale are size (spatial extent of continuous habitat), isolation (distance from the nearest

patch) and edge-effect (changes in habitat quality in proximity to the boundaries between habitat and non-habitat). More generally, this means that *patch*-scale and *landscape*-scale mechanisms are not necessarily interchangeable, although related, and they may need first to be investigated separately (Ruffell et al., 2016; Fletcher et al., 2018).

Although the consequences of habitat loss and fragmentation on biodiversity have been largely investigated over the last decades (see references in Fletcher et al., 2018), some controversies still exist. There is a general agreement that habitat loss has detrimental effect on biodiversity (Brooks et al., 2002; Schipper et al., 2008; Fletcher et al., 2018; Fahrig, 2017). In contrast, the extent to which habitat fragmentation has negative effect on biodiversity is still debated (Fahrig et al., 2019; Fletcher et al., 2018). For instance, a recent review of the literature by Fahrig et al., (2017) found that ecological responses to habitat fragmentation *per se* (fragmentation independent of habitat amount) were in >70% of study cases non-significant, and, when significant usually positive. In contrast, a later study by Fletcher et al., (2018) has argued that data from literature rather support a significant negative effect of habitat fragmentation *per se* on biodiversity. This controversy revealed several limitations in HL&F research that need to be overcome in order to resolve the scientific controversy and improve our understanding of biodiversity (Fletcher et al., 2018). We mention some of these limitations listed in Fletcher et al., (2018). First, habitat loss and habitat fragmentation are often correlated processes, that is, one rarely occurs without the other. Statistical analyses that do not explicitly take into account the collinearity of such processes can draw misleading conclusions on the effect of HL&F on species ecology (Didham et al., 2012). Second, ecological responses to HL&F are difficult to generalize across taxa and ecosystems. For instance, a positive response to HL&F by a given species can actually go in favor of a detrimental effect of HL&F, if this species is generalist or invasive (Pfeifer et al., 2017; Fletcher et al., 2018). Third, habitat and non-habitat categorization need to be carried out according to the species 'perspective (Betts et al., 2014). Ultimately, HL&F processes have to be regarded as spatio-temporal processes, where different mechanisms may act depending on the spatial scale,

generating ecological responses that rely on complex temporal dynamics (Hylander and Ehrlé, 2013; Ryo et al., 2019). Therefore, reconciling the scientific disagreement requires research that goes beyond statistical and correlative approaches, and focuses more accurately on the mechanisms acting across spatial-temporal scales (Fletcher et al., 2018).

At the *landscape*-scale, Fahrig, (2017) meta-analysis suggests that habitat fragmentation *per se* can have either non-significant (>70%), positive (>20%) or negative (<10%) effect on biodiversity (species abundance, occurrence and richness). Fahrig L. concluded that perhaps, habitat fragmentation *per se* does not influence biodiversity as much as loss of habitat amount. Accordingly, data on 5675 species from eight taxonomic groups (plants, fungi, gastropods, insects, amphibians, reptiles, birds and mammals) suggest that 'habitat amount' has a stronger negative effect on ecological response (species density; number of species in a plot of fixed size) than measures of habitat fragmentation *per se* (patch size and isolation) (Watling et al., 2020). These results have been interpreted as supporting the 'Habitat Amount Hypothesis' (HAH), which states that species density increases proportionally to the amount of habitat present in the area where individuals of a given species are likely to arrive (Fahrig 2013; Martin and Fahrig, 2012). This area can extend beyond the boundaries of the habitat patch from which ecological data have been retrieved, thus suggesting that patch-size may not be a full predictor of the measured ecological response. The HAH is still debated, in part because some empirical studies have demonstrated the importance of mechanisms occurring at the *patch*-scale, such as edge-effects and patch-size, to explain ecological patterns (e.g., Hanski et al., 1995; Ries et al. 2004; Prevedello and Vieira, 2010; Hanski et al., 2013; Rybicki and Hanski, 2013; Haddad et al., 2015; Pfeifer et al., 2017; Ries et al., 2017; Haddad et al., 2017; Betts et al., 2019; Bueno and Peres 2019). On the last point, indeed Watling et al., 2020 states that HAH should be seen as a parsimonious approach, where *patch*-scale variables (patch area and isolation) are replaced "with a single measure of habitat amount" that provide "a useful approximation for describing the key landscape driver of species density" in the study area. Therefore, if HAH represents a "useful approximation" of *patch*-scale

processes occurring in a fragmented landscape, understanding processes at the *patch*-scale would i) provide accurate insights on the HL&F mechanisms impacting biodiversity; ii) clarify possible controversies; and ii) elucidate how and to which extent extrapolate *patch*-scale mechanisms at a larger spatial scale (*landscape*-scale). In the present study we investigate two of the three main mechanisms occurring at the *patch*-scale: *patch-size* (habitat amount) and *edge-effect*. The meta-analysis of Watling et al., (2020) has shown that *patch-size* does significantly explain species density response (mean effect size = 0.351 ± 0.007), although at a lower extent than *habitat amount* computed at the *landscape*-scale (mean effect size = 0.606 ± 0.007). Previous works have shown that also *edge-effect* has great influence on ecological communities (Ries et al. 2004; Fletcher, 2005; Pfeifer et al., 2017; Ries et al., 2017). For instance, Pfeifer et al., (2017) have developed a statistical approach that quantify additive multiple edge-effect from landcover data ('Edge Influence') and proposed a simple measure ('Edge Sensitivity') estimating the sensitivity of a given species to edges. Using these statistics, Pfeifer et al., (2017) assessed edge-determined changes in abundance among 1,673 vertebrate species, showing that 85% are affected either positively (46%) or negatively (39%) by edge-effect. The species negatively impacted by edge-effect have 3.7 times more probability to be listed as threatened (according to IUCN classification), whereas species with positive response are typically invasive or generalist species. These studies provided clear evidence on the importance of *patch*-scale processes on ecological response.

Species ecology has a direct impact on species genetic diversity and differentiation. Effective population density (D_e) and dispersal (σ_e) are two variables typically included in population genetic models, able to capture some of the direct consequences of species ecology on neutral genetics (reviewed in Cayuela et al., 2018). D_e and σ_e are often different from census population density (D) or dispersal (σ) (e.g., Slatkin, 1987; Broquet and Petit, 2009; Hellegren and Galtier, 2016). This is because D_e and σ_e are proxies for processes involved on the reproduction of individuals and inheritance of genes, whereas D and σ are estimates based on the actual observation of individuals, which may not all contribute equally to the genetic

pool of the population (e.g., Cayuela et al., 2018). Beyond species ecology, genetic patterns are also influenced by the molecular properties of DNA, such as mutation rate (μ) and recombination rate (r). The commonalities between genetic and ecological processes suggest that HL&F may affect genetic diversity similarly to how it affects species ecological response (e.g., abundance). On the other hand, discrepancies between ecological and genetic processes (e.g., mutation rate) could point out to possible differences on the effect of HL&F on ecological and genetic responses.

Several empirical and theoretical studies have investigated the consequences of HL&F on population genetic variation. Meta-analyses on *landscape*-scale studies around the world and across several taxonomic groups have shown consistent evidence of an overall negative effect of anthropogenic disturbance and HL&F on species genetic diversity (Lino et al., 2019; Almeida-Rocha et al., 2020; Monteiro et al., 2019; Miraldo et al., 2016; Schlaepfer et al., 2018; Vranckx et al., 2012; Rivera-Ortíz et al., 2015; Aguilar et al., 2008; Honnay & Jacquemyn, 2007), although some have observed either neutral or non-significant effect (land-use intensity and human density; Milette et al., 2020). The recent global analysis of Almeida-Rocha et al., (2020) investigated the effect of *landscape*-scale and *patch*-scale disturbance on species genetic responses. They stated that *landscape*-scale measures of habitat loss and reduction in connectivity were the factors with most detrimental effect on genetic diversity (respectively, forest cover effect size: -0.67 [-0.89, -0.44]; fragments isolation effect size: -0.65 [-1.28, -0.02]). The *patch*-scale processes with a significant negative effect on genetic diversity were 'local quality' (effect size: -0.08 [-0.38, -0.21]) and 'patch size' (effect size: -0.42 [-0.65, -0.20]), although, we could consider also 'connectivity' as a *patch*-scale process, since Almeida-Rocha et al., (2020) defined it as the level of physical isolation between habitat fragments (see definition of *patch-isolation* in Fahrig, 2019). It is worth to note that in Almeida-Rocha et al., (2020), *edge-effect* did not appear to have a significant influence on species genetic response. This is probably due to the small sample size of *edge-effect* genetic studies included in the meta-analysis (N = 1; Quevedo et al., 2013), which is probably related to the fact that no

much research has been dedicated to such process yet (see Radespiel et al., 2019 for an exception). This reflects a general tendency in landscape genetic literature, where most studies are focused on the effect of HL&F on population genetic differentiation at the *landscape*-scale, and very few on population genetic diversity at the *patch*-scale (Storfer et al., 2010; DiLeo and Wagner, 2016).

Theoretical and simulation-based studies on the genetic consequences of HL&F are also skewed towards patterns of population genetic differentiation at the *landscape*-scale (Cushman et al. 2012; Mona et al., 2014; Jackson and Fahrig, 2014; Jackson and Fahrig, 2016). However, there are some exceptions. The earlier theoretical studies investigating the impact of *edge-effect* on genetic diversity date back to the work of Maruyama, (1972) and Tajima, (1990), who performed numerical computations of the expected amount of genetic diversity in several structured models. They both showed that lower dispersal rate generates lower genetic diversity, and assuming that edge populations experience lower dispersal rate than central populations, then it is expected that edge populations would have lower genetic diversity (Tajima, 1990). Later theoretical works have confirmed Tajima's predictions, showing that populations closer to the edges of a finite habitat (stepping-stone model) tends to have shorter coalescence times, and therefore, lower genetic diversity (Hey, 1991; Herbots, 1994; Wilkins and Wakeley, 2002; Wilkins, 2004). In particular, Wilkins (2004) has derived explicit expressions of the distribution of coalescence times for edge-impacted populations in a rectangular habitat. Following theoretical studies have also pointed out the impact of habitat boundaries on genetic variation (e.g., Leblois et al., 2006; Duforet-Frebourg and Slatkin, 2016). Leblois et al., (2006) performed spatially-explicit simulations to study the effect of habitat contraction on population genetic diversity. They observed that samples collected across the whole habitat were showing lower observed heterozygosity than samples collected exclusively from the centre. They imputed these results to the effect of edges at the boundaries of the simulated scenario. Recently, Ringbauer et al., (2017) have proposed a framework that model the effect of a linear barrier on the distribution of coalescence times, on the probability of identity-by-descent, and therefore on genetic diversity. The barrier would be

equivalent to the edges of a habitat patch. However, similarly to Wilkins, (2004), the use of such model might become computationally expensive in realistic and more complex habitat fragments, where the additive effect of multiple edges (= multiple barriers) is expected to take place (Pfeifer et al., 2017).

In the present study we explore the genetic consequence of habitat loss at the *patch*-scale, looking at the interplay between *patch-size* and *edge-effect* on spatial genetic diversity. We propose a statistical approach to estimate 'edge-impacted population size' from habitat cover information (proxy for D_e), and use this measure to predict spatial genetic diversity in both equilibrium and non-equilibrium populations. Our approach was in part inspired by the ecological measure of 'Edge Influence' introduced by Pfeifer et al., (2017), and by previous works on landscape genetics that uses 'neighbourhood-level' (or 'node-based') analysis to correlate habitat amount with genetic diversity (e.g., Keyghobadi et al., 2005). Pfeifer et al., (2017) results show that, across many taxa, 'Edge-Influence' statistics correlates with species abundance (ecological response), but remain currently unknown the performance of this approach on predictions of edge-determined changes in genetic diversity (genetic response). In conclusion, this work is set within the context of a true integration of population genetic and landscape ecology methods for landscape genetics, since often these disciplines use similar concept but little theoretical overlap (Storfer et al., 2006; Wagner and Fortin, 2013).

2.3 Material and Methods

2.3.1 Background

Pfeifer et al., (2017) have developed a metric called 'Edge Influence' (*I*) for quantifying the cumulative effects of multiple edges, including edge shape and patch-size, from maps of habitat cover (e.g., percentage tree cover). This measure provides quantitative information of *I* for a focal point in the map, and thus allow to relate 'edge-influence' with some other variable of interest, such as species abundance. It is defined as:

$$I = \max(\sigma_c, |\bar{C} - C|) \times sign(\bar{C} - C) \qquad \text{(Eq. 1)}$$

where C is the habitat cover at the focal point and, $\bar{C}$ and σ_c are, respectively, the average and standard deviation of habitat cover within an area A around the focal point. In Pfeifer et al., (2017), $\bar{C}$ and σ_c are estimated using a Gaussian filter, that is by weighting the contribution on $\bar{C}$ and σ_c of each data point within A by their distance from the focal point. The Gaussian filter weights each data point according to a Gaussian distribution:

$$G(x, y) = \frac{1}{2\pi r^2} e^{-\frac{x^2+y^2}{2r^2}} \qquad \text{(Eq. 2)}$$

where x and y are the distances, respectively, on the x and y direction from the focal point, and r the radius of the area A. Note that the Gaussian weight depends only on geographical distance and radius r. The value of I corresponds to the maximum between the regional (σ_c) and the point heterogeneity ($|\bar{C} - C|$), with a sign given by $\bar{C} - C$, which is negative if the focal point is near the matrix (non-habitat) and tends to zero while moving away from the matrix, toward the interior of the habitat.

2.3.2 Edge-corrected Effective Population Size

Similar to Pfeifer et al., (2017), we propose an approach that can quantify the additive effect of multiple edges at each point in space within a 2D stepping stone model with irregular shape. Such analysis provides an 'edge-influence index' for each deme in the grid, that is used to correct local carrying capacity and obtain an estimator of 'edge-corrected population size' (N_{edge}). The 'edge-influence index' can be interpreted as a measure of connectivity within habitat patch, assuming that proximity to habitat edges determines the extent to which each deme can receive alleles from any direction in the 2D space. Deme 'edge-influence index' was calculated from a grid representing the shape and the habitat structure (carrying capacity) of the patch. Carrying capacity information within habitat patch are considered to be a proxy for local census population size of the studied species, and could be retrieved, for instance, from data on habitat cover (e.g., percentage of tree cover, NDVI). For each deme, we computed the weighted mean of deme population size from the total number of demes included within a square matrix with side length '*R*' centred in the focal deme. The weights of the mean were defined using the correlation coefficient of allele frequencies between pairs of demes at a given distance '*k*' as detailed in Kimura and Weiss, (1964):

$$r(k) \sim \frac{e^{-\sqrt{\frac{4\mu}{m}}k}}{\sqrt{k}} \qquad (Eq.\ 3)$$

where μ is the per site mutation rate and m is the deme dispersal rate.

Thus the 'edge-influence index' of a deme with spatial coordinates 'x, y' ($E_{x,y}$) is computed as:

$$E_{x,y} = \frac{\sum_{i=1}^{n} N_i * r(k)}{\sum_{i=1}^{n} N_i} \qquad (Eq.\ 4)$$

where N_i is the population size of deme i and n is the total number of demes that fall within the square matrix area with side length '*R*', centred in 'x, y'. The smallest R that maximises the fitting between genetic diversity and 'edge-corrected population size' is selected as being the optimal radius, since it limits computational cost while being sufficiently accurate (see *Results*). The 'edge-influence index' for each deme is then normalized with respect to the 'edge-influence index' value in a 2D homogenous habitat (equal N_i across demes) with no boundaries ($\hat{E}$), $\tilde{E}_{x,y} = E_{x,y}/\hat{E}$. We then obtain the 'edge-corrected population size' of each deme in the habitat by $N_{edge} = N_i * \tilde{E}_{x,y}$. A graphical representation of this approach is presented in Fig. S1.

2.3.3 Two-dimensional spatial model at equilibrium

Similar to Barton et al., (2013), we model the observed heterozygosity (H_0) within a sub-population (deme) in a 2D stepping-stone by using the probability of identity-by-descent of two alleles randomly sampled within the same sub-population (F_0) :

$$1 - H_0(t) = f_0(t) + \big(1 - f_0(t)\big)(1 - H_{t*}) \qquad (Eq.\ 5)$$

where $H_0(t)$ is the observed heterozygosity at time t, $f_0(t)$ is the probability of identity-by-descent at time t and H_{t*} is the heterozygosity at a given time t^*, a timepoint beyond which geography (sampling location) is assumed to have little or no impact on the distribution of lineages. In other words, *Eq. 5* states that two alleles randomly sampled within subpopulations can be homozygous either due to identity-by-descent (first term) or by chance (second term). Given a sampling area A which diameter is much lower than $\sigma/\sqrt{\mu^{-1}}$, where σ is the dispersal distance of a studied species and μ is the per site mutation rate, Barton et al., (2013) define t^* as a time satisfying

$\mathrm{Diameter(A)}^2/\sigma^2 \ll t^* \ll \mu^{-1}$. Assuming a low mutation rate, in the time window between the present and t^* in the past, no new mutations have appeared and a considerable proportion of alleles are very closely related ('fast local coalescence') (Barton et al., 2013). As a consequence, the genetic diversity observed at this spatial and temporal scale refers to mutations that occurred more than t^* generations ago. Similarly, Wilkins (2004) defines a 'transition time' τ between two 'phases' of the coalescence process in spatially structured populations: '*scattering-phase*' (dependent on sampling location) and '*collecting-phase*' (independent from sampling location), respectively. Wilkins (2004) results were based on previous studies showing that the genealogical process in structured population can be modelled using the 'separation-of-timescales' approach (Wakeley, 1999; Wilkins and Wakeley, 2002). According to this approach, the genealogy of genes in a structured model can be divided in two time 'timescales': '*scattering phase*' and '*collecting phase*'. The '*scattering phase*' corresponds to the distribution of coalescence times in the recent past, which is dependent on the original sampling location and on the details of the structured model (dispersal rate, density, drift). In a spatially structured model, it can be modelled as a random walk of lineages, which coalesce once they are in the same deme, at a rate given by the inverse of the deme population size. The '*collecting phase*' corresponds to the distribution of coalescence times in the ancient past, which is independent from the original location of the sampled genes. In this phase, coalescence rate is dependent on the total population size of the 2D stepping-stone, given that it mostly includes lineages that have crossed the population multiple times since the time to their most common ancestor (Barton and Wilson, 1995). Referring to *Eq. 5*, the '*scattering phase*' is modelled by the probability of identity-by-descent (first term), whereas the '*collecting phase*' is represented by the heterozygosity at time t^* (second term).

At mutation-drift equilibrium ($t \to \infty$), the probability of identity-by-descent of two alleles randomly sampled within the same deme, can be

written using the Wright-Malécot approximation (Wright 1943; Malécot 1948; Barton et al., 2002):

$$F_{0(t=\infty)} \approx \frac{\ln\left(\frac{l_c}{K}\right)}{ln\left(\frac{l_c}{K}\right) + 2\pi N_e m} \qquad (Eq.\ 6)$$

where $l_c = \sqrt{m/4\mu}$ represents the balance between mutation and dispersal, m is the deme dispersal rate in a 2D stepping stone model and K is the 'local scale' constant, which in the case of a 2D stepping stone model is equal to $\sqrt{1/32}$. If we assume a dispersal distribution of Gaussian form (as it might be the case to assume with real data), m can be replaced by $2\sigma^2$ and $K = \sigma$.

As mentioned above, the heterozygosity at time t^* is determined by lineages coalescing more than t^* generations ago ('*collecting-phase*') and that had enough time to cross the whole population (Barton and Wilson, 1995). It has been shown that, under a range of parameter values, the genealogy of most lineages in the '*collecting-phase*' can be approximated by the coalescence process in a panmictic population, which size is equal to the total population size in a 2D stepping stone model (Cox and Durrett, 2002; Wilkins, 2004). This measure relates also to the average coalescence time of two alleles randomly sampled within the same deme (Strobeck 1987; Slatkin, 1991; Hey, 1991). Therefore, we can use the expression for the expected coalescence time of a panmictic population of size equal to the whole population to derive H_{t*}. Since in the present study we are measuring heterozygosity using microsatellite data, H_{t*} is written as the expected heterozygosity at mutation-drift equilibrium under a Stepwise Mutation Model (microsatellite) (Kimura and Ohta, 1973; Ohta and Kimura, 1973):

$$H_{t*} \approx 1 - \frac{1}{\sqrt{1 + 2\theta_0}} \qquad (Eq.\ 7)$$

where $\theta_0 \approx 2\bar{t_0}\mu$ is the expected number of differences between sampled alleles (or population mutation rate) under the infinite site model of mutation and $\bar{t_0}$ is the expected coalescent times of two alleles sampled from a panmictic population with size equal to the 2D stepping stone model. The expected coalescence time $\bar{t_0}$ can be written as $\bar{t_0} = 2N_e d_{tot}$. where N_e is the deme effective population size and d_{tot} is the total number of demes in the population.

So far, these expressions assume homogeneous deme effective population size across the 2D stepping-stone. To model the consequences of edge-effect on deme heterozygosity across a habitat fragment of irregular shape, we compute the edge-corrected probability of identity-by-descent at mutation-drift equilibrium for each deme in the grid by using in *Eq. 6* the estimated N_{edge} defined in the section above. Then, by rearranging *Eq. 5*, and using *Eq. 6* and *Eq. 7*, we can write the edge-corrected observed heterozygosity as:

$$H_{x,y} \approx \left[1 - \frac{1}{1 + \frac{2\pi N_{edge} m}{\ln\left(\frac{l_c}{K}\right)}}\right] H_{t*} \qquad (Eq.\ 8)$$

2.3.4 A statistical approach for testing molecular edge-effect

As an alternative approach to the spatial model presented above, we used a generalized additive model (GAM) to quantify the impact of the 'edge-corrected population size' on the observed heterozygosity. The GAM model was chosen because of its flexibility for both linear and non- linear relationships between predictor and response variable, and its capacity to capture patterns that classical linear models would miss (Hastie and Tibshirani, 1987). In particular, it estimates an unknown 'smooth' function of some predictor variable using non-parametric approaches to maximize the

fitting of the data. We used the *gam* function from the *mgcv* R package (Wood, 2011), setting REML as estimation method for the smooth parameter. We quantified the predictive power of the 'edge-corrected population size' by i) measuring the total 'Deviance explained by the model', which represents the proportion of the total differences between observed and predicted values; or ii) by quantifying the Normalized Root-Mean-Square Error (NRMSE) between predicted and observed values. NRMSE measures the average squared difference between estimated and observed values, normalized by the range of the observed data: $NRMSE = \sqrt{\sum_{i=1}^{N}\left(\widehat{H}_{x_i,y_i} - H_{x_i,y_i}\right)^2 / N} \Big/ \left(\max_{i \in N} H_{x_i,y_i} - \min_{i \in N} H_{x_i,y_i}\right)$. We computed NRMSE for both spatial and GAM models.

2.3.5 Two-dimensional spatial model at non-equilibrium

We use coalescent theory in a 2D stepping-stone model to describe the temporal dynamic of deme heterozygosity in habitat patches undergoing contraction. Our model considers a 2D stepping-stone model that has been evolving for τ_2 generations. If we consider the present time as $t = 0$, the individuals forming the initial population τ_2 generations ago are randomly sampled from a panmictic model with size N_0 and heterozygosity given by *Eq. 7*. Like the radiation model described in Slatkin (1993), our analytical model represents an instantaneous range expansion in a 2D stepping-stone model. The spatially-structured population is subjected to habitat contraction at time τ_1 in the past ($\tau_1 < \tau_2$), during which, as in real study cases, suitable demes are removed from the periphery of the habitat patch. The reduction in *patch-size* has two important consequences expected to influence deme heterozygosity: *i*) decrease in the total amount of habitat, due to loss of suitable habitat at the periphery; and *ii*) increase in edge-influence, caused by the increased proximity of demes to the habitat boundaries. Habitat loss is

modelled by considering the total population size before $N_{\tau_2} d_{tot(\tau_2)}$ and after $N_{\tau_1} d_{tot(\tau_1)}$ contraction. Changes in edge-influence for a given deme i are parametrized by two N_{edge} values, before $N_{edge(i,\tau_2)}$ and after $N_{edge(i,\tau_1)}$ contraction.

Using the cumulative density function (CDF) of probability of identity-by-descent for two alleles sampled from the same deme (Wilkins, 2004), we can then derive the expression for deme heterozygosity at each point in space and time. By rearranging *Eq. 5*, we obtain:

$$H_0(t) = \left(1 - \int_0^{\tau_2} f_0(t)\, dt\right) H_{t*} \qquad (Eq.\ 9)$$

In this expression the 'separation-of-timescales' is considered by modelling the '*scattering-phase*' with $\int_0^{\tau_2} f_0(t)\, dt$ and the '*collecting phase*' with H_{t*}. The '*scattering-phase*' can be decomposed in two components: the probability of identity-by-descent for two alleles coalescing before τ_1 (time of contraction) and probability of identity-by-descent for two alleles coalescing between τ_1 and τ_2 (time of instantaneous expansion):

Pr(coal. by τ_2 & no mutation) = Pr(coal. by τ_1 & no mutation) + Pr($\tau_1 <$ coal. $\leq \tau_2$ & no mutation); (*Eq. 10a*)

which can be written as:

Pr(coal. by τ_1 & no mutation) = $\int_0^{\tau_1} P\left(T_2 = t \middle| N_{edge(i,\tau_1)}\right) e^{-2\mu t}\, dt$; (*Eq. 10b*)

Pr($\tau_1 <$ coal. $\leq \tau_2$ & no mutation) = $\int_{\tau_1}^{\tau_2} P\left(T_2 = t \middle| N_{edge(i,\tau_2)}\right) e^{-2\mu t}\, dt$; (*Eq. 10c*)

where $P\left(T_2 = t \middle| N_{edge(i,\tau_2)}\right)$ or $P\left(T_2 = t \middle| N_{edge(i,\tau_1)}\right)$ is the probability of two alleles sampled from the same deme of coalescing at time t, given a deme population size of $N_{edge(i,\tau_2)}$ or $N_{edge(i,\tau_1)}$, respectively, before or after habitat contraction; $e^{-2\mu t}$ is the probability of no mutation since coalescence time t. Then, the CDF of probability of identity-by-descent for two alleles sampled from the same deme at time t_0 is:

$$F_0(t_0) = \int_0^{\tau_1} P\left(T_2 = t \middle| N_{edge(i,\tau_1)}\right) e^{-2\mu t}\, dt + \int_{\tau_1}^{\tau_2} P\left(T_2 = t \middle| N_{edge(i,\tau_2)}\right) e^{-2\mu t}\, dt; \qquad (\textit{Eq. 10d})$$

We write the probability that two lineages coalesce at time t in the past $P(T_2 = t)$ using *Equation A7* from Wilkins (2004), which is an approximation of the exact probability distribution in a 2D stepping-stone model (Barton and Wilson 1995, 1996). This approximation allows easier and faster computation while being sufficiently accurate:

$$P(T_2 = t) = \frac{1}{Nb\ t}\left(1 - \frac{1}{1 + (1/Nb)}\left(\frac{2Log(t - 1/2) + 2\gamma}{Nb} + \frac{i\, e^{-\sum_{k=1}^{i-3} k^2/t}\, Log^{i-1}(t + 3/2 - i)}{Nb^{i-1}}\right)\right) \qquad (\textit{Eq. 11a})$$

where γ is the Euler constant (~ 0.5772), Nb is the neighbourhood size ($2N_e m\pi$) and i is the number of additional terms for the approximation form, which are added up from three through forty-four with alternate sign. For instance:

$$P(T_2 = t) = \frac{1}{Nb\,t}\left(1 - \frac{1}{1+(1/Nb)}\left(\frac{2Log(t-1/2)+2\gamma}{Nb} - \frac{3\,Log^2(t-3/2)}{Nb^2}\right)\right) \quad \text{for } t = 3; \qquad (Eq.11b)$$

$$P(T_2 = t) = \frac{1}{Nb\,t}\left(1 - \frac{1}{1+(1/Nb)}\left(\frac{2Log(t-1/2)+2\gamma}{Nb} - \frac{3\,Log^2(t-3/2)}{Nb^2} + \frac{4\,e^{-1/t}\,Log^3(t-5/2)}{Nb^3}\right)\right) \text{ for } t = 4; \qquad (Eq.\ 11c)$$

$$P(T_2 = t) = \frac{1}{Nb\,t}\left(1 - \frac{1}{1+(1/Nb)}\left(\frac{2Log(t-1/2)+2\gamma}{Nb} - \frac{3\,Log^2(t-3/2)}{Nb^2} + \frac{4\,e^{-1/t}\,Log^3(t-5/2)}{Nb^3} - \frac{5\,e^{-5/t}\,Log^4(t-7/2)}{Nb^4}\right)\right) \text{ for } t = 5; \qquad (Eq.\ 11d)$$

For $t \geq 44$, all additional terms are included. *Eq. 11* can be substituted in *Eq. 10*, by using either $N_{edge(i,\tau_2)}$ or $N_{edge(i,\tau_1)}$.

Wilkins, (2004) has shown that the '*collecting phase*', under certain conditions, can be approximated by the standard neutral coalescent model (SNCM; Kingman, 1982a, b). As in Barton et al., (2013) and in agreement with the definition of '*collecting phase*' by Wakeley, (1999), H_{t*} represents the probability that two lineages are heterozygous, independently from geography since these two lineages have most likely crossed the whole habitat patch multiple times. Since we are modelling habitat contraction as a reduction in total patch size and given that the '*collecting phase*' can be approximated by SNCM, H_{t*} is derived from the expression of the expected heterozygosity for a panmictic population in a varying environment (Griffiths and Tavaré, 1994). In particular, since genetic diversity at time t is influenced by past events that decreased significantly the patch population size, we used

the harmonic mean of the varying population size over time (Griffiths and Tavaré, 1994) as effective population size from which estimate H_{t*} by using *Eq. 7*:

$$N_e d_{tot} = \frac{1}{\frac{1}{\tau}\sum_{i=1}^{\tau}\frac{1}{N_i d_{tot(i)}}} \qquad (Eq.\ 12)$$

A graphical representation of the non-equilibrium model is presented in Fig. S2.

2.3.6 Population genetics simulations

We used the computer program SINS (Simulating INdividuals In Space; Maié et al., 2021; Rasteiro et al., 2012) to simulate habitat patches undergoing contraction. SINS is an individual-based software that uses a forward-time approach to simulate diploid individuals (males and females) over a 2D grid of demes, similar to a classical 2D stepping-stone model (Kimura and Weiss, 1964) as in SPLATCHE software (Currat et al., 2004; Currat et al., 2019). Deme population size ($\widehat{N}_{i,t+1}$) is drawn from a Poisson distribution with a mean value ($N_{i,t+1}$) given by a corrected form of the Maynard-Smith and Slatkin, (1973) logistic growth (Rasteiro et al., 2012; Maié et al., 2021). The $N_{i,t+1}$ value in each deme is logistically regulated by deme-specific carrying capacity (K), growth rate (r) and deme number of reproductive females ($N_{f,t}$):

$$N_{i,t+1} = 2N_{f,t}\,\frac{1+r}{1+r\,\frac{2N_{f,t}}{K}} \qquad (Eq.\ 13)$$

Mating is random within demes, namely one reproductive individual from each sex is randomly chosen and generate a new individual (offspring) which inherit, for each locus, one allele from each parent. At each generation, mating occurs iteratively until $\widehat{N}_{i,t}$ is reached. Each deme exchanges migrants with the four neighbouring demes. The number of individuals that will emigrate from the focal deme at each generation is drawn from a Poisson distribution with mean $M = N_{i,t} m n_d / 4$, where m is the dispersal rate and n_d is the number of available neighbouring demes, which is equal to 2 for a corner deme and 4 for central deme. The number of migrants will then be stochastically distributed among the neighbouring demes using a binomial distribution with probability $P_{dir} = (1 - F_{dir})/(n_d - F_t)$, where dir is one of the available directions (min=1; max=4), F_{dir} is a deme-specific friction parameter defining the difficulty to move from the focal deme into that direction and F_t is the sum of the frictions of the $n_d$receiving demes. SINS is able to simultaneously simulate demographic and genetic data over non-overlapping discrete generations and through space. Several types of molecular markers (sequences, SNPs and microsatellites) and chromosome objects (X and Y chromosomes, autosomes and mitochondrial DNA) can be simulated. In the present study we will exclusively use independent microsatellite loci, which mutations are modelled according to the Stepwise Mutation Model (Ohta and Kimura, 1973). Similar to SPLATCHE (Currat, Ray and Excofier, 2004; Currat et al., 2019), changes from one habitat configuration to another (i.e., habitat contraction) are defined by K and F maps describing, respectively, the carrying capacity and friction value of each deme in the two configurations. Demes that are unsuitable after an 'environmental event' are set with K = 0 and F = -99, such that no individuals can grow and no migrants can move within these unsuitable demes.

Simulations start with all demes having the same initial genetic diversity, which is defined by a file for each locus specifying the initial number of alleles and frequency of the simulated markers. In this study, we used initial

allele frequency for each locus by sampling 100 individuals from a Wright-Fisher (WF) population simulated in *fastsimcoal* (Excoffier and Foll, 2011) using the *fastsimcoal* function in the *strataG* R package (Archer, Adams and Schneiders, 2017). The size of the WF population was set equal to 10,000 and mutation rate according to the value specified in the SINS simulation.

2.3.7 Simulated scenarios

Equilibrium: We simulated mainly two habitat configurations, i) a 9 x 9 squared grid (d_{tot} = 81 demes) and ii) a 23 x 23 irregular grid (d_{tot} = 382 demes) (Fig. 1a, b). Three other habitat configurations of 23 x 23 demes were also simulated to show consistency of our results across habitats of different shapes (Fig. S6). For each habitat configuration we tested several values of K (50, 100, 200, 300) and m (0.02, 0.04, 0.06, 0.08, 0.1), which were kept constant across all demes in the habitat. We also tested several values of mutation rate, μ (5 x 10^{-3}, 5 x 10^{-4}, 5 x 10^{-5}, 5 x 10^{-6}, 5 x 10^{-8}) and number of genetic markers (15, 30, 100, 400 microsatellites). The simulations were run for over 10,000 generations, such to ensure that the whole population had reached a 'quasi-genetic equilibrium' (i.e., $\Delta He/t \approx 0$). Demographic data (deme population size) were recorded every 50 generations, except when specified otherwise.

Non-Equilibrium: We simulated a 13 x 13 squared grid (d_{tot} = 169 demes) that undergoes habitat contraction after 10,000 generations. Habitat contraction was modelled as an instantaneous removal of the outer demes from the 13 x 13 grid, leading to a 9 x 9 squared grid (Fig. 1c). After habitat contraction, simulations were run for more 4,000 generations. Because of the computational cost of the simulations, most *non-equilibrium* scenarios were simulated only under the 13 x 13 squared grid configuration. However, we also tested the effect of habitat contraction on grid configurations with larger number of demes and more complex shape (Fig. S6). As for the *Equilibrium*

section, we investigated the impact of K (50, 100, 200, 300), m (0.02, 0.04, 0.06, 0.08, 0.1) and μ (5 x 10^{-3}, 5 x 10^{-4}, 5 x 10^{-5}, 5 x 10^{-6}, 5 x 10^{-8}) on pattern of genetic diversity, in space and time. Across all *non-equilibrium* simulations, we simulated thirty independent microsatellite loci and sampled fourteen individuals per deme every 50 generations.

For both *Equilibrium* and *Non-Equilibrium* settings, we ran 30 independent replicates for each simulated scenario.

2.3.8 Genetic diversity

Measures of genetic diversity were estimated in R (R Core Team, 2015) using the *summary* function of the *adegenet* R package (Jombart, 2008; Jombart and Ahmed, 2011). We computed three summary statistics typically used in population genetic studies: observed heterozygosity (H_O), expected heterozygosity (H_{exp}) and mean number of alleles (*MNA*), which here is equivalent to allelic richness (A_r) since sample size is the same across demes (Allendorf and Luikart, 2009). Diversity was estimated for each independent genetic marker and then averaged across loci. Individual-based heterozygosity (H_eL) was derived by the 'Homozygosity by loci' index proposed by Aparicio et al., 2006 and implemented in the *Rhh* R package (Alho et al., 2010). 'Heterozygosity by loci' can be computed as $H_eL = 1 - HL = 1 - \sum E_h / \left(\sum E_h + \sum E_j\right)$, where E_h and E_j are the expected heterozygosities of the loci that an individual has in homozygosity (*h*; the two alleles are equal) and in heterozygosity (*j*; the two alleles are different). This index weights locus homozygosity by the level of genetic variability at that locus in the population, such that a locus will contribute more to HL if its allele distribution is less skewed (more uniform) and more variable (multiple-alleles). We chose HL since it appeared to correlate better with genome-wide homozygosity and inbreeding coefficients in natural populations (Aparicio et al., 2006).

2.3.9 Demographic analysis

During the spatio-temporal simulation, census population size was recorded for each deme in the habitat. Data were retrieved every 50 generations for the whole simulation, and the harmonic mean of the population size over time was calculated as in *Eq. 12*, but replacing $N_i d_{tot(i)}$ with $N_{i,t}$, the census population size of deme i at generation t. The harmonic mean of the population size can be used to represent the rate of genetic drift of a given population, when population size varies over time due, for instance, to environmental oscillations (Griffiths and Tavaré, 1994).

2.4 Results

2.4.1 Equilibrium

We modelled a 2D stepping stone with a total of 81 demes on a 9 x 9 squared grid under homogenous carrying capacity and dispersal rate (Fig. 1a). After 10,000 generations, we sampled 5 individuals from each deme located along the diagonal of the simulated grid. Thirty microsatellites per individuals were 'genotyped' and used for estimating deme genetic diversity. In Fig. 2, we show the spatial distribution of deme observed heterozygosity (H_O), assessing the impact of population size, dispersal rate and mutation rate on this measure of genetic diversity. We observed a significant decrease in H_O in demes located on the corner of the habitat (x = 0 and x = 8) compared to central demes (x = 4). Differences in H_O between central and corner demes were stronger in scenarios with low population size (K = 50, $\Delta H_0 = \sim 0.20$; Fig. 2a), low dispersal rate (m = 0.02, $\Delta H_0 = \sim 0.15$; Fig. 2b) and mutation rate ($10^{-4} \leq \mu \leq 10^{-5}$, $\Delta H_0 = 0.2$; Fig. 2c). At these parameter values, we also observed a stronger loss of H_O relative to the maximum level of H_O measured in the habitat (central demes) $\Delta H_0 / \max_{i \in n} H_{0(i)}$: ~ 30% for K = 50, ~ 25% for m = 0.02 and ~ 30% for $\mu = 10^{-4}$ (Fig. S4). While results on expected heterozygosity are consistent with those on observed heterozygosity, MNA shows rather different patterns (Fig. S3 and S4). In particular, it is observed that the extent on which central and corner demes differ in diversity is no longer dependent on population size nor on dispersal rate, whereas is still dependent on mutation rate (lower row in Fig. S3).

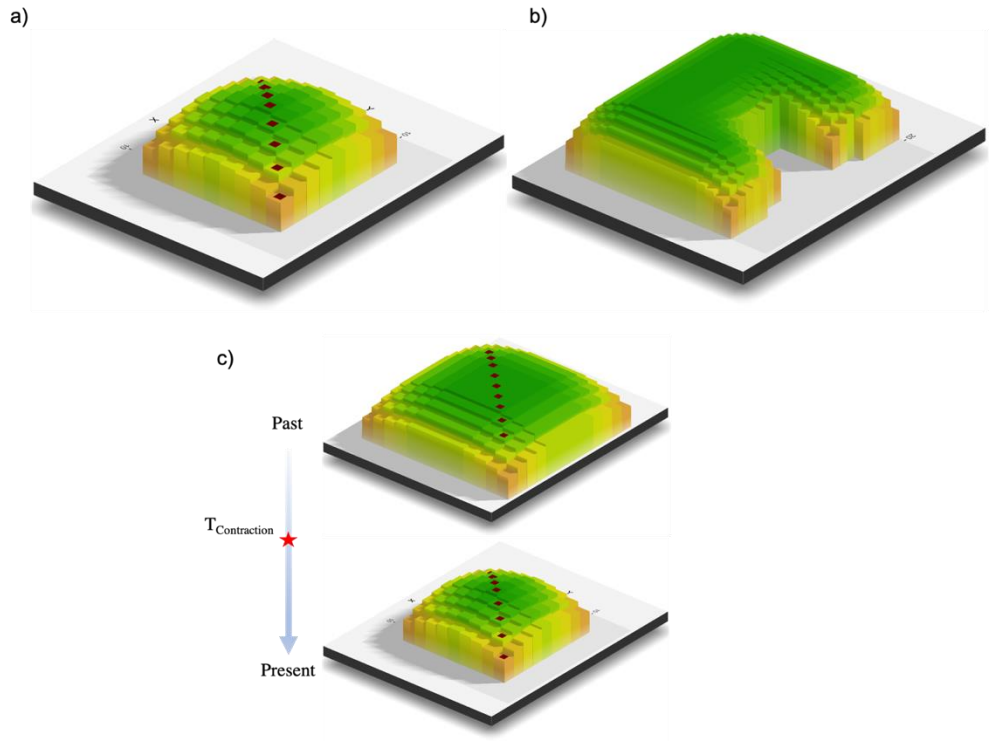


Figure 1. Habitat configurations modelled in this study. a) 9 x 9 square grid of demes (d_{tot}=81); **b)** 23 x 23 grid of irregular shape (d_{tot}=382); **c)** instantaneous habitat contraction from a 13 x 13 (d_{tot}=169) to a 9 x 9 grid of demes (d_{tot}=81). Within each habitat configuration, demes have homogenous carrying capacity and dispersal rate. Colours and deme bars are proportional to the amount of edge-effect at which each deme is subjected to. Brown squares indicate sampled demes.

We tested whether differences in genetic diversity between corner and central demes of the same scenario were attributable to differences in population size (due to demographic stochasticity). Since SINS software simulate population size as a stochastic process, thus allowing some variation in $\widehat{N}_{i,t}$ over time, we computed the harmonic mean of the recorded population sizes (HMP) for each deme in the habitat. Fig. S5 shows that HMP remain largely unchanged across the whole habitat, an exception being the simulation performed with K = 50 and m = 0.02 (Fig. S5a), where corner demes experience lower HMP than central demes ($\overline{HMP}_{corner}$: 40,

$\overline{HMP}_{central}$: 45), with large fluctuations across simulation replicates (SD_{corner}: 15, $SD_{central}$: 8.5). Note also that HMP was lower than the carrying capacity set at each scenario, approaching to the expected value in simulations with higher carrying capacity.

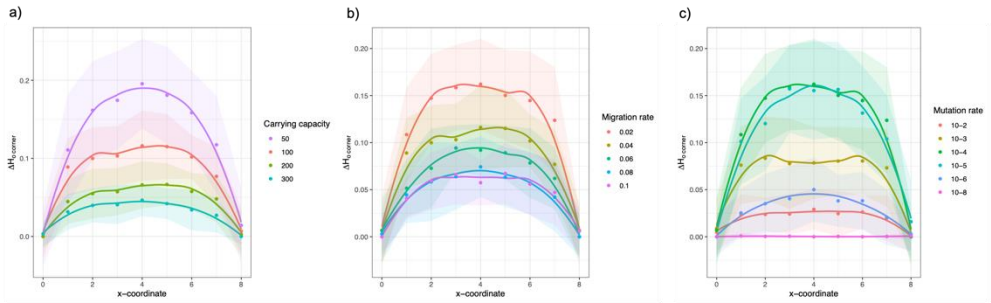


Figure 2. Spatial distribution of differences in observed heterozygosity (H_0) relative to corner demes ($\Delta H_{o\ corner}$) along the diagonal of a 9 x 9 stepping-stone model. Impact of **a)** carrying capacity (fixed parameters, m: 0.04; μ: 5 x 10^{-4}); **b)** migration rate (fixed parameters, K: 100; μ: 5 x 10^{-4}); and **c)** mutation rate (fixed parameters, K: 100; m: 0.02). Central demes (x = 4) show significantly higher H_o with respect to corner demes (x = 0 and x = 8) at lower carrying capacity (K: 50; $\Delta H_{o\ corner}$: ~ 0.20), lower migration rate (m: 0.02; $\Delta H_{o\ corner}$: ~ 0.15) and intermediate mutation rate (μ: 5 x 10^{-4} / 5 x 10^{-5}; $\Delta H_{o\ corner}$: ~ 0.20).

The results on the 9 x 9 grid provide a concise overview of how edge-effects can influence the spatial distribution of genetic diversity in a finite two-dimensional habitat. Studies that have investigated the influence of proximity to edges on some biological response (ecological or genetic), typically use discrete and simplified definition of edge-effect, such as 'distance to the nearest-edge' (see Pfeifer et al., 2017 otherwise). Such approaches do not incorporate the cumulative effect of multiple edges, which is relevant in natural habitat given their common irregular shape. In the present study, we have developed a statistical approach that quantify the additive effect of multiple edges on demes genetic diversity in topologically complex 2D habitat (see *Material and Methods*), similarly to Pfeifer et al., (2017) work, and provided a link with spatial population genetics theory. Using (*Eq. 8*), we

assessed the performance of 'edge-corrected population size' in predicting the spatial distribution of observed heterozygosity in the habitat of Fig. 1b. In Fig. 3, we compare demes observed heterozygosity with its expected value according to the Wright-Malécot approximation (*Eq. 8*), using as N_e the proposed 'edge-corrected population size' (N_{edge}). As an alternative approach, we perform model-fitting under a GAM model using the 'edge-corrected population size' as predictive variable and assess how each independent simulation replicate deviates from the fitted model. Both approaches suggest little deviation between observed and expected values, regardless which population size, dispersal rate or mutation rate is simulated. Results were consistent when simulating other habitat shapes (Fig. S6 and Fig. S7). Note that across all simulated shapes, we observed a strong deviation between theoretical predictions and simulated data when K: 50 and m: 0.02 (first row, Fig. S8). Accordingly, demographic data shows large variability in deme population size at these parameter values, and a lower HMP compared to higher dispersal rates (second row, Fig. S8).

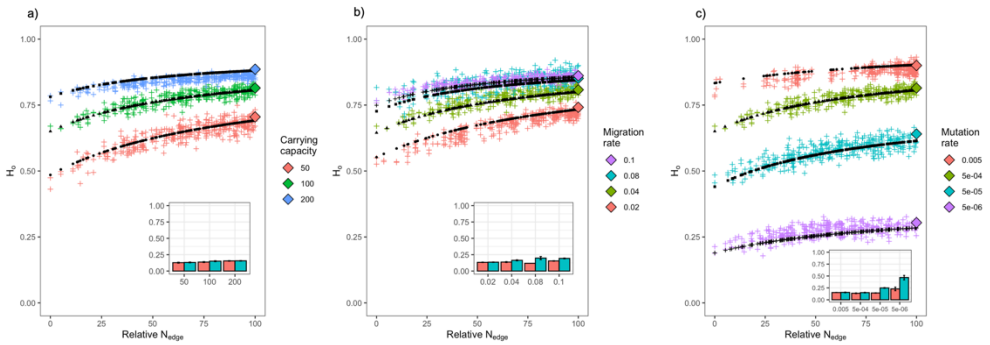


Figure 3. Comparison between edge-corrected theoretical results and simulations on a 23 x 23 stepping-stone model of irregular shape. Impact of **a)** carrying capacity (fixed parameters, m: 0.04; μ: 5 x 10^{-4}); **b)** migration rate (fixed parameters, K: 100; μ: 5 x 10^{-4}); and **c)** mutation rate (fixed parameters, K: 100; m: 0.04) on the predictability of edge-impacted changes in genetic diversity. Each cross represents the average deme H_o across 30 simulation replicates and 30 microsatellites. Relative N_{edge} is a standardized edge-corrected population size (between 0 and 100) for visualization purposes. Black dots define the theoretical predictions using N_{edge} in the Wright-Malécot approximation (*Eq. 8*). The insert bar plots

quantify the NRMSE under the Wright-Malécot approximation (light blue) and GAM model (red). Colored diamonds indicate the expected H_o using the Wright-Malécot approximation and ignoring edge-effect. Under the tested combination of parameters, the results suggest good agreement between the theoretical predictions and simulation results.

One of the parameters that needs to be chosen for quantifying edge-effect is the side length of the area from which habitat information are used to compute the weighted average deme populations size. We show that the side length value can be easily selected and it has little effect on the performance of the statistics (Fig. S9). This is because the mean is weighted using a correlation measure (*Eq. 3*) that decays quickly with geographical distance.

We also assessed the effect of the number of genetic markers (15, 30, 100 and 400 loci) on the detectability of edge-impacted changes in population- and individual-based heterozygosity. In Fig. S9a-b, we plot the observed heterozygosity of each sampling unit (population or individuals) measured from one simulation replicate with its estimated 'edge-corrected population size'. We then compare the fitted GAM model from one simulation replicate (coloured line) with i) the fitted GAM model obtained by averaging population or individual-based heterozygosity across simulation replicates (black line); and ii) the theoretical prediction based on the Wright-Malécot approximation (dashed blue line). Overall, the results show a good fit between observed heterozygosity and theoretical prediction (Fig. S10c). However, lower number of genetic markers increases the variance around the expected value, and therefore decreases the percentage of genetic variance explained by N_{edge} (Fig. S10d). The effect of the number of genetic markers is stronger with individual-based heterozygosity (9% - 43% variance explained) than with population-based one (27% - 73% variance explained). Since often in real datasets, few individuals are sampled from the same location, we assessed whether binning individual heterozygosity by value of N_{edge} would improve the percentage of variance explained by the model. The results suggest that individual-binning data analysis improves the explanatory power of N_{edge} at

individual level (from ~9% to ~40% with 15 loci; Fig. S11), and therefore could be used when both number of markers and sampled individuals is limited.

2.4.2 Non-Equilibrium

Since genetic patterns measured from natural populations are unlikely to be at equilibrium, we also investigated the spatio-temporal dynamics of neutral genetic patterns in a habitat undergoing contraction (Fig. 1c). We were particularly interested in exploring the effect of edges on genetic diversity over time, and identify the time required to observe significant differences between central and the edge populations. In Fig. 4 and Fig. S12, we show the temporal changes in heterozygosity for central, edge and corner demes, and assess the impact of population size, dispersal rate and mutation rate on these patterns. Overall, we observed that the genetic diversity of corner and edge demes were consistently decreasing over time, with corner demes being more affected than edge demes. Instead, the heterozygosity of central demes did not change much across most parameter combinations of the simulated scenario. However, our simulations show that when dispersal rate is rather high (e.g., m > 0.08) or carrying capacity and mutation rate are small (e.g., K< 50; $\mu < 10^{-4}$), central demes are also affected by a decrease in heterozygosity (Fig. 4). The effect of high dispersal rate on central deme genetic diversity is stronger when 'mean number of alleles' (MNA) is used as measure of genetic diversity (Fig. S13b). Similar to heterozygosity, the MNA of central demes remains unchanged at higher mutation rate ($\mu \geq 10^{-4}$), whereas it decreases for lower values of mutation rate (Fig. S13c). Unlike dispersal and mutation rate, deme carrying capacity appears to not influence the spatial distribution of MNA and the absolute differences in MNA between central and corner demes (Fig. S12a).

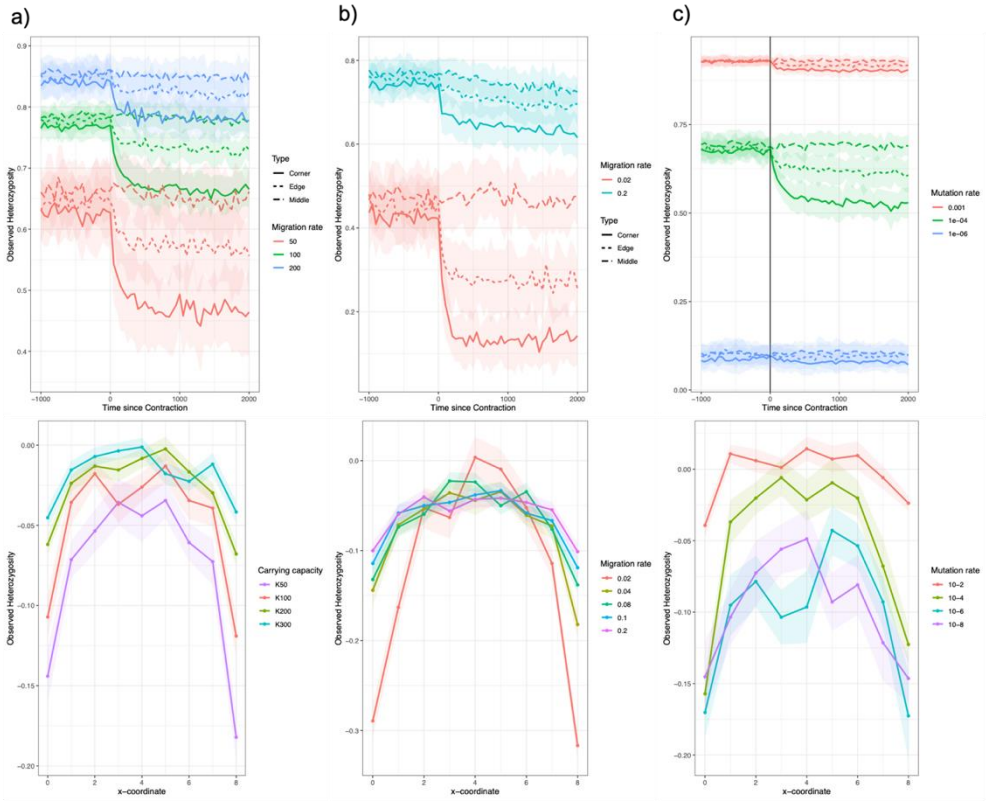


Figure 4. Temporal dynamics of deme H_0 in a 13 x 13 stepping-stone model undergoing habitat contraction. Impact of **a)** carrying capacity (fixed parameters, m: 0.04; μ: 5 x 10^{-4}); **b)** migration rate (fixed parameters, K: 100; μ: 5 x 10^{-4}); and **c)** mutation rate (fixed parameters, K: 100; m: 0.04) on temporal changes in deme H_0. The first-row shows $H_0(t)$ for middle, edge and corner demes. The second row represents the difference in deme H_0 before and after contraction for demes sampled along the diagonal. Across all parameter's combinations, we sampled 14 individuals per deme, 30 microsatellite per individual. Dots and shade indicate, respectively, the mean and standard deviation of deme H_0 based on 30 simulation replicates.

The simulation results show *i*) how patterns of genetic diversity are affected over time by a decrease in total population size (habitat contraction) and *ii*) by an increase in edge-influence. They provide useful information on the effect of population size, dispersal rate and mutation rate on the temporal dynamic of deme genetic diversity. In addition, we have developed an

analytical model that describe how the probability of identity-by-descent of two alleles sampled in the same deme changes over time under habitat contraction ($F_0(t)$; see *Material and Methods*). From $F_0(t)$, we can derive $H_0(t)$, and thus predict how genetic diversity change during habitat contraction. The model considers a 2D stepping-stone model that has been evolving for τ_2 generations, subjected to habitat contraction at τ_1 generations in the past ($\tau_1 < \tau_2$). At τ_2 generations in the past, we assume an instantaneous range expansion as in the radiation model described in Slatkin (1993). See Fig. S2 for a graphical representation of the analytical model.

Habitat contraction decreases total population size and increases deme edge-influence, affecting the expected pattern of deme $H_0(t)$ across the habitat patch. The model implements these processes using four parameters: deme edge-corrected population size before ($N_{edge(i,\tau_2)}$) and after ($N_{edge(i,\tau_1)}$) contraction, and total population size before ($K\ d_{tot(\tau_2)}$) and after ($K\ d_{tot(\tau_1)}$) contraction. Assuming that we have prior knowledge on mutation rate, dispersal rate and patch-size, it is only required to estimate $N_{edge(i,\tau_2)}$ and $N_{edge(i,\tau_1)}$ from habitat cover data (e.g., remote-sensing data) using the proposed edge-corrected population size approach (see above, *Material and Methods*). To describe the dynamic of deme genetic diversity during habitat contraction, we computed numerically $F_0(t)$ from *Eq. 10d* and calculated $H_0(t)$ from *Eq. 9*. In Fig. 5, we compare simulations and analytical results for three demes with different edge-influence (corner, edge and centre demes) across several values of deme carrying capacity, dispersal and mutation rate. The results refer to the scenario defined in Fig. 1c and suggest good agreement between theoretical $H_0(t)$ and simulated deme $\widehat{H_0}(t)$. We assessed the theoretical predictions of the analytical model in habitat patches of different shapes and size (Fig. S14-S16), randomly sampling fifteen demes across the habitat patches. Except few instances, most demes showed good fit between simulation and analytical results.

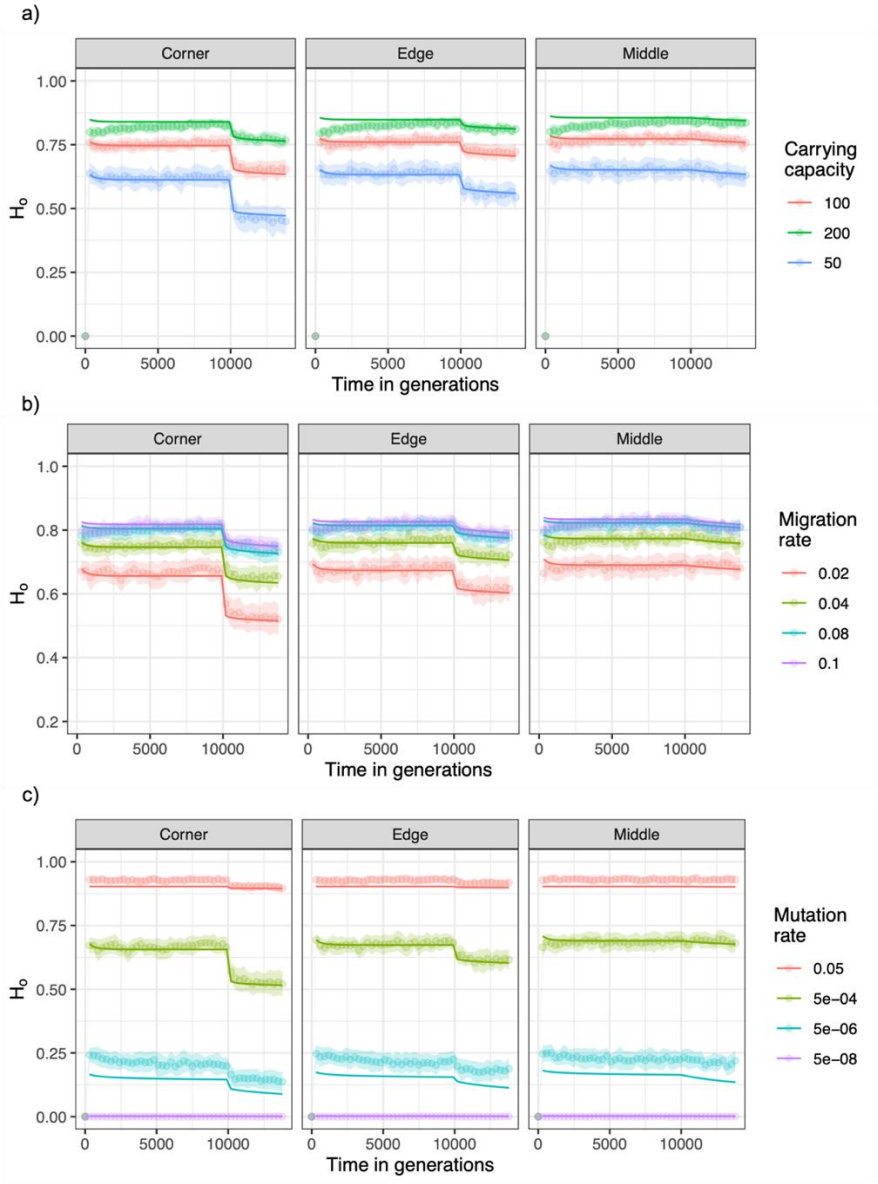


Figure 5. Comparison between analytical and simulation results of a spatially-structured model undergoing habitat contraction. The solid lines represent theoretical predictions based on numerical integration of (*Eq. 10d*). The dots indicate the average deme $H_0(t)$ based on 30 simulation replicates by genotyping 14 individuals per deme at 30 microsatellite loci. We investigated the agreement of analytical and simulation results at various values of **a)** carrying capacity (fixed parameters, m: 0.04; μ: 5 x 10^{-4}), **b)** migration rate (fixed parameters, K: 100; μ: 5 x 10^{-4}) and **c)** mutation rate (fixed parameters, K: 50; m: 0.04).

2.5 Discussion

In this work we investigated the effect of habitat contraction on the spatial distribution of genetic diversity in species with geographically-limited dispersal. During such process, *habitat-amount* and *edge-effect* play a fundamental role in determining genetic diversity within habitat, and those were the major focus of our work. First, we provided an overview of how edges influence genetic diversity depending on deme population size, migration rate and mutation rate. Second, we proposed a measure called *'edge-corrected population size*' (N_{edge}) that takes into account population size, migration rate and mutation rate to predict the additive impact of multiple edges on genetic diversity, a pattern expected to be common in natural population. Ultimately, we proposed and explored a theoretical framework describing non-equilibrium deme heterozygosity in populations undergoing habitat contraction, pointing out to the role of *edge-effect* in determining the tempo and mode of changes in genetic diversity. Broadly speaking, the intent of the present study was in part to link some aspects of species ecology with population genetic theory, proposing several approaches that can be used to predict and test how habitat structure affect intra-species genetic diversity.

2.5.1 Edge-effect in equilibrium populations (habitat patches)

Although often acknowledged, the importance of edge-impacted changes in genetic diversity has been largely overlooked in theoretical and empirical studies. Early works by Maruyama, (1972), Tajima, (1990), have suggested that subpopulations in proximity to habitat edges should hold lower genetic diversity than central subpopulations. However, their conclusions were based on indirect evidence, that is on deductions under the hypothesis that edge-subpopulations would have lower dispersal rate. Later theoretical studies (e.g., Leblois et al., 2006) have shown that observed heterozygosity estimated from samples collected close to habitat boundaries are lower than for samples collected in the centre of the habitat. Similarly, Duforet-Frebourg and Slatkin, (2016) have shown that in a finite 2D stepping-stone model edge-

effect increases the correlation of allele frequency between demes compared to an unbounded 2D stepping-stone model. They also noted that edge-effect is less strong at lower migration rate, that is deviations between observed gene correlation in a bounded habitat and expectations under an unbounded model are more limited in space, being significant in close proximity to the habitat boundary and negligible moving towards to centre. Here, we investigated with further details the influence of edge-effect on genetic diversity (deme H_o, *MNA* and H_e). Our results have shown that deme H_o and H_e are negatively affected by edge-effect, at higher extent in population with low population size or migration rate (K: 50; m: 0.02;), but also in populations with high values of those demographic parameters (e.g., K: 300; m: 0.1) (Fig. 2a, b; Fig. S3a, b first row). Mutation rate showed a 'non-linear' behaviour, since edge-effect was stronger at intermediate values $10^{-4} \leq \mu \leq 10^{-5}$, but still detectable at higher and lower mutation rate ($\mu < 10^{-5}$; $\mu > 10^{-4}$) (Fig. 2c; Fig. S3c first row). Mean number of alleles (MNA), showed a rather different behaviour. Edge-effect appeared to impact deme MNA independently from population size and migration rate, with an average diversity loss in corner demes of 1.5 alleles (Fig. S3a, b bottom row). Deme MNA showed a linear behaviour in relation to mutation rate, with stronger edge-effect at high mutation rate (Fig. S3c bottom row). Overall, our results have shown that corner demes can lose 5% - 33% in heterozygosity (H_o and H_e) and 17% - 33% in allelic diversity compared to central demes (Fig. S4), suggesting a strong impact of edges on the number of unique genetic variants (alleles). It is well-established that allelic richness (or MNA) can more easily detect differences in effective population size than heterozygosity (Nei et al., 1975; Allendorf, 1986; Greenbaum et al., 2014; Hoban et al., 2014), since the latter depends on an average of allele frequencies across many alleles (in the case of multi-allelic microsatellites). Corroborating our and previous findings, there is also the recent meta-analysis of Almeida-Rocha et al., (2020), showing that MNA was the statistic detecting the most negative effect of habitat disturbance on genetic response (effect size: -0.65 [-0.80, -0.50]) compared

to observed and expected heterozygosity (effect size H_{obs}: -0.30 [-0.38, -0.22]; effect size H_{exp}: -0.36 [-0.42, -0.30]). Our findings on the effect of edges on allelic diversity are even more relevant if we consider that microsatellite MNA appears to be a better proxy for genome-wide SNP diversity (Ryynänen et al., 2007; Väli et al., 2008; Vilas et al., 2015; Fischer et al., 2017) and a better predictor of long-term adaptation than heterozygosity (Allendorf, 1986; Hughes et al., 2008; Caballero and Garcia-Dorado, 2013). Therefore, when genome-wide molecular markers (genome-wide) are not available for the species of interest, which is often the case for non-model organisms, MNA can be used as a summary statistic to test molecular edge-effects.

2.5.2 Edge-corrected population size: a proxy of edge-determined changes in genetic diversity

Natural habitat fragments have often irregular shape and heterogenous structure. Simple measures for quantifying edge-effect, such as distance to nearest habitat edge, may not rigorously capture the additive effects of multiple edges, likely occurring in natural habitats. This was the central idea behind the 'edge-influence' statistics developed by Pfeifer et al., (2017), which they used to assess the impact of multiple additive habitat edges on species abundance across 1,673 vertebrate taxa. Their work was preceded by a global analysis of habitat fragmentation across the world, showing that nearly 70% of remaining forests today are at 1 km of an edge (Haddad et al., 2015), and therefore pointing out the likely pervasive impact of habitat edges on biodiversity. Pfeifer et al., (2017) and other previous studies have investigated the ecological response of species to habitat edges, and currently little is known on edge-impacted changes in genetic diversity on natural populations. Reasons for this knowledge gap are probably related to the necessity of a spatially-restricted sampling design, which may not correspond with the primary objectives of most conservation genetic studies, often looking at *landscape*-scale genetic pattern (DiLeo and Wagner, 2016), but also to the need of a large number of genetic markers for reliably capture

fine-scale genetic variation. Few studies, to our knowledge, have explicitly investigated molecular edge-effect in wild populations, namely Radespiel et al., (2019) and Quevedo et al., (2013), although suitable datasets to address this question are likely available in literature.

In this study, we have proposed a framework that uses information on local carrying capacity, species dispersal rate and mutation rate to estimate the additive effect of multiple edges on genetic diversity ($\tilde{E}_{x,y}$). Such measure is then used to correct local carrying capacity (N_{edge}), and predict edge-impacted changes in heterozygosity by implementing N_{edge} in the Malécot-Wright approximation for the probability of identity-by-descent of two alleles sampled from the same location ($F_{0(t=\infty)}$ in *Eq. 6*). Our results have shown that N_{edge} can be a good predictor of spatial genetic variation in finite habitats, under various parameters' combinations (population size, dispersal and mutation rate) (Fig. 3; Fig. S7). The accuracy of the theoretical predictions decreased significantly at lower migration rate (K: 50; m: 0.02; Fig. S8), probably due to the large fluctuations in population size over time at this parameters' combination (Fig. S5a or Fig. S7). Another explanation could be given by the behaviour of the Malécot-Wright approximation for two alleles sampled from the same location. In fact, it has been shown that Malécot-Wright approximation deviates significantly from exact results when neighbourhood size ($2Nm\pi$) is smaller than 10 (Wilkins, 2004). An alternative explanation refers to the observation of Wilkins, (2004) showing that, when $2Nm\pi$ < ~12.5, the genealogical process at the collecting phase can significantly deviate from the standard coalescent. Nevertheless, this parameters' combination did not have an effect on the fitted GAM model, suggesting that when observed data deviates significantly from theoretical predictions, the N_{edge} can still be used as explanatory variable. The analytical model performed well also when observed heterozygosity was measured at the individual level ('Heterozygosity-by-loci'; Aparicio et al., 2006), although, with a lower explanatory power than deme observed heterozygosity (Fig.

S10c, d). In both types of sampling unit, the number of genotyped markers had a strong impact on the percentage of variance explained by the model, with values of 9% and 27% with 15 loci, and 43% and 73% with 400 loci, respectively at the individual and deme-level (Fig. S10d). Binning individual heterozygosity by value of N_{edge} increases the explanatory power of N_{edge} on individual based heterozygosity (Fig. S11), and thus could be used when sample size is limited in real study cases.

The 'edge-corrected population size' proposed in the present study could also be interpreted as a measure of connectivity. As detailed in *Material and Methods*, N_{edge} is estimated by using the weighted mean of deme carrying capacity within an area (A) centred at the focal point. The weight for a deme i included in A is proportional to the expected genetic correlation between the focal point and deme i. The underlying assumption is that deme effective population size at each location in space depends on the census deme-population size, but also on the surrounding amount of habitat from which a given deme can receive migrants at each generation (connectivity). Connectivity indices have a long history in spatial ecology (Moilanen and Nieminen, 2002; Balkenhol et al., 2009; Pfluger and Balkenhol, 2014). In landscape genetics, connectivity measures from spatial ecology have been used to link environmental variables with neutral and adaptive genetic variation (e.g., Keyghobadi et al., 2005; Taylor and Hoffman, 2014). A class of those measures are called 'neighborhood-based' approaches which, similarly to our framework, weight the influence of habitat variables by geographical and dispersal distance. However, although spatial ecology and genetics share fundamental mechanisms, they also differ in some processes. In particular, population genetic theory predicts that neutral genetic variation fluctuates over spatial scales of $l_c \sim \sigma/\sqrt{2\mu}$ (or $\sim\sqrt{m/4\mu}$ for a 2D stepping-stone model) (Malécot, 1948; Barton et al., 2002), which depends on the balance between dispersal and mutation rate. Since N_{edge} is measured by explicitly taking into account the 'spatial scale of variation' (l_c) of the focal

species (*Eq. 3*), it is expected to be a better predictor of genetic diversity than measures of 'ecological' connectivity.

In literature, the necessity of a landscape genetic theory that unifies "*ecology and evolution in a truly interdisciplinary way*" has been stated in several occasions (Dyer, 2015; DiLeo and Wagner, 2016; Balkenhol et al., 2015, 2017). With the present work, we aimed to partially contribute to this knowledge gap and, in addition, provide a null-model at equilibrium for molecular edge effect, which to our knowledge has never been proposed.

2.5.3 Non-Equilibrium pattern of genetic diversity during habitat contraction

The results discussed above make the important assumption that populations are at 'quasi-equilibrium', namely that genetic variation remain nearly unchanged over time, representing the local environmental conditions of the population as we see it today. This might be a robust assumption if, as pointed out by Slatkin and Barton, (1989) and Barton et al. (2013), samples are collected over intermediate spatial scale (area A), such that $\mathrm{Diameter(A)}^2/\sigma^2 \ll t^* \ll \mu^{-1}$, where t^*can be interpreted as a time window over which we are sufficiently confident that habitat structure has remained unchanged. However, predictions under the 'quasi-equilibrium' assumption do not provide, by definition, an understanding of the temporal dynamics of genetic variation.

Here, we have considered a non-equilibrium spatial model where a geographically-structured population undergoes habitat contraction at a given point in time. First, we have explored through spatial simulations the impact of carrying capacity, dispersal rate and mutation rate on the dynamic of genetic diversity. Our results have shown that genetic diversity is influenced by changes in *patch-size* and *edge-effect* relatively quickly, within few hundreds of generations since contraction. The joint effect of habitat reduction and *edge-effect* on genetic diversity were observed across all parameters'

combinations, including large carrying capacity, high migration rate and high mutation rate. Subpopulations in proximity to habitat edges were the most affected by the contraction event, whereas central subpopulations remained nearly unaffected. This is true for most parameters' combination except when dispersal rate is high ($m > 0.02$) and mutation rate is low ($\mu \leq 10^{-6}$). In fact, under these ranges, central demes showed a decrease in genetic diversity. As observed at *equilibrium* condition, MNA is the measure showing the largest changes in genetic diversity, even for central subpopulations (Fig. S13b). Although initially surprising, the larger impact of edges on central demes for highly dispersing populations is deducible from the theory. In fact, the Malécot-Wright approximation includes the quantity l_c, the 'spatial scale of genetic variation', which is proportional to the dispersal rate of the studied species. This implies that changes in neutral allele frequency at a given point in space can fluctuate over relatively large distances in high dispersing populations. These results suggest also that species with high dispersal may in principle be more affected by edge-effect than geographically-restricted populations. Similar observations were reported in Duforet-Frebourg and Slatkin, (2016), showing that edge-effect increases the correlation of allele frequency between demes, at larger extent with high than low migration rate.

In our work, we have in addition developed a theoretical framework describing and predicting the non-equilibrium dynamic of deme genetic diversity following changes in *patch-size* and *edge-effect*. We have shown good agreement between simulation and analytical results, suggesting that such framework could be used for predicting the effect of habitat disturbance on genetic diversity. One of the aspects to stress about our spatial model is that both *patch*-level processes (*edge-effect* and *patch-size*) plays a fundamental role in determining spatial genetic diversity. *Edge-effect* influences deme heterozygosity relatively fast after contraction, depending on deme position relative to edges (Fig. 5; Fig. S14-S16). Instead, reduction in *patch-size* impacts deme heterozygosity indistinctly across demes and at a

slower rate than *edge-effect*. This reflects the separation-of-timescales in the genealogy of structured models, characterized by a short-term (*scattering-phase*) and a long-term (*collecting phase*) coalescent process (Wakeley, 1999; Wilkins, 2004). We have shown that the short-term process can be modelled using the estimated edge-corrected population size (N_{edge}), whereas the long-term process can be modelled as the harmonic mean of the total population size (*patch-size*) over time. A similar theoretical problem has been addressed by Alcala et al., (2013), where they analytically characterized the dynamics of genetic diversity following a change in migration rate between populations in a finite island-model. Ours and Alcala et al., (2013) results on within-population genetic diversity are qualitatively concordant. That is, in both structured models, within-population genetic diversity quickly decays over time following an exponential distribution. The rate of such decay is decomposed in two temporal dynamics, a short-term dynamic dependent on μ, m, local N and total population size; and a long-term dynamic dependent only on μ and total population size. However, here we take into account an explicit spatial model (2D stepping stone), since many species are indeed spatially structured (Sexton et al., 2014).

The non-equilibrium framework here proposed has been developed to investigate and predict the effect of habitat contraction (*edge-effect* and *patch-size*) on genetic diversity. In this framework, changes in genetic diversity have been modelled as the results of changes in local and total effective population size. Therefore, we speculate that our non-equilibrium model could be adapted to investigate recent changes in population size in geographically-structured populations. Several approaches have been developed to model changes in population size or/and connectivity in structured models, although often ignoring the effect of space (e.g., Nielsen and Wakeley 2001; Hey and Nielsen 2004; Gutenkunst et al. 2009; Costa and Wilkinson-Herbots 2016; Arredondo et al., 2021). On the other hand, multiple studies have investigated non-equilibrium genetic patterns in homogenous geographically-structured populations (Nagylaki 1974, 1977; Slatkin, 1991;

Slatkin, 1993; Duforet-Frebourg and Slatkin, 2016; Ortega-Del Vecchyo and Slatkin, 2019), but to our knowledge not much has been done on non-equilibrium spatial models in varying environments (e.g., habitat contraction).

2.5.4 Possible applications and future implementations

The 'edge-corrected population size' statistic described in the present study has been thought as an approach to test edge-impacted changes in genetic diversity. The equilibrium and non-equilibrium frameworks rely on information that are not necessarily available, but for which sufficiently good proxy can be found. For instance, information of patch habitat quality can be retrieved from tree cover datasets (e.g., Hansen et al., 2013) or from Normalized Difference Vegetation Index (NDVI) (e.g., https://earthdata.nasa.gov/lance). Further extension of our work could include the exploration and validation of an additional statistic quantifying the 'genetic sensitivity' of species to edge-effects (see Pfeifer et al., 2017 for details). This would require genetic data across several patches, with sufficient variability in carrying capacity (habitat amount) and 'genetic edge-influence'. Moreover, estimates of 'edge-corrected population size' could provide some insights on how (and if) edge-impacted genetic diversity influences the success of local adaptation in perturbed environments that may arise from the creation of habitat edges. Accordingly, Polechova, (2018), has shown that the success for range expansion or colonization of new environments in a spatially-restricted species is strongly dependent on genetic drift. In particular, Polechova, (2018) define an 'expansion threshold' below which adaptation cannot occur due to the strength of genetic drift, given that it reduces genetic diversity below that required for adaptation to a heterogeneous environment.

Other applications of the 'edge-corrected population size' can be imagined. For instance, our approach could be used to correct measure of pairwise genetic differentiation by controlling for the increased genetic drift

due to edge-effect. In fact, pairwise F_{st} computed including edge-populations are significantly higher than pairwise comparisons with no edge populations (Fig. S17, S18). Lastly, further exploration of the non-equilibrium model could reveal whether this could be used in an inferential framework to estimate past changes in population size in geographically-structured populations. At this stage, this is largely speculative and would require further investigation.

2.6 References

Aguilar, R., Quesada, M., Ashworth, L., Herrerias-Diego, Y., Lobo, J., 2008. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. Molecular Ecology 17, 5177–5188.

Alcala, N., Streit, D., Goudet, J., Vuilleumier, S., 2013. Peak and Persistent Excess of Genetic Diversity Following an Abrupt Migration Increase. Genetics 193, 953–971.

Alho, J.S., Välimäki, K., Merilä, J., 2010. Rhh: an R extension for estimating multilocus heterozygosity and heterozygosity–heterozygosity correlation. Molecular Ecology Resources 10, 720–722.

Allendorf, F.W., 1986. Genetic drift and the loss of alleles versus heterozygosity. Zoo Biology 5, 181–190.

Allendorf, F.W., Luikart, G., 2009. Conservation and the Genetics of Populations. John Wiley & Sons.

Almeida-Rocha, J.M., Soares, L.A.S.S., Andrade, E.R., Gaiotto, F.A., Cazetta, E., 2020. The impact of anthropogenic disturbances on the genetic diversity of terrestrial species: A global meta-analysis. Molecular Ecology 29, 4812–4822.

Aparicio, J.M., Ortego, J., Cordero, P.J., 2006. What should we weigh to estimate heterozygosity, alleles or loci? Mol Ecol 15, 4659–4665.

Archer, F.I., Adams, P.E., Schneiders, B.B., 2017. stratag: An r package for manipulating, summarizing and analysing population genetic data. Molecular Ecology Resources 17, 5–11.

Balkenhol, N., Cushman, S., Storfer, A., Waits, L., 2015. Landscape

Genetics: Concepts, Methods, Applications. John Wiley & Sons.

Balkenhol, N., Dudaniec, R.Y., Krutovsky, K.V., Johnson, J.S., Cairns, D.M., Segelbacher, G., Selkoe, K.A., von der Heyden, S., Wang, I.J., Selmoni, O., Joost, S., 2019. Landscape Genomics: Understanding Relationships Between Environmental Heterogeneity and Genomic Characteristics of Populations, in: Rajora, O.P. (Ed.), Population Genomics: Concepts, Approaches and Applications, Population Genomics. Springer International Publishing, Cham, pp. 261–322.

Balkenhol, N., Waits, L.P., Dezzani, R.J., 2009. Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. Ecography 32, 818–830.

Barlow, J., Lennox, G.D., Ferreira, J., Berenguer, E., Lees, A.C., Nally, R.M., Thomson, J.R., et al., 2016. Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. Nature 535, 144–147.

Barton, N.H., Depaulis, F., Etheridge, A.M., 2002. Neutral Evolution in Spatially Continuous Populations. Theoretical Population Biology 61, 31–48.

Barton, N.H., Etheridge, A.M., Kelleher, J., Véber, A., 2013. Inference in two dimensions: Allele frequencies versus lengths of shared sequence blocks. Theoretical Population Biology, Coalescent Theory 87, 105–119.

Barton, N.H., Wilson, I., Harvey, P.H., Leigh Brown, A.J., Smith, J.M., 1995. Genealogies and geography. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 349, 49–59.

Betts, M.G., Fahrig, L., Hadley, A.S., Halstead, K.E., Bowman, J., Robinson, W.D., Wiens, J.A., Lindenmayer, D.B., 2014. A species-centered approach for uncovering generalities in organism responses to habitat loss and fragmentation. Ecography 37, 517–527.

Betts, M.G., Wolf, C., Pfeifer, M., Banks-Leite, C., Arroyo-Rodríguez, V., Ribeiro, D.B., Barlow, J., Eigenbrod, F., Faria, D., Fletcher, R.J., Hadley, A.S., Hawes, J.E., Holt, R.D., Klingbeil, B., Kormann, U., Lens, L., Levi, T., Medina-Rangel, G.F., Melles, S.L., Mezger, D., Morante-Filho, J.C., Orme, C.D.L., Peres, C.A., Phalan, B.T., Pidgeon, A., Possingham, H., Ripple, W.J., Slade, E.M., Somarriba, E., Tobias, J.A., Tylianakis, J.M., Urbina-Cardona, J.N., Valente, J.J., Watling, J.I., Wells, K., Wearn, O.R., Wood, E., Young, R., Ewers, R.M., 2019. Extinction filters mediate the global effects of habitat fragmentation on animals. Science 366, 1236–1239.

Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., Fonseca, G.A.B.D., Rylands, A.B., Konstant, W.R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G., Hilton-Taylor, C., 2002. Habitat Loss and Extinction in the Hotspots of Biodiversity. Conservation Biology 16, 909–923.

Bueno, A.S., Peres, C.A., 2019. Patch-scale biodiversity retention in fragmented landscapes: Reconciling the habitat amount hypothesis with the island biogeography theory. Journal of Biogeography 46, 621–632.

Caballero, A., García-Dorado, A., 2013. Allelic Diversity and Its Implications for the Rate of Adaptation. Genetics 195, 1373–1384.

Cayuela, H., Rougemont, Q., Prunier, J.G., Moore, J.-S., Clobert, J., Besnard, A., Bernatchez, L., 2018. Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review. Molecular Ecology 27, 3976–4010.

Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M., Palmer, T.M., 2015. Accelerated modern human–induced species losses: Entering the sixth mass extinction. Science Advances 1, e1400253.

Ceballos, G., Ehrlich, P.R., Dirzo, R., 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and

declines. PNAS 114, E6089–E6096.

Cox, J.T., Durrett, R., 2002. The stepping stone model: New formulas expose old myths. The Annals of Applied Probability 12, 1348–1377.

Currat, M., Arenas, M., Quilodràn, C.S., Excoffier, L., Ray, N., 2019. SPLATCHE3: simulation of serial genetic data under spatially explicit evolutionary scenarios including long-distance dispersal. Bioinformatics 35, 4480–4483.

Currat, M., Ray, N., Excoffier, L., 2004. splatche: a program to simulate genetic diversity taking into account environmental heterogeneity. Molecular Ecology Notes 4, 139–142.

Curtis, J.T., 1971. THE MODIFICATION OF MID-LATITUDE GRASSLANDS AND FORESTS BY MAN.

Curtis, J.T., 1956. The modification of mid-latitude grasslands and forests by man., in: Man's Role in Changing the Face of the Earth. University of Chicago Press., Chicago, IL, pp. 721–736.

Cushman, S.A., Shirk, A., Landguth, E.L., 2012. Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex landscapes. Landscape Ecol 27, 369–380.

Didham, R.K., Kapos, V., Ewers, R.M., 2012. Rethinking the conceptual foundations of habitat fragmentation research. Oikos 121, 161–170.

DiLeo, M.F., Wagner, H.H., 2016. A Landscape Ecologist's Agenda for Landscape Genetics. Curr Landscape Ecol Rep 1, 115–126.

Duforet-Frebourg, N., Slatkin, M., 2016. Isolation-by-distance-and-time in a stepping-stone model. Theoretical Population Biology 108, 24–35.

Dyer, R.J., 2015. Population Graphs and Landscape Genetics. Annual Review of Ecology, Evolution, and Systematics 46, 327–342.

Excoffier, L., Foll, M., 2011. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics 27, 1332–1334.

Fahrig, L., 2019. Habitat fragmentation: A long and tangled tale. Global Ecology and Biogeography 28, 33–41.

Fahrig, L., 2017. Ecological Responses to Habitat Fragmentation Per Se. Annual Review of Ecology, Evolution, and Systematics 48, 1–23.

Fahrig, L., 2013. Rethinking patch size and isolation effects: the habitat amount hypothesis. Journal of Biogeography 40, 1649–1663.

Fahrig, L., 2003. Effects of Habitat Fragmentation on Biodiversity. Annu. Rev. Ecol. Evol. Syst. 34, 487–515.

Fischer, M.C., Rellstab, C., Leuzinger, M., Roumet, M., Gugerli, F., Shimizu, K.K., Holderegger, R., Widmer, A., 2017. Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in Arabidopsis halleri. BMC Genomics 18, 69.

Fletcher, R.J., 2005. Multiple edge effects and their implications in fragmented landscapes. Journal of Animal Ecology 74, 342–352.

Fletcher, R.J., Didham, R.K., Banks-Leite, C., Barlow, J., Ewers, R.M., Rosindell, J., Holt, R.D., Gonzalez, A., Pardini, R., Damschen, E.I., Melo, F.P.L., Ries, L., Prevedello, J.A., Tscharntke, T., Laurance, W.F., Lovejoy, T., Haddad, N.M., 2018. Is habitat fragmentation good for biodiversity? Biological Conservation 226, 9–15.

Greenbaum, G., Templeton, A.R., Zarmi, Y., Bar-David, S., 2014. Allelic

Richness following Population Founding Events – A Stochastic Modeling Framework Incorporating Gene Flow and Genetic Drift. PLOS ONE 9, e115203.

Griffiths, R.C., Tavare, S., 1994. Sampling Theory for Neutral Alleles in a Varying Environment. Philosophical Transactions: Biological Sciences 344, 403–410.

Groombrifge, G., 1992. Global Biodiversity: Status of the Earth's Living Resources. Springer Netherlands.

Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, et al., 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. Science Advances 1, e1500052.

Haddad, N.M., Gonzalez, A., Brudvig, L.A., Burt, M.A., Levey, D.J., Damschen, E.I., 2017. Experimental evidence does not support the Habitat Amount Hypothesis. Ecography 40, 48–55.

Hanski, I., Pakkala, T., Kuussaari, M., Lei, G., 1995. Metapopulation Persistence of an Endangered Butterfly in a Fragmented Landscape. Oikos 72, 21–28.

Hanski, I., Zurita, G.A., Bellocq, M.I., Rybicki, J., 2013. Species–fragmented area relationship. PNAS 110, 12715–12720.

Hastie, T., Tibshirani, R., 1987. Generalized Additive Models: Some Applications. Journal of the American Statistical Association 82, 371–386.

Herbots, H.M.J.D., 1994. Stochastic Models in Population Genetics: Genealogy and Genetic Differentiation in Structured Populations (Thesis).

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907–913.

Hey, J., 1991. A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. Theoretical Population Biology 39, 30–48.

Hoban, S., Arntzen, J.A., Bruford, M.W., Godoy, J.A., Hoelzel, A.R., Segelbacher, G., Vilà, C., Bertorelle, G., 2014. Comparative evaluation of potential indicators and temporal sampling protocols for monitoring genetic erosion. Evolutionary Applications 7, 984–998.

Honnay, O., Jacquemyn, H., 2007. Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. Conserv Biol 21, 823–831.

Hughes, A.R., Inouye, B.D., Johnson, M.T.J., Underwood, N., Vellend, M., 2008. Ecological consequences of genetic diversity. Ecol Lett 11, 609–623.

Hylander, K., Ehrlén, J., 2013. The mechanisms causing extinction debts. Trends in Ecology & Evolution 28, 341–346.

Jackson, N.D., Fahrig, L., 2016. Habitat amount, not habitat configuration, best predicts population genetic structure in fragmented landscapes. Landscape Ecol 31, 951–968.

Jackson, N.D., Fahrig, L., 2014. Landscape context affects genetic diversity at a much larger spatial extent than population abundance. Ecology 95, 871–881.

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.

Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070–3071.

Keyghobadi, N., Roland, J., Matter, S.F., Strobeck, C., 2005. Among- and

within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. Proceedings of the Royal Society B: Biological Sciences 272, 553–560.

Kimura, M., Ohta, T., 1973. Mutation and evolution at the molecular level. Genetics 73, Suppl 73:19-35.

Kimura, M., Weiss, G.H., 1964. The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. Genetics 49, 561–576.

Kingman, J.F.C., 1982a. The coalescent. Stochastic Processes and their Applications 13, 235–248.

Kingman, J.F.C., 1982b. On the genealogy of large populations. Journal of Applied Probability 19, 27–43.

Leblois, R., Estoup, A., Streiff, R., 2006. Genetics of recent habitat contraction and reduction in population size: does isolation by distance matter? Mol Ecol 15, 3601–3615.

Lino, A., Fonseca, C., Rojas, D., Fischer, E., Ramos Pereira, M.J., 2019. A meta-analysis of the effects of habitat loss and fragmentation on genetic diversity in mammals. Mammalian Biology 94, 69–76.

Malécot, G., 1948. Les Mathématiques de L'hérédité. Masson.

Martin, A.E., Fahrig, L., 2012. Measuring and selecting scales of effect for landscape predictors in species–habitat models. Ecological Applications 22, 2277–2292.

Maruyama, T., 1972. Rate of Decrease of Genetic Variability in a Two-Dimensional Continuous Population of Finite Size. Genetics 70, 639–651.

Millette, K.L., Fugère, V., Debyser, C., Greiner, A., Chain, F.J.J., Gonzalez, A., 2020. No consistent effects of humans on animal genetic diversity worldwide. Ecology Letters 23, 55–67.

Miraldo, A., Li, S., Borregaard, M.K., Flórez-Rodríguez, A., Gopalakrishnan, S., Rizvanovic, M., Wang, Z., Rahbek, C., Marske, K.A., Nogués-Bravo, D., 2016. An Anthropocene map of genetic diversity. Science 353, 1532–1535.

Moilanen, A., Nieminen, M., 2002. Simple Connectivity Measures in Spatial Ecology. Ecology 83, 1131–1145.

Mona, S., Ray, N., Arenas, M., Excoffier, L., 2014. Genetic consequences of habitat fragmentation during a range expansion. Heredity 112, 291–299.

Monteiro, W.P., Veiga, J.C., Silva, A.R., Carvalho, C. da S., Lanes, É.C.M., Rico, Y., Jaffé, R., 2019. Everything you always wanted to know about gene flow in tropical landscapes (but were afraid to ask). PeerJ 7, e6446.

Moore, N.W., 1962. The Heaths of Dorset and their Conservation. Journal of Ecology 50, 369–391.

Nagylaki, T., 1977. Decay of genetic variability in geographically structured populations. Proc Natl Acad Sci U S A 74, 2523–2525.

Nagylaki, T., 1974. The Decay of Genetic Variability in Geographically Structured Populations*. Proc Natl Acad Sci U S A 71, 2932–2936.

Nei, M., Maruyama, T., Chakraborty, R., 1975. The Bottleneck Effect and Genetic Variability in Populations. Evolution 29, 1–10.

Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Lysenko, I., Senior, R.A., Börger, L., Bennett, D.J., et al., 2015. Global effects of land use on local terrestrial biodiversity. Nature 520, 45–50.

Ohta, T., Kimura, M., 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet Res 22, 201–204.

Ortega-Del Vecchyo, D., Slatkin, M., 2019. F ST between archaic and present-day samples. Heredity 122, 711–718.

Pardini, R., Bueno, A. de A., Gardner, T.A., Prado, P.I., Metzger, J.P., 2010. Beyond the Fragmentation Threshold Hypothesis: Regime Shifts in Biodiversity Across Fragmented Landscapes. PLOS ONE 5, e13666.

Pfeifer, M., Lefebvre, V., Peres, C.A., Banks-Leite, C., Wearn, O.R., Marsh, C.J., Butchart, S.H.M., et al., 2017. Creation of forest edges has a global impact on forest vertebrates. Nature 551, 187–191.

Pflüger, F.J., Balkenhol, N., 2014. A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. Molecular Ecology 23, 2146–2156.

Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M., Sexton, J.O., 2014. The biodiversity of species and their rates of extinction, distribution, and protection. Science 344.

Polechová, J., 2018. Is the sky the limit? On the expansion threshold of a species' range. PLOS Biology 16, e2005372.

Prevedello, J.A., Vieira, M.V., 2010. Does the type of matrix matter? A quantitative review of the evidence. Biodivers Conserv 19, 1205–1223.

Quevedo, A.A., Schleuning, M., Hensen, I., Saavedra, F., Durka, W., 2013. Forest fragmentation and edge effects on the genetic structure of Clusia sphaerocarpa and C. lechleri (Clusiaceae) in tropical montane forests. Journal of Tropical Ecology 29, 321–329.

Radespiel, U., Schulte, J., Burke, R.J., Lehman, S.M., 2019. Molecular edge effects in the Endangered golden-brown mouse lemur Microcebus ravelobensis. Oryx 53, 716–726.

Rasteiro, R., Bouttier, P.-A., Sousa, V.C., Chikhi, L., 2012. Investigating sex-biased migration during the Neolithic transition in Europe, using an explicit spatial simulation framework. Proceedings of the Royal Society B: Biological Sciences 279, 2409–2416.

Ries, L., Fletcher, R.J., Battin, J., Sisk, T.D., 2004. Ecological Responses to Habitat Edges: Mechanisms, Models, and Variability Explained. Annu. Rev. Ecol. Evol. Syst. 35, 491–522.

Ries, L., Murphy, S.M., Wimp, G.M., Fletcher, R.J., 2017. Closing Persistent Gaps in Knowledge About Edge Ecology. Curr Landscape Ecol Rep 2, 30-41.

Ringbauer, H., Kolesnikov, A., Field, D.L., Barton, N.H., 2018. Estimating Barriers to Gene Flow from Distorted Isolation-by-Distance Patterns. Genetics 208, 1231–1245.

Rivera Ortíz, F.A., Aguilar, R., Arizmendi, M.D.C., Quesada, M., Oyama, K., 2015. Habitat fragmentation and genetic variability of tetrapod populations.

Ruffell, J., Banks-Leite, C., Didham, R.K., 2016. Accounting for the causal basis of collinearity when measuring the effects of habitat loss versus habitat fragmentation. Oikos 125, 117–125.

Rybicki, J., Hanski, I., 2013. Species–area relationships and extinctions caused by habitat loss and fragmentation. Ecology Letters 16, 27–38.

Ryo, M., Aguilar-Trigueros, C.A., Pinek, L., Muller, L.A.H., Rillig, M.C., 2019. Basic Principles of Temporal Dynamics. Trends in Ecology & Evolution 34, 723–733.

Ryynänen, H.J., Tonteri, A., Vasemägi, A., Primmer, C.R., 2007. A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (Salmo salar). J Hered 98, 692–704.

Schipper, J., Chanson, J.S., Chiozza, F., Cox, N.A., Hoffmann, M., Katariya, V., Lamoreux, J., et al., 2008. The Status of the World's Land and Marine Mammals: Diversity, Threat, and Knowledge. Science 322, 225–230.

Schlaepfer, D.R., Braschler, B., Rusterholz, H.-P., Baur, B., 2018. Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. Ecosphere 9, e02488.

Sexton, J.P., Hangartner, S.B., Hoffmann, A.A., 2014. Genetic Isolation by Environment or Distance: Which Pattern of Gene Flow Is Most Common? Evolution 68, 1–15.

Slatkin, M., 1993. Isolation by Distance in Equilibrium and Non-Equilibrium Populations. Evolution 47, 264–279.

Slatkin, M., 1991. Inbreeding coefficients and coalescence times. Genetics Research 58, 167–175.

Slatkin, M., 1987. Gene flow and the geographic structure of natural populations. Science 236, 787–792.

Slatkin, M., Barton, N.H., 1989. A COMPARISON OF THREE INDIRECT METHODS FOR ESTIMATING AVERAGE LEVELS OF GENE FLOW. Evolution 43, 1349–1368.

Smith, J.M., Slatkin, M., 1973. The Stability of Predator-Prey Systems. Ecology 54, 384–391.

Storfer, A., Murphy, M.A., Evans, J.S., Goldberg, C.S., Robinson, S., Spear,

S.F., Dezzani, R., Delmelle, E., Vierling, L., Waits, L.P., 2007. Putting the ‘landscape’ in landscape genetics. Heredity 98, 128–142.

Storfer, A., Murphy, M.A., Spear, S.F., Holderegger, R., Waits, L.P., 2010. Landscape genetics: where are we now? Molecular Ecology 19, 3496–3514.

Strobeck, C., 1987. Average Number of Nucleotide Differences in a Sample from a Single Subpopulation: A Test for Population Subdivision. Genetics 117, 149–153.

Tajima, F., 1990. Relationship between migration and DNA polymorphism in a local population. Genetics 126, 231–234.

Väli, U., Einarsson, A., Waits, L., Ellegren, H., 2008. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? Mol Ecol 17, 3808–3817.

Vancutsem, C., Achard, F., Pekel, J.-F., Vieilledent, G., Carboni, S., Simonetti, D., Gallego, J., Aragão, L.E.O.C., Nasi, R., 2021. Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. Science Advances 7, eabe1603.

Vilas, A., Pérez-Figueroa, A., Quesada, H., Caballero, A., 2015. Allelic diversity for neutral markers retains a higher adaptive potential for quantitative traits than expected heterozygosity. Mol Ecol 24, 4419–4432.

Vranckx, G., Jacquemyn, H., Muys, B., Honnay, O., 2012. Meta-analysis of susceptibility of woody plants to loss of genetic diversity through habitat fragmentation. Conserv Biol 26, 228–237.

Wagner, H.H., Fortin, M.-J., 2013. A conceptual framework for the spatial analysis of landscape genetic data. Conserv Genet 14, 253–261.

Wakeley, J., 1999. Nonequilibrium Migration in Human History. Genetics 153,

1863–1871.

Watling, J.I., Arroyo-Rodríguez, V., Pfeifer, M., Baeten, L., Banks-Leite, C., Cisneros, L.M., et al., 2020. Support for the habitat amount hypothesis from a global synthesis of species density studies. Ecology Letters 23, 674–681.

Wilkins, J.F., 2004. A Separation-of-Timescales Approach to the Coalescent in a Continuous Population. Genetics 168, 2227–2244.

Wilkins, J.F., Wakeley, J., 2002. The Coalescent in a Continuous, Finite, Linear Population. Genetics 161, 873–888.

Wood, S.N., 2001. mgcv: GAMs and Generalized Ridge Regression for R. R News 1, 20–25.

Wright, S., 1943. Isolation by Distance. Genetics 28, 114–138.

2.7 Supplementary Figures

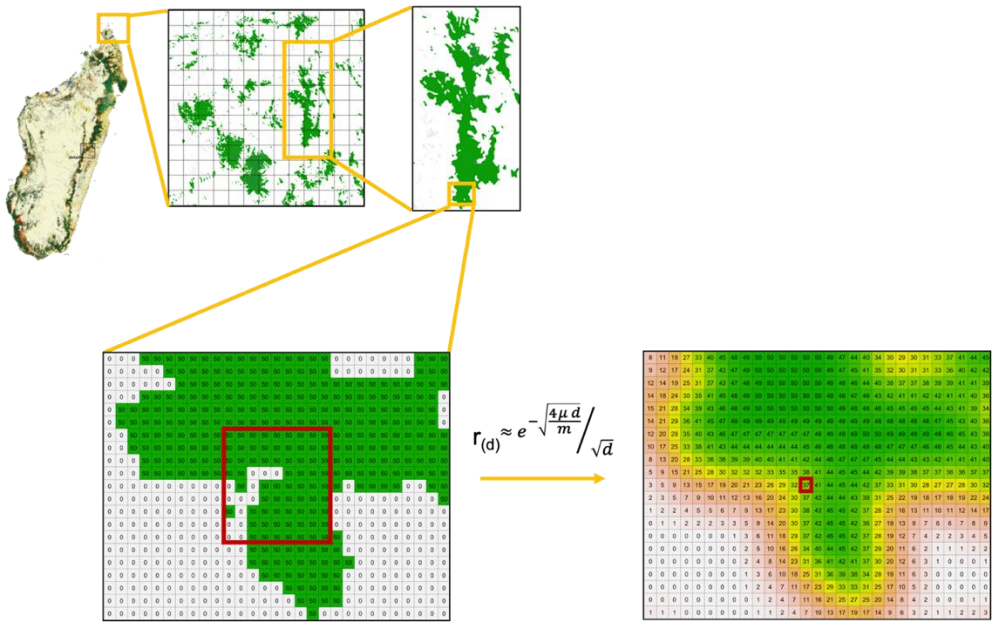
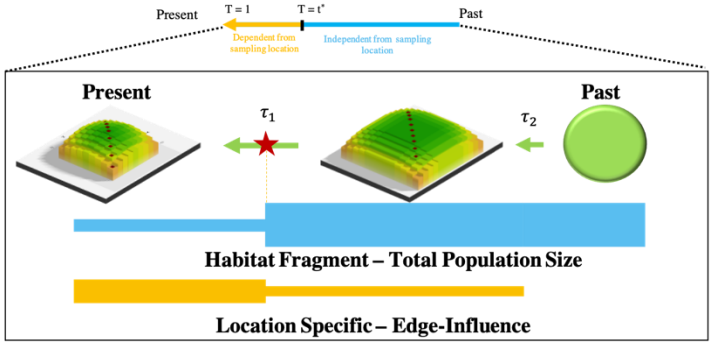


Figure S1. Illustration of the spatial filtering for computing 'edge-influence index'. As an example, it is shown the percentage of tree cover for several forest fragments in the Loky-Manambato region, northern Madagascar. We removed detailed information on tree cover by setting the cells with tree cover < 60 to 0, and to 50 otherwise. The solid red line shows a square matrix with area A and side length R = 9. The focal point (in the centre of the square matrix) will get the weighted average of the cells within A. The weights are calculated using the genetic correlation equation from Kimura and Weiss, 1964. The genetic correlation equation show that the focal point has the heaviest weight and other cells within A have smaller weights as their distance from the focal point increases.



$$\boldsymbol{H_0(t)} = \left(1 - \int_0^{\tau_2} \boldsymbol{f_0(t)}\, \boldsymbol{dt}\right) \boldsymbol{H_{t*}}$$

$$\int_0^{\tau_2} \boldsymbol{f_0(t)}\, \boldsymbol{dt} = \Pr(\text{coal. by } \tau_2 \text{ \& no mutation}) = \Pr(\text{coal. by } \boldsymbol{\tau_1} \text{ \& no mutation}) + \Pr(\boldsymbol{\tau_1} < \text{coal.} \leq \boldsymbol{\tau_2} \text{ \& no mutation})$$

Figure S2. Illustration of the non-equilibrium spatial analytical model of habitat contraction. The genealogical process of a spatially structured model can be divided in two phases: *scattering phase* (dependent on alleles sampling location) and *collecting phase* (mostly independent on sampling location). The transition time between these two phases is indicated by t^*. Thinking backward in time, two alleles are sampled within subpopulation in the present ($t = 0$). The habitat fragment has undergone contraction at τ_1 and an instantaneous range expansion at τ_2. The analytical model assumes that edge effect influences the *scattering phase*, whereas habitat loss influences the *collecting phase* of the genealogical process.

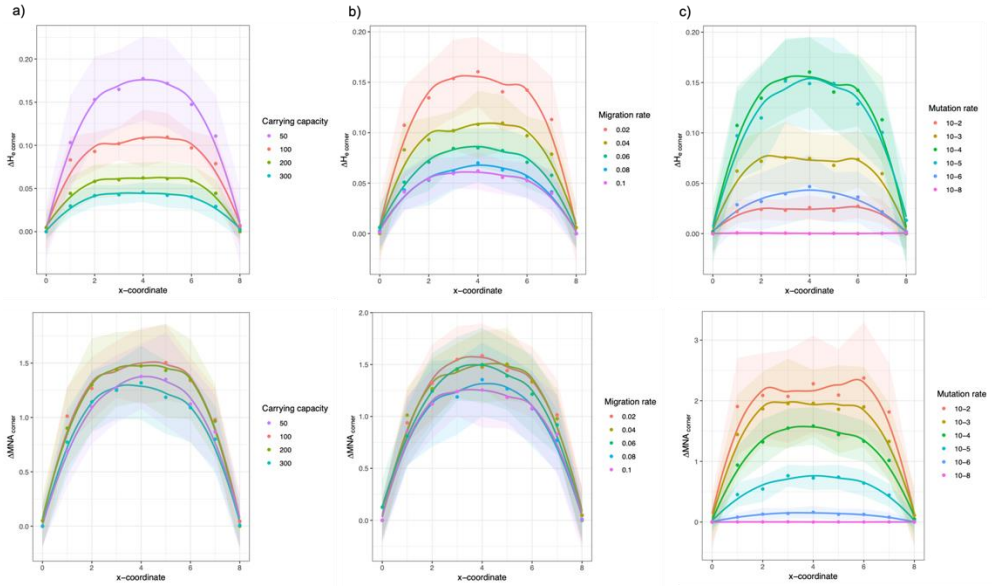


Figure S3. Spatial distribution of differences in expected heterozygosity ($\Delta H_{e\ corner}$) and mean number of alleles (ΔMNA_{corner}) relative to corner demes along the diagonal of a 9 x 9 stepping-stone model. Impact of **a)** carrying capacity (fixed parameters, m: 0.04; μ: 5 x 10^{-4}); **b)** migration rate (fixed parameters, K: 100; μ: 5 x 10^{-4}); and **c)** mutation rate (fixed parameters, K: 100; m: 0.02). Central demes (x = 4) show significantly higher H_e with respect to corner demes (x = 0 and x = 8) at lower carrying capacity (K: 50; $\Delta H_{o\ corner}$: ~ 0.17), lower migration rate (m: 0.02; $\Delta H_{o\ corner}$: ~ 0.15) and intermediate mutation rate (μ: 5 x 10^{-4} / 5 x 10^{-5}; $\Delta H_{o\ corner}$: ~ 0.15). ΔMNA_{corner} does not change significantly with carrying capacity and migration rate, whereas it is more pronounced at higher mutation rate (e.g., μ: 5 x 10^{-2}; ΔMNA_{corner}: ~ 2.2).

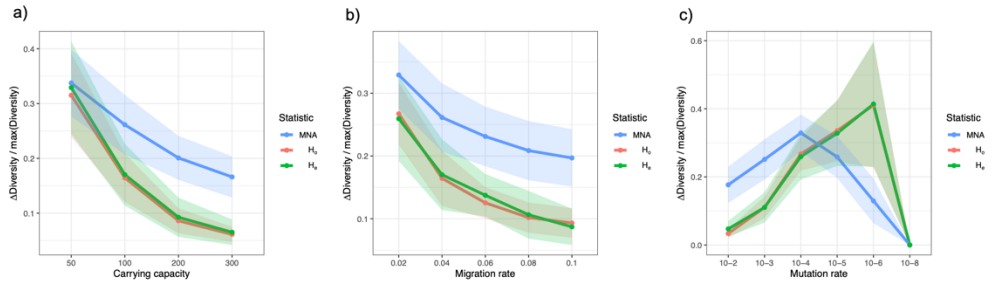


Figure S4. Relative difference in genetic diversity (ΔMNA_{corner}, $\Delta H_{o\,corner}$, $\Delta H_{e\,corner}$) between central and corner demes sampled in a 9 x 9 stepping-stone model. Corner demes shows a decrease in genetic diversity of **a)** ~33% or ~5-17% at, respectively, K=50 and K=300 (fixed parameters, m: 0.04; μ: 5 x 10^{-4}); **b)** ~25-33% or ~10-20% at, respectively, m: 0.02 and m: 0.1 (fixed parameters, K: 100; μ: 5 x 10^{-4}); and **c)** ~5-20% or ~17-40% at, respectively, μ: 5 x 10^{-2} and μ: 5 x 10^{-6} (fixed parameters, K: 100; m: 0.02). $\Delta H_{o\,corner}$ and $\Delta H_{e\,corner}$ increases linearly with decreasing mutation rate up to μ: 5 x 10^{-8}, since no genetic diversity was present at this value of mutation rate. ΔMNA_{corner} shows a non-linear effect with respect to mutation rate, being more affected by edge-effect at intermediate mutation rate (10^{-3}-10^{-5}).

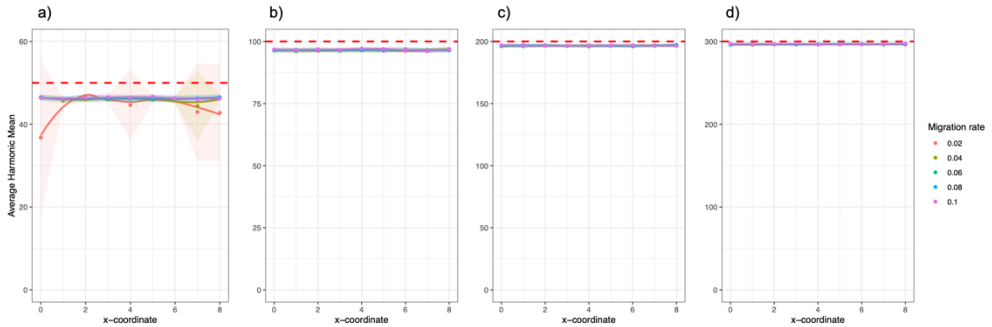


Figure S5. Harmonic mean of deme population size (HMP) across 10,000 generations in a 9 x 9 stepping-stone model for several values of migration rate. a) K: 50; **b)** K: 100; **c)** K: 200; **d)** K: 300. Dashed red line indicates the simulated carrying capacity for each scenario. Thick lines represent the average HMP across 30 demography simulations, with standard deviation given by the shade around the mean. Note that the harmonic mean was always lower than the value of carrying capacity set at each scenario, with larger deviation for smaller values of carrying capacity. The results do not show differences in HMP along diagonal demes, with the exception of K: 50, m: 0.02 scenario where corner demes show lower HMP than central demes.

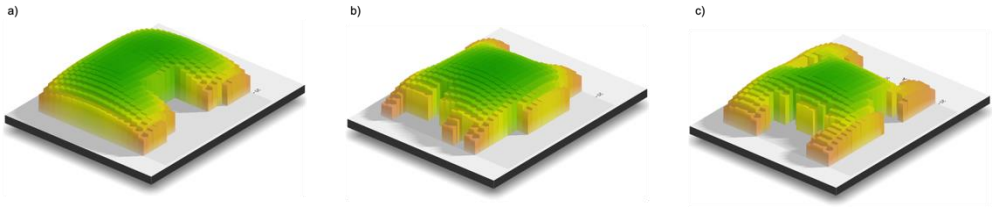


Figure S6. Simulated scenario for the edge-corrected population size. a) 'scenario 1' including 382 demes, also present in main text; **b)** 'scenario 2' including 318 demes; **c)** 'scenario 3' including 305 demes. Colors and bars refer to deme-specific edge influence, that for visualization purposes was estimated assuming K: 50, m: 0.04 and μ: 5 x 10^{-4}.

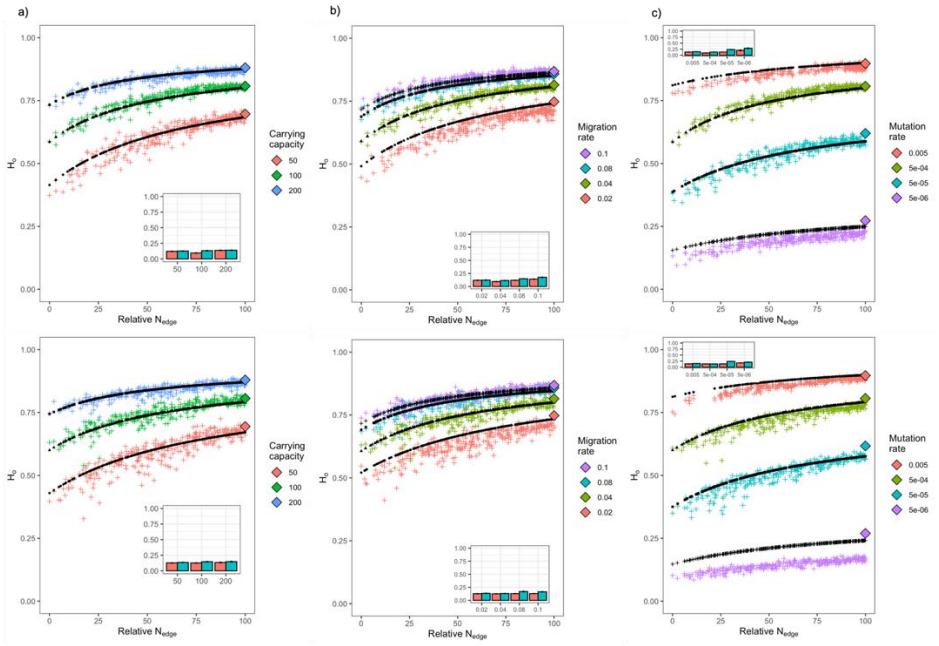


Figure S7. Comparison between edge-corrected theoretical results and simulations on two 23 x 23 stepping-stone model of irregular shape. Impact of **a)** carrying capacity (fixed parameters, m: 0.04; μ: 5 x 10^{-4}); **b)** migration rate (fixed parameters, K: 100; μ: 5 x 10^{-4}); and **c)** mutation rate (fixed parameters, K: 100; m: 0.04) on the predictability of edge-impacted changes in genetic diversity. Each cross represents the average deme H_o across 30 simulation replicates. Relative N_{edge} is a standardized edge-corrected population size (between 0 and 100) for visualization purposes. Black symbols define the theoretical predictions using N_{edge} in the Wright-Malécot approximation (Eq. 6). The insert bar plots quantify the NRMSE under the Wright-Malécot approximation (light blue) and GAM model (red). Colored diamonds indicate the expected H_o using the Wright-Malécot approximation and ignoring edge-effect.

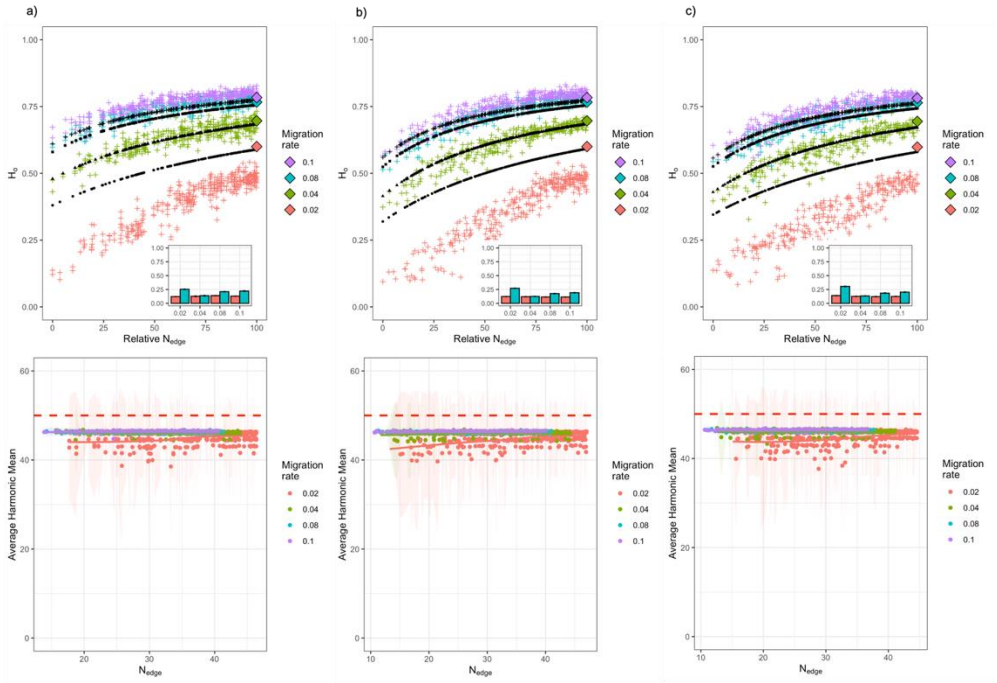


Figure S8. Comparison between edge-corrected theoretical results and simulations on three 23 x 23 stepping-stone model of irregular shape with K: 50. First row: the effect of varying migration rate on the observed heterozygosity predicted for **a)** scenario 1; **b)** scenario 2; and **c)** scenario 3. See Fig. S5 for a visual representation of the simulated habitat shapes. The insert bar plots quantify the NRMSE under the Wright-Malécot approximation (light blue) and GAM model (red). Second row: harmonic mean of deme population size (HMP) across 10,000 generations with respect to N_{edge}. Dashed red line indicates K: 50. Standard deviation is given by the shade around the mean. K: 50 and m: 0.02 shows large variability and lower HMP than higher migration rate, which can explain the poor predictability of N_{edge} at these parameter values.

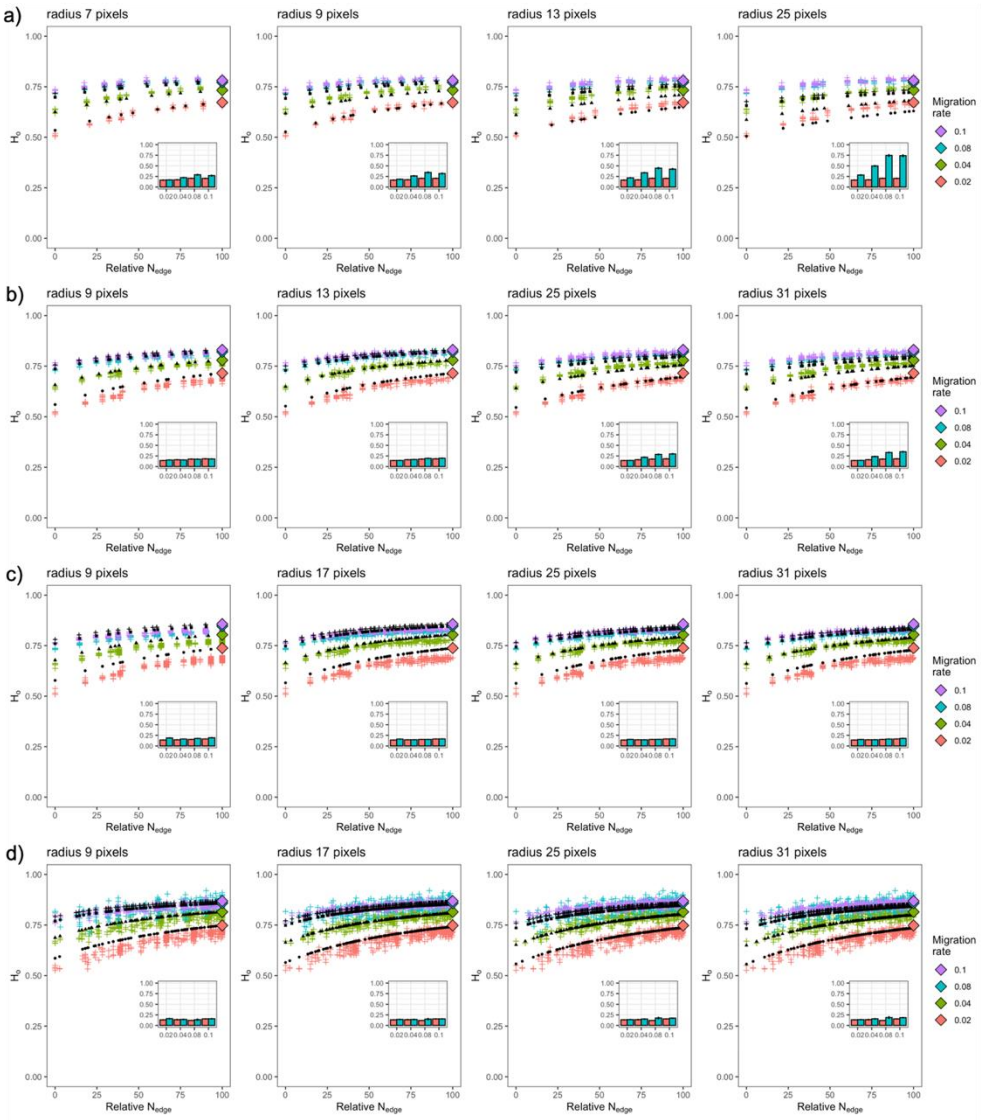


Figure S9. Effect of radius selection on the edge-corrected population size (N_{edge}). These results compare theoretical expectations and simulations for **a)** 9 x 9; **b)** 13 x 13; **c)** 17 x 17 square grids; and **d)** 'scenario 1' from Fig. S1 under different migration rate values and pixel size for the radius selection (fixed parameters, K: 100 and μ: 5 x 10^{-4}). Each cross represents the average deme H_o across 30 simulation replicates. Colored diamonds indicate the expected H_o using the Wright-Malécot approximation and ignoring edge-effect. The insert bar plots quantify the NRMSE under the Wright-Malécot approximation (light blue) and GAM model (red). We conclude that the radius size has little effect on the performance of N_{edge}.

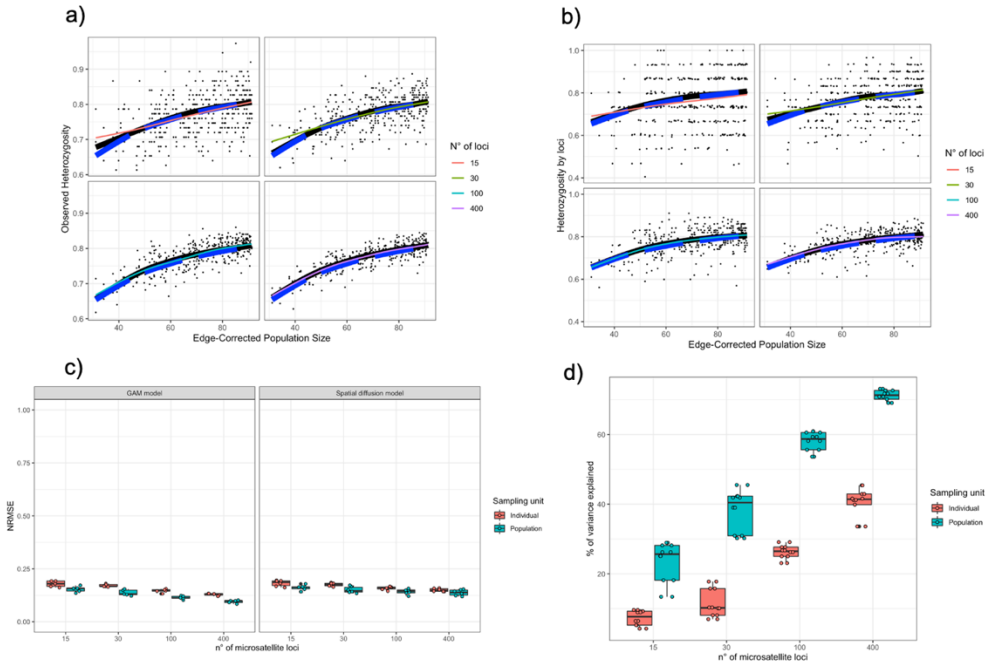


Figure S10. Effect of the number of genetic markers and sampling unit (deme or individual) on the detection of edge-determined changes in genetic diversity. The impact of 15, 30, 100 or 400 microsatellite markers on the agreement between theoretical and simulation results for **a)** deme H_o and **b)** individual 'Heterozygosity by loci'. Each point represents genetic diversity estimated from samples collected from one deme in one simulation replicate. Dashed blue and solid black line indicate, respectively, theoretical results and average value across 30 simulations. Colored line is a GAM model obtained from one simulation replicate. **c)** NRMSE and **d)** Percentage of variance explained by the 'edge-corrected population size' for deme H_o (light blue) and individual 'Heterozygosity by loci' (red) when varying number of genetic markers. The results suggest good agreement between theoretical and estimates of genetic diversity from one simulation replicate (as proxy of what can be obtained from field samples). The power of detection is greatly affected by the number of microsatellite markers, if below 100.

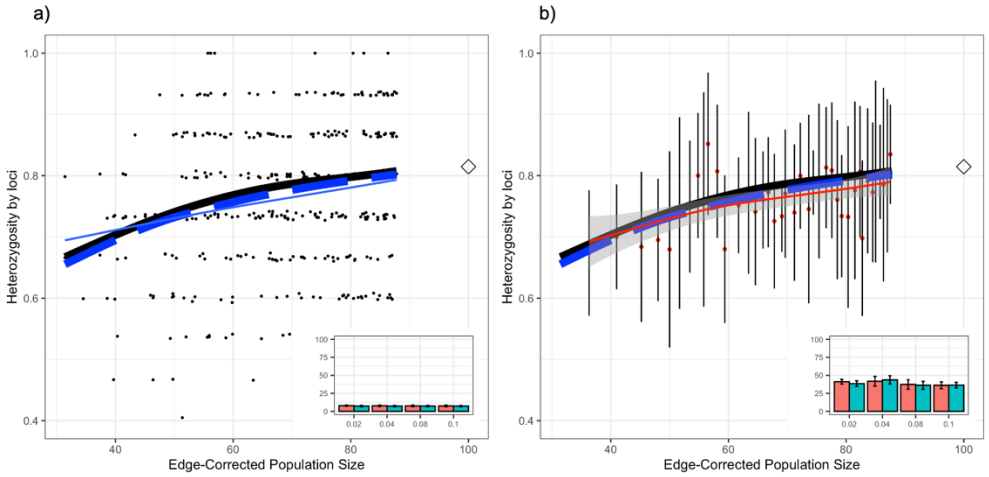


Figure S11. Effect of edge-corrected population size data binning on individual-based heterozygosity. a) raw individual-based heterozygosity; **b)** individual heterozygosity grouped in 40 bins of equal sample size based on the N_{edge} value. Heterozygosity by loci was measured across 30 microsatellites by sampling 5 individuals per deme in 'scenario 1' from Fig. S1 (d_{tot} = 382 demes) with K: 100, m: 0.04 and μ: 5 x 10^{-4}. Data points refer to one single simulation replicate. Dashed blue and solid black line indicate, respectively, theoretical results based on the estimated deme N_{edge} and average value across 30 simulations. Solid thin line is the GAM model obtained from one simulation replicate. The insert bar plots quantify the percentage of variance explained by the 'edge-corrected population size' under the Wright-Malécot approximation (light blue) and GAM model (red). White diamonds indicate the theoretical H_o using the Wright-Malécot approximation and ignoring edge-effect.

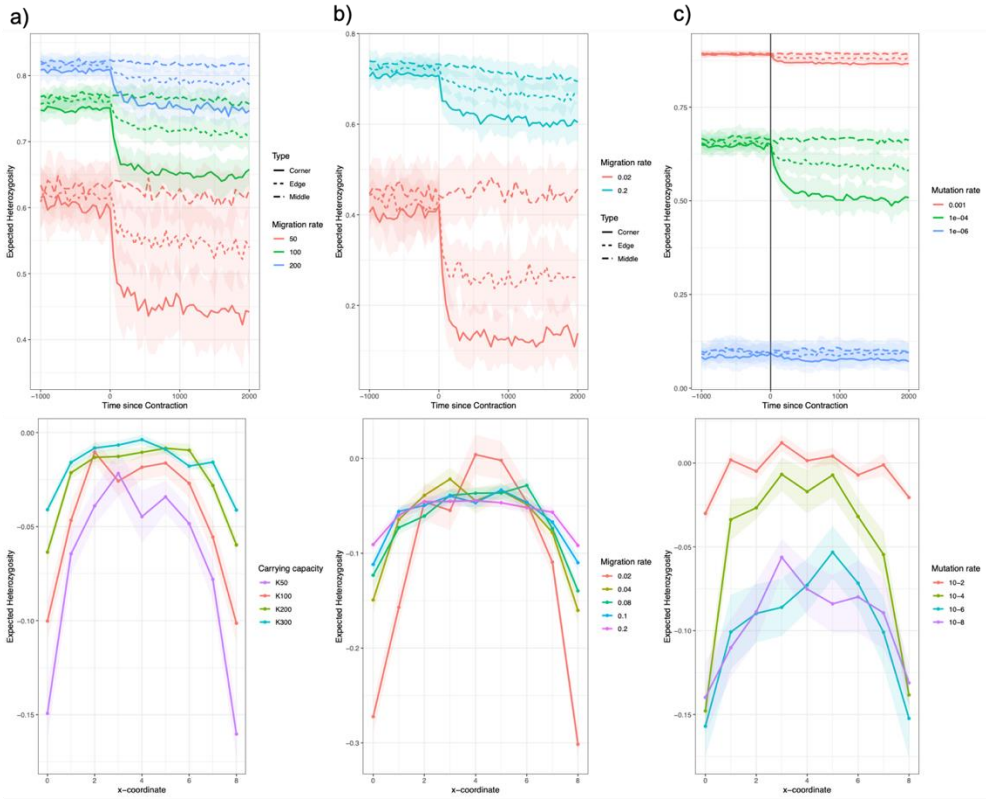


Figure S12. Dynamic of deme H_0 in a 13 x 13 stepping-stone model undergoing habitat contraction. Impact of **a)** carrying capacity (fixed parameters, m: 0.04; μ: 5 x 10^{-4}); **b)** migration rate (fixed parameters, K: 100; μ: 5 x 10^{-4}); and **c)** mutation rate (fixed parameters, K: 100; m: 0.04) on temporal changes in deme H_0. The first-row shows $H_0(t)$ for centre, edge and corner demes. The second row represents the difference in deme H_0 before and after contraction for demes sampled along the diagonal. Across all parameter's combinations, we sampled 14 individuals per deme, 30 microsatellite per individual. Dots and shade indicate, respectively, the mean and standard deviation of deme H_0 based on 30 simulation replicates.

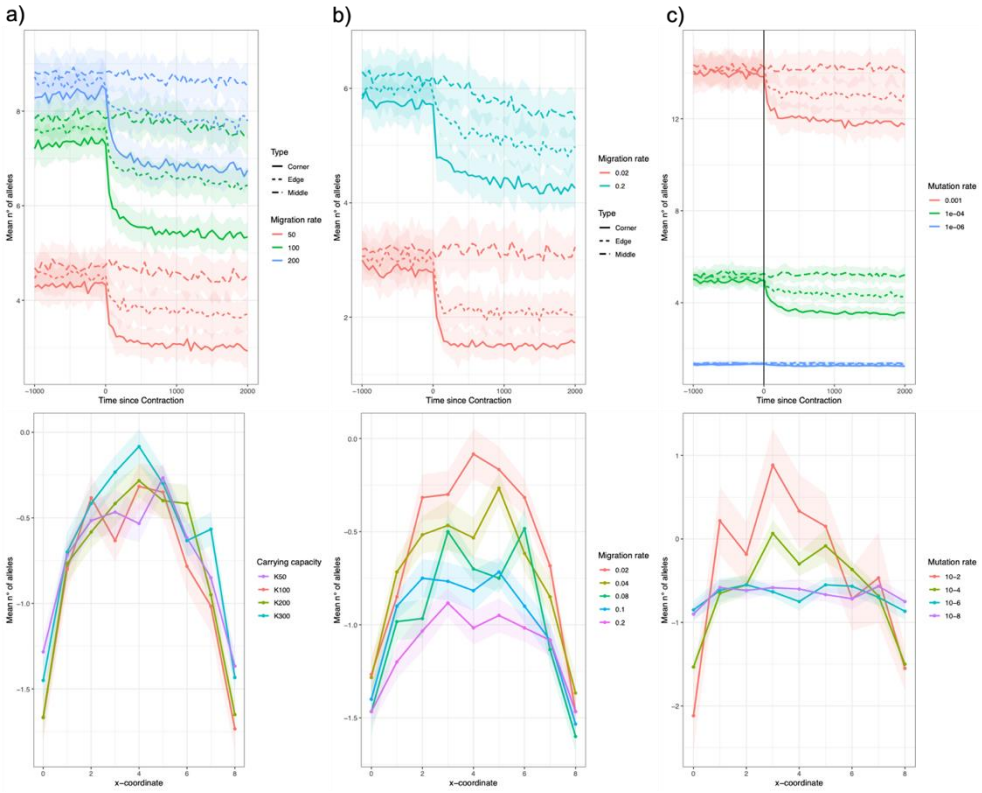


Figure S13. Dynamic of deme H_0 in a 13 x 13 stepping-stone model undergoing habitat contraction. Impact of **a)** carrying capacity (fixed parameters, m: 0.04; μ: 5 x 10^{-4}); **b)** migration rate (fixed parameters, K: 100; μ: 5 x 10^{-4}); and **c)** mutation rate (fixed parameters, K: 100; m: 0.04) on temporal changes in deme H_0. The first-row shows $H_0(t)$ for centre, edge and corner demes. The second row represents the difference in deme H_0 before and after contraction for demes sampled along the diagonal. Across all parameter's combinations, we sampled 14 individuals per deme, 30 microsatellite per individual. Dots and shade indicate, respectively, the mean and standard deviation of deme H_0 based on 30 simulation replicates.

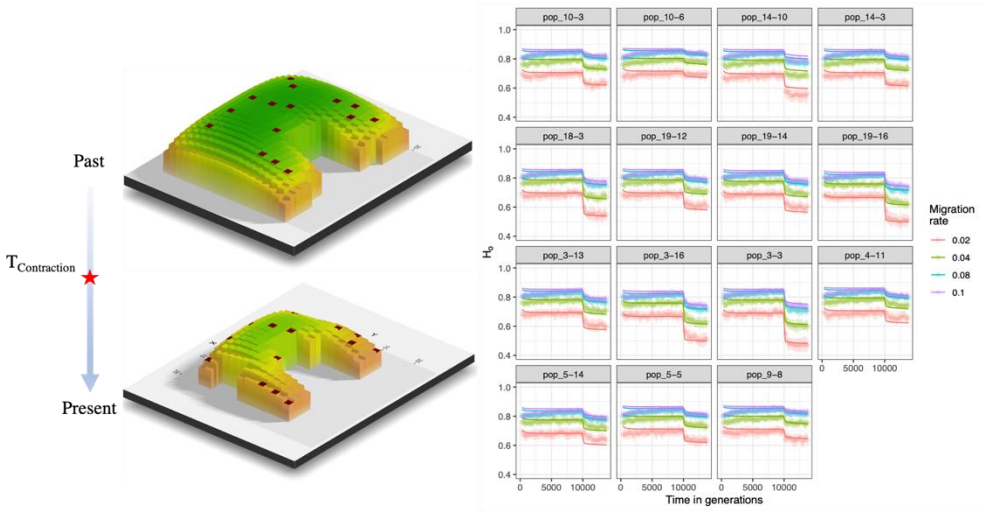


Figure S14. Non-equilibrium deme H_0 in habitat patch (scenario 1) undergoing instantaneous habitat contraction. Left column shows the simulated scenario. Colours and deme bars are proportional to the amount of edge-effect at which each deme is subjected to. Brown squares indicate sampled demes. Right column presents the comparison between simulation and analytical results for different values of migration rate, across demes subjected to various level of edge-effect. The solid lines represent theoretical predictions based on numerical integration of (*Eq. 8d*). Simulation results are represented by coloured dots and refer to the average deme H_0 across 30 simulation replicates, computed by sampling 14 individuals per deme at 30 microsatellite loci (fixed parameters, K: 100; μ: 5 x 10^{-4}). The results indicate good agreement between theoretical predictions and simulation results.

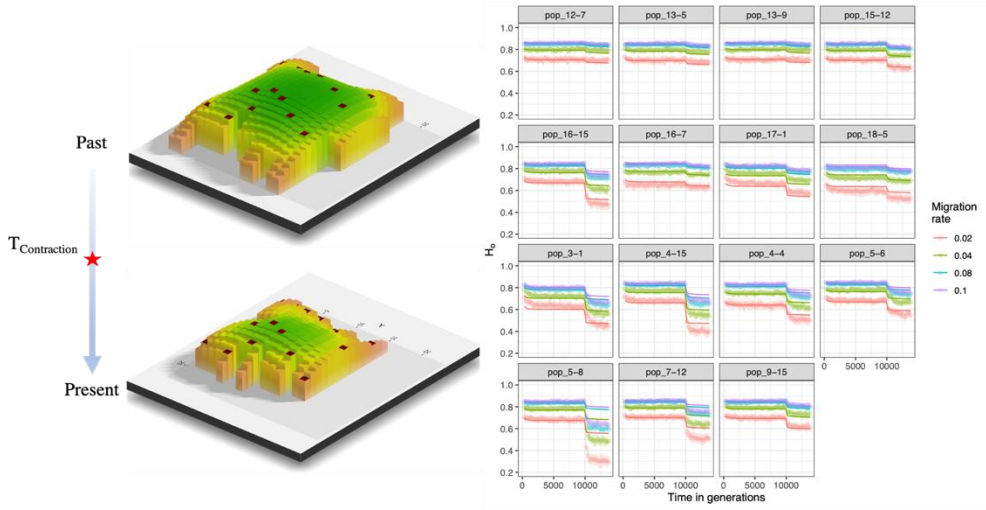


Figure S15. Non-equilibrium deme H_0 in habitat patch (scenario 2) undergoing instantaneous habitat contraction. Left column shows the simulated scenario. Colours and deme bars are proportional to the amount of edge-effect at which each deme is subjected to. Brown squares indicate sampled demes. Right column presents the comparison between simulation and analytical results for different values of migration rate, across demes subjected to various level of edge-effect. The solid lines represent theoretical predictions based on numerical integration of (*Eq. 8d*). Simulation results are represented by coloured dots and refer to the average deme H_0 across 30 simulation replicates, computed by sampling 14 individuals per deme at 30 microsatellite loci (fixed parameters, K: 100; μ: 5 x 10^{-4}). The results indicate good agreement between theoretical predictions and simulation results, except in few instancew (i.e., pop_5-8).

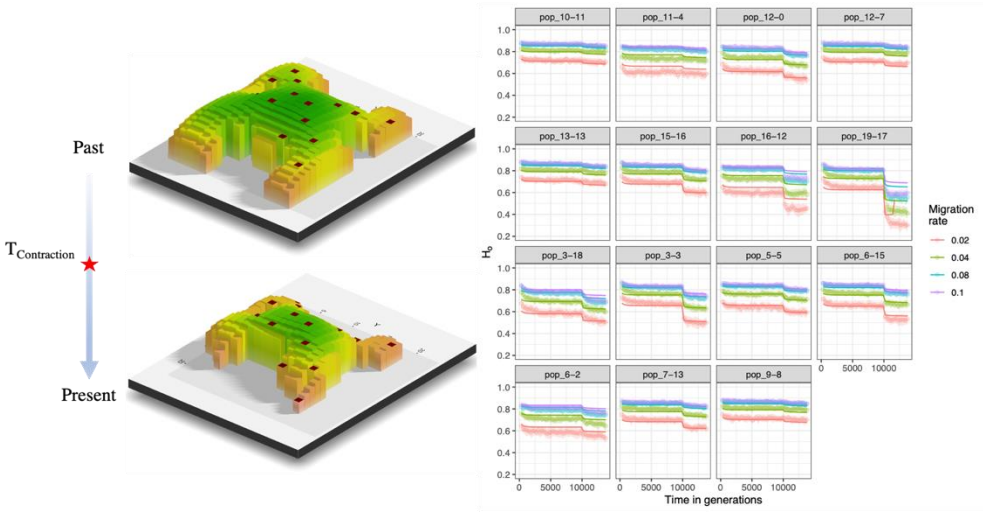


Figure S16. Non-equilibrium deme H_0 in habitat patch (scenario 3) undergoing instantaneous habitat contraction. Left column shows the simulated scenario. Colours and deme bars are proportional to the amount of edge-effect at which each deme is subjected to. Brown squares indicate sampled demes. Right column presents the comparison between simulation and analytical results for different values of migration rate, across demes subjected to various level of edge-effect. The solid lines represent theoretical predictions based on numerical integration of (*Eq. 8d*). Simulation results are represented by coloured dots and refer to the average deme H_0 across 30 simulation replicates, computed by sampling 14 individuals per deme at 30 microsatellite loci (fixed parameters, K: 100; μ: 5 x 10^{-4}). The results indicate good agreement between theoretical predictions and simulation results.

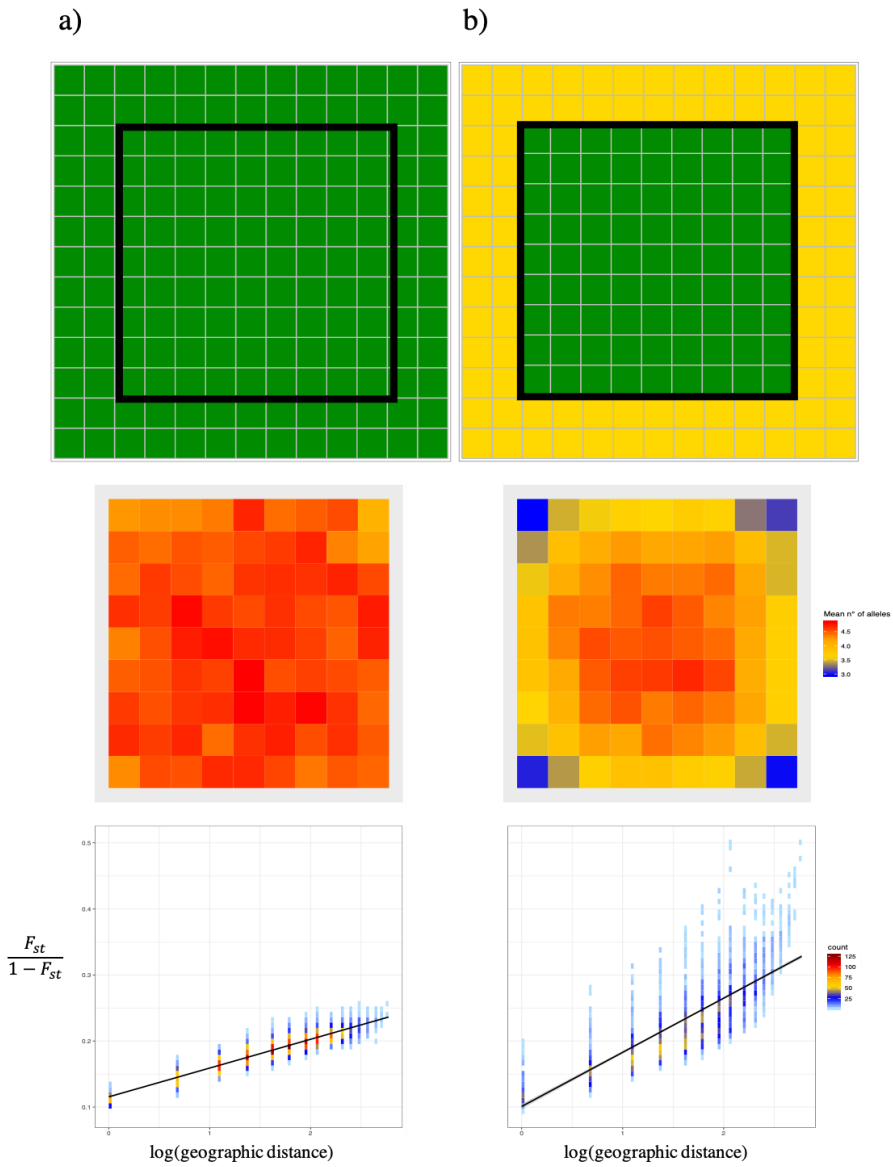


Figure S17. The influence of edge-effect on genetic diversity and differentiation. a) Before contraction; b) After contraction. First row: simulated scenario (13 x 13 demes) with $K = 100$, $m = 0.04$ and $\mu = 10^{-5}$, where dark green or yellow cells refer to habitat or no habitat. Dark line delimits the sampled demes. Second row: distribution of deme genetic diversity (MNA) across sampled demes. Third row: isolation-by-distance (IBD) pattern across all pairwise comparisons. The results show higher IBD slope after habitat contraction due to increase in edge effect. Five diploid individuals were sampled from each cell and simulated 30 microsatellite loci.

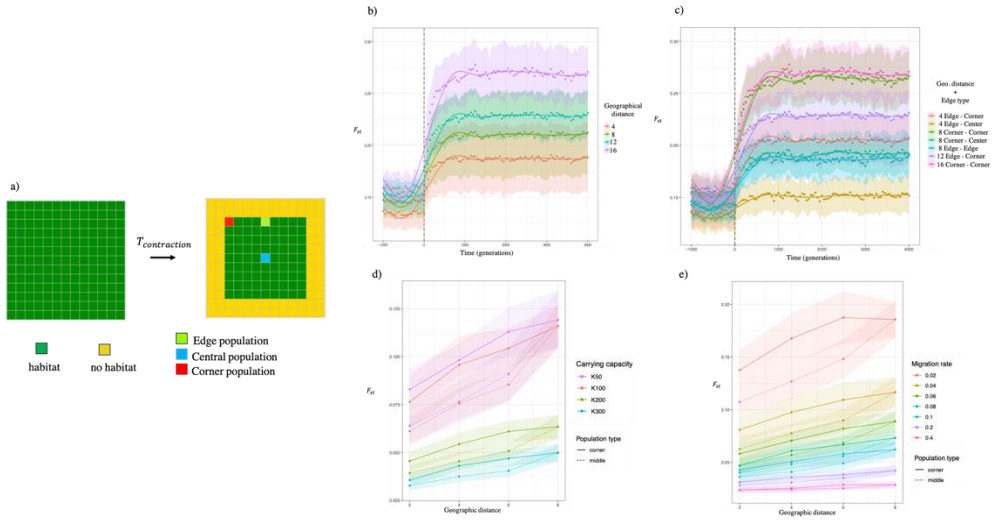


Figure S18. The influence of edge-effect on isolation-by-distance. a) simulated scenario (13 x 13 demes), where dark green or yellow cells refer to habitat or no habitat, while red, blue and light green cells correspond to corner, edge and central demes, respectively; b) Temporal dynamic of pairwise F_{st} for deme pairs at different geographical distances (4, 6, 12, 16), dashed line represents the time of contraction; c) Temporal dynamic of pairwise F_{st} for deme pairs at different geographical distances (4, 6, 12, 16) and including different edge types (Center, Edge, Corner), dashed line represents the time of contraction; d) the effect of 'Population type' on isolation-by-distance pattern, for several values of carrying capacity. Solid or dashed lines refer to pairwise comparisons including middle (=center) or corner demes, respectively. Panel b) and c) show that 'Edge type' can influence F_{st} as much as geographical distance, as shown for comparisons at distance 8. Panel d) and e) show that F_{st} comparisons including corner demes are higher than center ones.

Chapter 3

3 The effect of habitat loss and fragmentation on isolation-by-distance and time

Gabriele Maria Sgarlata[1], Tiago Maié[1], Tiago de Zoeten[1], Rita Rasteiro[1], Lounès Chikhi[1, 2]

1. Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, 2780-156 Oeiras, Portugal;
2. Laboratoire Evolution & Diversité Biologique (EDB UMR 5174), CNRS, IRD, UPS, Université de Toulouse Midi-Pyrénées, Toulouse, France;

AUTHOR CONTRIBUTIONS: GMS, LC and TM designed the project. LC, RR and TM designed the genetic simulation framework, TM and RR developed the genetic simulation framework, GMS performed spatial genetic simulations and data analyses, and developed the numerical framework. TZ contributed to data analysis. GMS drafted the manuscript and LC revised it.

Highlights

- Spatially-explicit simulations show that 'isolation-by-distance' (IBD) pattern among isolated habitat fragments can be maintained for thousands of generations after instantaneous habitat loss and fragmentation (HL&F)
- Gradual habitat fragmentation and range expansion prior to instantaneous habitat fragmentation are two processes commonly observed in nature that significantly increase the persistence of IBD pattern after HL&F
- We propose a theoretical model of *isolation-by-distance in time* in the context of HL&F, obtained by modifying the original theorical results on isolation-by-distance (Slatkin, 1991; Slatkin, 1993) and apply them to a toroidal stepping-stone model in the context of HL&F
- Simulation and numerical results show that dispersal rate and habitat fragment size are the two most important factors influencing the persistence of IBD pattern, and therefore the relative contribution of geography on pairwise F_{st}, after HL&F

3.1 Abstract

Throughout Earth's natural history, habitats have undergone drastic changes in quality and extent, influencing the distribution of species. In the last few hundred years, human activities have destroyed natural habitats at an unprecedent rate, converting continuous habitat into fragmented and isolated patches. Recent global metanalyses suggest that habitat loss and fragmentation (HL&F) has negatively impacted the genetic diversity of many taxa across the world. These conclusions have been drawn by comparing present-day genetic patterns from populations occurring in continuous and fragmented landscapes. In this work, we attempted to go beyond 'pattern' and investigate through simulations some of the 'processes' that influence genetic variation in the context of HL&F. Since most species have a geographically restricted dispersal (known as "isolation-by-distance", IBD), we studied the impact of HL&F on isolation-by-distance. We characterised the behaviour of IBD in the case of i) instantaneous HL&F, ii) gradual (two-steps) HL&F, and iii) instantaneous HL&F following range expansion. In addition, we propose a spatially-explicit theoretical framework by modifying the original theorical results on isolation-by-distance (Slatkin, 1991; Slatkin, 1993) and apply them to a toroidal stepping-stone model in the context of HL&F. Our results suggest that isolation-by-distance can be maintained for relatively long time after HL&F, thus pointing to the long-term importance of spatial genetic structure in species genetic diversity. In addition, our results may explain why present-day fragmented population still show significant IBD pattern although being disconnected.

3.2 Introduction

Environmental changes have influenced the habitat of species worldwide, across both wide and short spatial and temporal scales (Hewitt, 2000; Newbold et al., 2015; Ceballos et al., 2017; Cooper et al., 2021). Consequently, species' habitats have contracted and expanded over hundreds, thousands of years. At relatively small spatial scales, such events have reduced and/or increased connectivity among populations of the same species. Reconstructing the natural history of species and landscapes holds an interest on its own, but in addition it could provide useful insights on how landscapes have changed and how species have responded to such changes (Fordham et al., 2016; Román-Palacios and Wiens, 2020; Mitchell and Rawlence, 2021). For instance, having independent sources of information on how a given landscape has changed in the recent past (i.e., palynological, geological data, etc..) and how a studied species has changed its distribution during the same time scale (i.e., fossils, sub-fossils data, etc..) might be informative on how that species may respond to habitat loss and fragmentation (Hewitt, 1996; Barak et al., 2016; Lyman, 2017; Grace et al., 2019). This is reasonable under the assumption of 'niche conservatism', defined as the species retention over time of ecological traits related to its current niche (Wiens et al., 2010). In other words, under the assumption that the present habitat preferences of the studied species are the same as in the past. One could ask why not carry out lab experiments to investigate species response to habitat and environmental change. Indeed, some species can be relatively easy to manipulate in laboratory or common garden experiment (e.g., Cayuela et al., 2018; Haddad et al., 2015; Exposito-Alonso et al., 2018). However, not all species can be brought in a lab or are easy to manipulate. Moreover, responses to long-time scale processes or to the variety of environmental changes that have occurred in nature might not be reproducible in the life-span of few human researchers (Kawecki et al., 2012;

Bailey and Bataillon, 2016). Thus, alternative, but complementary, approaches and sources of information might be necessary to address these questions.

Genetic data are one of the sources of information that can uncover the past of species (Nielsen and Beaumont, 2009; Beichman et al., 2018; Mitchell and Rawlence, 2021; Loog, 2021). Reconstructing the history of species can provide useful insights in conservation biology, by contributing to our understanding of species response to habitat loss and fragmentation (HL&F), a major threat to biodiversity especially in the last decades (Haddad et al., 2015; Silva Junior et al., 2020; Vancutsem et al., 2021). Species demographic history can also elucidate on the causing drivers of ancient HL&F and/or past fauna extinction, by comparing genetic-based inference with archaeological, fossil and paleoenvironmental data. This is particularly relevant for the ongoing debate on to which extent natural environmental change and/or human activities have shaped present biodiversity or contributed to past megafauna extinctions (Sandom et al., 2014; Saltré et al., 2016; Louys et al., 2021; Stewart et al., 2021).

Several approaches have been developed for making sense of observed genetic pattern and infer the causal processes (e.g., Beichman et al., 2018). Some statistical approaches compare observed patterns of genetic variation with those produced by several hypothetical models of species demographic history, and the model that best-fits with the observed data is chosen (Beaumont et al., 2002; Gutenkunst et al., 2009; Excoffier and Foll, 2011). Since these are models, which are simplistic but useful representation of the real world, there is no certainty on their accuracy in relation to the real demographic history of the studied species. Yet, they allow to identify the most likely model(s) among those considered.

There are frameworks assuming that populations are not structured (e.g., Beaumont, 1999; Li and Durbin, 2011; Nikolic and Chevalet, 2014; Liu and Fu, 2015), namely that individuals reproduce at random and that mating can occur with equal probability with any of the individuals in the population.

Other methods consider explicitly that species might be subdivided in smaller unit of panmictic subpopulations (‘islands’ or ‘demes’), connected with other subpopulations by gene flow (e.g., Nielsen and Wakeley, 2001; Hey and Nielsen, 2004; Gutenkunst et al., 2009; Costa and Wilkinson-Herbots, 2017; Arredondo et al., 2021). However, relatively few inferential approaches that explicitly take into account *spatial population structure* have been developed (e.g., Rousset and Leblois, 2012; Barton et al., 2013; Ralph and Coop, 2013; Guindon et al., 2016).

Many empirical studies have shown that, across several species, the movement of individuals is limited by geographical distance, meaning that individuals tend, on average, to reproduce at relatively short-distances from where they were born (Meirmans, 2012; Sexton et al., 2014; Aguillon et al., 2017). These observations suggest, then, that *spatial population structure* is probably the most common form of population structure in the wild. Geographically-limited dispersal generates a spatial genetic pattern called ‘isolation-by-distance’ (IBD) (Wright, 1943; Malécot, 1948). It is typically visualized by correlating $F_{st}/(1-F_{st})$ with the natural logarithm of geographical distance (Rousset, 1997, 2000), but similar other approaches do exist (Kimura and Weiss, 1964; Hardy and Vekemans, 1999; Vekemans and Hardy, 2004; Cayuela et al., 2018). The processes that can lead to IBD pattern in natural populations are being continuously investigated since the seminal works of Wright, (1940, 1943, 1946) and Malécot, (1948), and several models have been proposed to explain genetic variation over short (Barton and Wilson, 1995; Wilkinson-Herbots, 1998; Barton et al., 2002; Cox and Durrett, 2002) and large spatial scales (Barton et al., 2010).

Implementations of spatial genetic models typically assume continuous habitat and, homogenous population density and dispersal. Some exceptions are, for instance, Robledo-Arnuncio and Rousset, (2010), which derived exact recursions for probabilities of identity in spatially continuous populations, accounting for fluctuation in local population density, and Barton, (2008), which constructed a theoretical framework describing the effect of a

barrier on isolation-by-distance pattern. More recent studies have open new avenues of research in spatial population genetics, going beyond some of the assumptions initially formalized in the classical 'isolation-by-distance' model (e.g., McRae, 2006), and recovering information on genetic ancestry at finer geographical- and temporal-scale (Ralph and Coop, 2013; Ringbauer et al., 2017). For instance, Duforet-Frebourg and Slatkin, (2016) have recently introduced the theory of isolation-by-distance and time (IBDT), a generalization of the theory of IBD, by allowing differences in sampling time of the genetic data. In particular, they extended previous results by Kimura and Weiss, (1964) on the correlation of allele frequency and by Slatkin, (1991, 1993) on expected coalescence time, accounting for both geographical distance and time. Ringbauer et al., (2018) have proposed an analytical framework that model isolation-by-distance patterns when a barrier is present in a two-dimensional population. Other studies have proposed alternative statistical frameworks for modelling genetic variation in space (and in few cases in time) (e.g., Baharian et al., 2016; Petkova et al., 2016; Bradburd et al., 2018; Al-Asadi et al., 2019; Racimo et al., 2020). These recent studies reveal a general trend in spatial population genetics which aims at improving our understanding of the distribution of genetic variation in space and time (Bradburd and Ralph, 2019).

The importance of accounting for population structure in population genetic models has been acknowledged in several theoretical studies (Städler et al., 2009; Chikhi et al., 2010; Mazet et al., 2016; Chikhi et al., 2018). For instance, Wakeley, (2000) and Arbogast et al., (2002) have shown that population structure (e.g., spatial and non-spatial models) influences genealogical structure, even at high dispersal rate, in a way that differ from an 'equivalent' non-structured model. In particular, Wakeley, (2000) showed that dispersal within subdivided populations alters both the genealogical topology and the time scale of the coalescent process, such that genetic-based approaches ignoring population structure do overestimate the divergence time among subdivided populations. Such effect becomes even

stronger with lower dispersal rate. Other theoretical studies have further confirmed these conclusions (e.g., Wilkins and Wakeley, 2002; Wilkins, 2004). Accordingly, several studies have shown that ignoring spatial genetic structure can have strong implication for demographic inference, estimates of genetic diversity and genome-wide association studies (GWAS) (Mazet et al., 2016; Chikhi et al., 2018; Battey et al., 2020).

In the present study, we investigated non-equilibrium IBD patterns in a two-dimensional population (*landscape*) undergoing HL&F. This work was motivated by the observation that genetic data sampled across fragmented subpopulations can show significant IBD signatures, although violating the assumption of continuous habitat. For instance, Schneider et al., (2010) correlated pairwise *Phi-st* with geographical distance from genetic data of grey mouse lemurs (*Microcebus murinus*) collected either in connected or fragmented subpopulations. They found significant IBD pattern in both connected and fragmented subpopulations, with an IBD steeper and characterised by higher pairwise *Phi-st* values in fragmented populations than in connected ones. Similar examples can be found, for instance, in northern rufous mouse lemurs in northern Madagascar (Aleixo-Pais et al., 2019), marsh fritillary in Denmark (Pertoldi et al., 2021), white carob tree in Argentina (Roser et al., 2017), Asian rice in China (Zhao et al., 2013), in blue duck in New Zealand (Grosser et al., 2017) and in the long-lived subalpine conifer in western Alps (Tóth et al., 2019). These observations may suggest that genetic data keep 'memory' of a past in which the landscape was continuous and not-fragmented. To put it in another way, we could say that not enough time has passed since HL&F, such that genetic patterns have not yet been altered by the new landscape configuration. In fact, according to Hutchison and Templeton (1999), we would expect a lack of correlation between genetic and geographical distance, due to the stronger influence of drift compared to gene-flow among fragmented populations (case III in Hutchins and Templeton, 1999).

The idea that genetic data may take some generations since an environmental change in order to show pattern that represent contemporary landscapes (lag-time) has been discussed in several occasions in the literature (e.g., Keyghobadi et al., 2005; Landguth et al., 2010; Epps and Keyghobadi, 2015). For example, Landguth et al., (2010) showed that F_{st} between populations separated by a barrier can take few hundred generations since the appearance of a barrier before reaching the new equilibrium value. Similarly, Mona et al., (2014) simulated a two-dimensional grid of demes subdivided in habitat patches, among which gene flow was impaired (habitat fragmentation), and found that, in their simulated scenario, habitat fragmentation required over 100 generations to significantly influence pattern of population structure (F_{st}). Thus, genetic time-lag could explain the discrepancy between present-landscape configuration and the spatial genetic pattern observed in real fragmented populations.

Additional spatial processes occurring during the evolutionary history of species could enhance possible inconsistencies between contemporary landscapes and detected genetic structure. For instance, Mona et al., (2014) explored the effect of habitat fragmentation in a species undergoing range expansion, showing that recent range expansion can further increase population structure compared to a fragmented landscape at equilibrium (no expansion). It is indeed well-established that range expansion processes can intensify spatial differences in genetic diversity and differentiation compared to a continuous population at equilibrium (Ibrahim et al., 1996; Edmonds et al., 2004; Hallatschek et al., 2007; Excoffier and Ray, 2008). In conclusion, the pervasive impact of geography-limited dispersal across many wild species and the time-lag between contemporary landscape features and observed genetic structure point out to the necessity of including both spatial and temporal processes when investigating the impact of HL&F on genetic diversity.

In this study we used a simulation-based approach to explore the behaviour of isolation-by-distance over time under different scenarios of

habitat loss and fragmentation. More specifically, we ask *i)* what is the spatio-temporal dynamic of isolation by distance in a two-dimensional landscape undergoing *instantaneous HL&F*, and *iii)* how range expansion and variation in tempo and mode of HL&F influences the persistence of IBD patterns. We conclude proposing a theoretical framework that consider both the effect of space and time in shaping population genetic variation in the context of HL&F. In particular, we modify the original theorical results on isolation-by-distance (Slatkin, 1991, 1993) and apply them to a toroidal stepping-stone model undergoing habitat loss and fragmentation. Our analytical model opens the possibility for future development of an inferential-approach allowing to interpret F_{st} pattern considering both space and time.

3.3 Materials and Methods

3.3.1 Population genetics simulations

We used the computer program SINS (Simulating INdividuals In Space; Rasteiro et al., 2012) to model a two-dimensional landscape experiencing habitat loss and fragmentation. SINS is an individual-based software that uses a forward-time approach to simulate diploid individuals (males and females) over a 2D grid of demes, similar to a classical 2D stepping-stone model (Kimura and Weiss, 1964) and inspired by the original coalescent-based software SPLATCHE (Currat et al., 2004). SINS and SPLATCHE share several features, but they are also complementary since they differ in some properties. First of all, SINS is a forward-in time simulator, whereas SPLATCHE simulate backward in time. Also, SINS can simulate several types of molecular markers at the same time and to account for variance in reproductive success, whereas this is not possible with SPLATCHE2. More details on the comparison between SINS and SPLATCHE can be found in Rasteiro et al., (2012).

In SINS, deme population size ($\widehat{N}_{i,t}$) is drawn from a Poisson distribution with a mean value ($N_{i,t}$) given by a corrected form of the (Maynard-Smith and Slatkin, 1973) logistic growth. The $N_{i,t}$ value in each deme is logistically regulated by deme-specific carrying capacity (K), growth rate (r) and deme number of reproductive females (N_f). Mating is random within demes, namely one reproductive individual from each sex is randomly chosen and generate a new individual (offspring) which inherit, for each locus, one allele from each parent. At each generation, mating occurs iteratively until $\widehat{N}_{i,t}$ is reached. Each deme exchanges migrants with the four neighbouring demes. The number of individuals that will emigrate from the focal deme at each generation is drawn from a Poisson distribution with mean $M = N_{i,t} m\, n_d/4$, where m is the dispersal rate and n_d is the number of available

neighbouring demes, which is equal to 2 for a corner deme and 4 for central deme. The number of migrants will then be stochastically distributed among the neighbouring demes using a binomial distribution with probability $P_{dir} = (1 - F_{dir})/(n_d - F_t)$, where dir is one of the available directions (min=1; max=4), F_{dir} is a deme-specific friction parameter defining the difficulty to move from the focal deme into that direction and F_t is the sum of the frictions of the n_d receiving demes. Similar to SPLATCHE, changes from one landscape configuration to another (i.e., continuous to fragmented) are defined by K and F maps describing, respectively, the carrying capacity and friction value of each deme in the two configurations. Demes that are unsuitable after an 'environmental event' are set with K = 0 and F = -99, such that no individuals can grow and no migrants can move within these unsuitable demes. The growth rate was set to 0.8, meaning that fraction of the carrying capacity available for further growth will generate 0.8 offspring per generation.

SINS is able to simultaneously simulate demographic and genetic data over non-overlapping discrete generations and through space. Several types of molecular markers (sequences, SNPs and microsatellites) and chromosome objects (X and Y chromosomes, autosomes and mitochondrial DNA) can be simulated. In the present study we will exclusively use independent microsatellite loci, which mutations are modelled according to the Stepwise Mutation Model (Ohta and Kimura, 1973). Simulations start with all demes having the same initial genetic diversity, which is defined by a file for each locus specifying the initial number and frequency of alleles for the simulated markers. We set initial allele frequency for each locus by sampling 100 individuals from a Wright-Fisher (WF) population simulated in *fastsimcoal* (Excoffier and Foll, 2011) using the *fastsimcoal* function in the *strataG* R package (Archer et al., 2017). The size of the WF population was set equal to 10,000 and mutation rate according to the value specified in the SINS simulation. We decided to define the initial deme allele frequencies in order

to ensure that dispersal-mutation-drift equilibrium would be reached quicker than if we had to wait for mutations to occur.

3.3.2 Simulated scenarios

Instantaneous HL&F: We simulated a two-dimensional grid of 13 x 13 demes (d_{tot} = 169 demes), with homogenous deme-population size (K) and deme-specific friction (F). After 10,000 generations, more than the time required for reaching a 'quasi-genetic equilibrium' (i.e., $\Delta He/t \approx 0$), the landscape is subjected to habitat loss and fragmentation, leading to nine remaining patches of demes (Fig. 1a). After that, simulations were run for over 4,000 generations. We tested several values of K (50, 100, 200) and m (0.02, 0.04, 0.06, 0.1), which were kept constant across all demes in the landscape. We also tested several values of mutation rate, μ (5 x 10^{-3}, 5 x 10^{-4}, 5 x 10^{-5}, 5 x 10^{-6}) and number of genetic markers (15, 30, 100, 400 microsatellites).

Range expansion prior to instantaneous HL&F: Range expansion is an evolutionary process that has shaped the genomic diversity of many taxa across the world (Hewitt, 2000; Eckert et al., 2008; Pauls et al., 2013). Interpreting spatial genetic pattern in the context of HL&F requires thus to understand how prior range expansion genetic signatures influence genetic pattern generated by HL&F. The simulated scenario is equivalent to the *instantaneous HL&F*, apart from the initial conditions. The 13 x 13 grid is initially unoccupied by individuals except in the lower-left corner deme. From such deme, individuals are sent to the neighbouring demes according to the simulated dispersal rate (m). We assessed the impact of past range expansion before the instantaneous HL&F event by setting the time of HL&F at different time points since the onset of the range expansion (the starting of the simulation; Δt_{exp}): 750, 1000 and 1500 generations (Fig. 2a). The

minimum Δt_{exp} value was chosen such to ensure complete colonisation of the 13 x 13 habitat before HL&F, across all tested parameters.

Two-steps HL&F: Habitat loss and fragmentation may not necessarily occur instantaneously in real landscape. Instead, complex landscape cover changes are more likely to occur, where habitat fragments are formed gradually following fine-scale spatial-temporal processes. Therefore, we assessed the impact of gradual HL&F by simulating a scenario where HL&F occurs as a two-steps event. The initial two-dimensional habitat is equivalent to the initial conditions of the *Instantaneous HL&F* scenario, before HL&F. After 10,000 generations, the landscape is subject to a first HL&F event ($T_{1^{st}HL\&F}$), forming three large isolated habitat fragments that undergo a second HL&F event after Δt_{2nd} generations. The second HL&F event produces a fragmented landscape with nine patches, equivalent to the landscape configuration in the *instantaneous HL&F* scenario (Fig. 2b). We considered three values of Δt_{2nd}: 500, 1000 and 1500 generations.

Across all simulations, we sampled 14 individuals at 30 microsatellites from each remaining patch (N_{tot}: 14 x 9 = 126 individuals), before and after HL&F. Two sampling strategies are considered: *local sampling*, where patch-sampled individuals come from the same deme; and *random sampling*, where patch-sampled individuals are randomly collected across demes within the patch. We ran 100 independent simulation replicates for each scenario and recorded genetic data every 50 generations.

3.3.3 Measuring spatial genetic patterns

Pairwise genetic differentiation was estimated using the R_{st} statistic, since it is more appropriate for microsatellite data mutating according to the Stepwise Mutation Model (Slatkin, 1995). We computed R_{st} using the *polysat*

R package (Clark and Jasieniuk, 2011). Genetic differentiation was estimated for each independent microsatellite locus and then averaged across loci. Isolation-by-distance (IBD) was estimated by correlating pairwise $R_{st}/(1-R_{st})$ with the log-transformed geographic distance as proposed in Rousset, (1997) for two-dimensional habitats. The significance of IBD was assessed using Mantel test. Pairwise geographic distances were calculated using the *euclidean* distance. For comparison, we also estimated genetic differentiation using Wright's F_{st}.

The quantification of the time of IBD loss (T_{IBD}) was carried out on the time series IBD p-value for each simulation replicate by using a changepoint method based on PELT-algorithm, implemented in the *changepoint* R package (Killick and Eckley, 2014). It estimates the point at which the statistical properties of a series of observations (p-value in this case) change. We used the *cpt.mean* function (Killick et al., 2011) to identify a single change point of the mean signal over time. The function segments the data in intervals of a user-defined size (penalty value) and, for each interval, computes the average data value. Different penalty values (1 – 10) are iteratively tested until a detection point is found (T_{IBD}).

Relative Weights Analysis (RWA) was used to quantify the relative contribution of geography and time since HL&F on pairwise R_{st} (or F_{st}). RWA is a statistical method that computes the relative importance of predictor variables on the response variable of a regression model, accounting for possible correlations among predictor variables (Tonidandel et al., 2009). RWA was carried out using the *rwa* R package (Chan, 2022).

3.3.4 A toroidal stepping-stone model under HL&F

Isolation-by-distance can be visualized by linear regression of pairwise F_{st} (or R_{st} if using microsatellites) between subpopulations and geographical distance. Alternatively, but formally more appropriate, $\eta = F_{st}/(1-F_{st})$ can be used instead of F_{st} (Rousset, 1997). Under the infinite

sites mutation model, pairwise F_{st} can be expressed in terms of the mean number of differences between sampled DNA sequences, called Hudson's F_{st} estimator (Hudson et al., 1992). Thus, given two subpopulations A and B, we can estimate pairwise F_{st} as:

$$F_{st} = 1 - \frac{H_w}{H_b} \qquad \text{(Eq. 1)}$$

where H_w is the mean number of differences between DNA sequences sampled within subpopulations, $H_w = (H_w^A + H_w^B)/2$, and H_b is the mean number of differences between DNA sequences sampled between subpopulations.

Earlier work by Slatkin, (1991) has shown that F_{st} can be expressed in terms of average coalescence times for a large variety of structured models, included a toroidal 2D stepping-stone, one type of spatial structured population models. According to the derivation of Slatkin, (1991), pairwise F_{st} for two alleles in subpopulations A and B at *i* and *j* steps apart, respectively, in the two directions *x* and *y* can be expressed as:

$$F_{st(i,j)} = \frac{\bar{t}_{ij} - \bar{t}_0}{\bar{t}_{ij}} \qquad \text{(Eq. 2)}$$

where $\bar{t}_{ij}$ is the average coalescence time for two alleles sampled *i* and *j* steps apart, and $\bar{t}_0$ is the average coalescence time for two alleles sampled within the same subpopulation. The equivalence between *Eq. 1* and *Eq. 2* becomes clearer if we consider that expected $\bar{H}_w$ and $\bar{H}_b$ can be expressed in terms of average coalescence times between two randomly sampled DNA sequences. In particular, $\bar{H}_w \approx [(2\mu\bar{t}_A + 2\mu\bar{t}_B)/2]$ and $\bar{H}_b \approx 2\mu\bar{t}_{AB}$, where $\bar{t}_A$ and $\bar{t}_B$ are the average coalescence times for two alleles randomly sampled within subpopulation A and subpopulation B, respectively, while $\bar{t}_{AB}$ is the average coalescence times for two alleles sampled in A and B (Slatkin and Hudson,

1991; Ortega-Del Vecchyo and Slatkin, 2019). Thus, $\bar{t}_{ij} = \bar{t}_{AB}$ and $\bar{t}_0 = 2(\bar{t}_A + \bar{t}_B)/2$. Assuming that subpopulation size is homogeneous across the toroidal 2D stepping-stone model, then $\bar{t}_A = \bar{t}_B$ and $\bar{t}_0$ simplifies to $\bar{t}_0 = \bar{t}_A = \bar{t}_B$. An important consideration to be made is that these notations are derived under the assumption of small per-site mutation rate μ (Slatkin, 1991).

In the case of a 2D stepping-stone model, it has been shown that, under a general model of dispersal, $\bar{t}_0$ is independent from dispersal (m) and it can be estimated by only considering the total number of individuals in the population, $\bar{t}_0 = 2N_e d_x d_y$, where N_e is the subpopulation size and d_x or d_y are the number of subpopulations in dimension x and y, respectively (Strobeck, 1987; Hey, 1991). Instead, $\bar{t}_{ij}$ can be expressed by dividing the coalescence process in two phases: *i)* the average amount of time required for two alleles sampled *i* and *j* steps apart to be in the same subpopulation (S_{ij}); and *ii)* the average time to coalesce once the two alleles occupy the same subpopulation ($\bar{t}_{00}$) (Slatkin, 1991). Note that $\bar{t}_{ij}$ depends, separately, on the details of the dispersal process (S_{ij}) and on the number of individuals in the population ($\bar{t}_{00}$), such that $\bar{t}_{ij} = \bar{t}_{00} + S_{ij}$. Then:

$$F_{st(i,j)} = \frac{\left[\left(\bar{t}_{00} + S_{ij}\right) - \bar{t}_0\right]}{\bar{t}_{00} + S_{ij}} \qquad \text{(Eq. 3)}$$

Given that the model assumes a homogenous subpopulation size, $\bar{t}_{00} = \bar{t}_0$ and therefore $F_{st(i,j)} = S_{ij}/(\bar{t}_{00} + S_{ij})$. According to Slatkin, (2005), the term S_{ij} is defined as:

$$S_{ij} = \sum_{l=0}^{d_x/2} \sum_{k=0}^{d_y/2} \frac{f_{kl}\left[1 - \cos\left(2\pi ik/d_y\right)\cos(2\pi jl/d_x)\right]}{\delta(k)\delta(l)(1 - f_{kl})} \qquad \text{(Eq. 4)}$$

where $f_{kl} = \left[1 - m\left(1 - \cos\left(2\pi k/d_y\right)\right)\right]^2 [1 - m(1 - \cos(2\pi l/d_x))]^2$, $f_{kl} = 0$ if $k = 0$ and $l = 0$. The coefficients $\delta(k) = 1$ or $\delta(l) = 1$ if $k = 0$ or $l = 0$, respectively, and $\delta(k) = 1/2$ or $\delta(l) = 1/2$ if $k > 0$ or $l > 0$. As pointed out by Slatkin, (2005), this expression is a correction for typographical errors of *Eq. 8b* in Slatkin, (1993) and of *Eq. A9* in Slatkin, (1991).

In the context of habitat loss and fragmentation, that is when a continuous habitat is suddenly broken apart in several habitat fragments, among which gene-flow is completely impaired, the expression for $F_{st(i,j)}$ needs to be modified. To address this problem, we used the approach proposed in Duforet-Frebourg and Slatkin, (2016) for modelling the expected coalescence time ($\bar{t}_{ij}$) and $F_{st(i,j)}$ for two alleles sampled at different times, let's say in the present (modern DNA) and in the ancient past (ancient DNA). Our model assumes an initial two-dimensional population at dispersal-mutation-drift equilibrium which undergoes HL&F at time t_*, such that individuals occurring in each isolated habitat fragment can no longer disperse beyond the boundaries of the defined fragment. In order to derive an expression for $F_{st(i,j)}$ in terms of expected coalescence times, we need to describe the properties of the coalescence process in the context of HL&F. It is useful then to think backward in time, considering that HL&F has occurred t_* generations in the past and that alleles are sampled in the present, $t = 0$ (see Fig. S1 for a graphical visualization of the model).

In particular, we decompose the coalescent process for two alleles sampled at *i* and *j* steps apart $\bar{t}_{ij(t_*)}$, as a three phases process: *i*) the possible paths travelled by the ancestor of the two sampled alleles since t_* across the available subpopulations within the corresponding habitat fragment, $P\left(\boldsymbol{A}_1^{t_*}, \boldsymbol{B}_1^{t_*} | \boldsymbol{A}_1^{t_0}, \boldsymbol{B}_1^{t_0}\right)$; *ii*) the amount of time required for the two ancestral lineages to occupy the same deme in the continuous habitat at the time of HL&F (t_*), $S_{ij(t_*)}$ where $i = \left|x_{A_1}^{t_*} - x_{B_1}^{t_*}\right|$ and $j = \left|y_{A_1}^{t_*} - y_{B_1}^{t_*}\right|$; and *iii*) the average

time to coalesce once the two lineages occupy the same subpopulation ($\bar{t}_{00}$). Thus, the new expression for $\bar{t}_{ij}$ is:

$$\bar{t}_{ij(t_*)} = \bar{t}_{00} + \bar{S}_{ij(t_*)} + t_* = \bar{t}_{00} + \sum_{x_{A_2}=0}^{d_{x(A)}-1} \sum_{y_{A_2}=0}^{d_{y(A)}-1} \sum_{x_{B_2}=0}^{d_{x(B)}-1} \sum_{y_{B_2}=0}^{d_{x(B)}-1} S_{ij}\, P\left(\boldsymbol{A}_1^{t_*}, \boldsymbol{B}_1^{t_*} | \boldsymbol{A}_1^{t_0}, \boldsymbol{B}_1^{t_0}\right) + t_* \quad \text{(Eq. 5)}$$

where S_{ij} is calculated from *Eq. 4* and $P\left(\boldsymbol{A}_1^{t_*}, \boldsymbol{B}_1^{t_*} | \boldsymbol{A}_1^{t_0}, \boldsymbol{B}_1^{t_0}\right)$ is the probability that two alleles sampled at $\boldsymbol{A}_1^{t_0} = (x_{A_1}^{t_0},\ y_{A_1}^{t_0})$ for habitat fragment A and at $\boldsymbol{B}_1^{t_0} = (x_{B_1}^{t_0},\ y_{B_1}^{t_0})$ for habitat fragment B were at position $\boldsymbol{A}_1^{t_*} = (x_{A_1}^{t_*},\ y_{A_1}^{t_*})$ and $\boldsymbol{B}_1^{t_*} = (x_{B_1}^{t_*},\ y_{B_1}^{t_*})$ at the time of HL&F (t_*). Such that:

$$P\left(\boldsymbol{A}_1^{t_*}, \boldsymbol{B}_1^{t_*} | \boldsymbol{A}_1^{t_0}, \boldsymbol{B}_1^{t_0}\right) = p\left(x_{A_1}^{t_*}, y_{A_1}^{t_*} | x_{A_1}^{t_0}, y_{A_1}^{t_0}\right) p\left(x_{B_1}^{t_*}, y_{B_1}^{t_*} | x_{B_1}^{t_0}, y_{B_1}^{t_0}\right) \quad \text{(Eq. 6)}$$

Using the method detailed in Duforet-Frebourg and Slatkin, (2016), we show how to compute the probability that an allele sampled in a given position $\boldsymbol{A}_1^{t_0} = (x_{A_1}^{t_0},\ y_{A_1}^{t_0})$ at time $t = 0$ was at $\boldsymbol{A}_1^{t_*} = (x_{A_1}^{t_*},\ y_{A_1}^{t_*})$ at time t_* back in the past. We consider a discrete random walk in a torus-like space with transition matrix:

$$M = \begin{bmatrix} 1-m & m/4 & 0 & \cdots & 0 & m/4 \\ m/4 & 1-m & m/4 & \cdots & 0 & 0 \\ 0 & m/4 & 1-m & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ m/4 & 0 & 0 & \cdots & m/4 & 1-m \end{bmatrix} \quad \text{(Eq. 7)}$$

In *Eq. 7*, we indicate the possible state transitions in one unit of time (e.g., one generation). Then, using classical results on Markov chains, we calculate, for instance, $p\left(x_{A_1}^{t_*}, y_{A_1}^{t_*} | x_{A_1}^{t_0}, y_{A_1}^{t_0}\right)$ by taking the element $[(x_{A_1}, y_{A_1}), (x_{A_1}, y_{A_1})]$ from the time t_* convolution of the transition matrix (the power t_* of the matrix M). For instance, assuming a toroid with three demes in each of x and y direction, and considering $m = 0.04$ we obtain the following transition matrix:

$$M = \begin{array}{c|cccccccccccccccc}
 & 0,0 & 1,0 & 2,0 & 3,0 & 0,1 & 1,1 & 2,1 & 3,1 & 0,2 & 1,2 & 2,2 & 3,2 & 0,3 & 1,3 & 2,3 & 3,3 \\
\hline
0,0 & 0.96 & 0.01 & 0 & 0.01 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 0 & 0 & 0 \\
1,0 & 0.01 & 0.96 & 0.01 & 0 & 0 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 0 & 0 \\
2,0 & 0 & 0.01 & 0.96 & 0.01 & 0 & 0 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 0 \\
3,0 & 0.01 & 0 & 0.01 & 0.96 & 0 & 0 & 0 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 \\
0,1 & 0.01 & 0 & 0 & 0 & 0.96 & 0.01 & 0 & 0.01 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1,1 & 0 & 0.01 & 0 & 0 & 0.01 & 0.96 & 0.01 & 0 & 0 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 \\
2,1 & 0 & 0 & 0.01 & 0 & 0 & 0.01 & 0.96 & 0.01 & 0 & 0 & 0.01 & 0 & 0 & 0 & 0 & 0 \\
3,1 & 0 & 0 & 0 & 0.01 & 0.01 & 0 & 0.01 & 0.96 & 0 & 0 & 0 & 0.01 & 0 & 0 & 0 & 0 \\
0,2 & 0 & 0 & 0 & 0 & 0.01 & 0 & 0 & 0 & 0.96 & 0.01 & 0 & 0.01 & 0.01 & 0 & 0 & 0 \\
1,2 & 0 & 0 & 0 & 0 & 0 & 0.01 & 0 & 0 & 0.01 & 0.96 & 0.01 & 0 & 0 & 0.01 & 0 & 0 \\
2,2 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 0 & 0 & 0.01 & 0.96 & 0.01 & 0 & 0 & 0.01 & 0 \\
3,2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 0.01 & 0 & 0.01 & 0.96 & 0 & 0 & 0 & 0.01 \\
0,3 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 0 & 0 & 0 & 0.96 & 0.01 & 0 & 0.01 \\
1,3 & 0 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 0 & 0 & 0.01 & 0.96 & 0.01 & 0 \\
2,3 & 0 & 0 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 0 & 0 & 0.01 & 0.96 & 0.01 \\
3,3 & 0 & 0 & 0 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 0.01 & 0 & 0.01 & 0.96
\end{array}$$

In this transition matrix, the row's indexes on the left side indicate the coordinates of the demes, namely $x \in \{0, 1, 2, 3\}$ and $y \in \{0, 1, 2, 3\}$, which are the same as the column's indexes. As an example, note that in the first column we show the probabilities of transition of a lineage in $x_{A_1} = 0$ and $y_{A_1} = 0$ to the other demes of the 4 x 4 torus-like space. In particular, a lineage in $x_{A_1}^{t_0} = 0$ and $y_{A_1}^{t_0} = 0$ has a probability $1 - m$ (=0.96) of staying in $x_{A_1}^{t^*=1} = 0$ and $y_{A_1}^{t^*=1} = 0$, a probability $m/4$ (=0.01) of moving to $x_{A_1}^{t^*=1} = 1$ and $y_{A_1}^{t^*=1} = 0$ (on the right), a probability $m/4$ (=0.01) of moving to $x_{A_1}^{t^*=1} = 3$ and $y_{A_1}^{t^*=1} = 0$ (on the left), a probability $m/4$ (=0.01) of moving to $x_{A_1}^{t^*=1} = 0$ and $y_{A_1}^{t^*=1} = 1$ (upward) and a probability $m/4$ (=0.01) of moving to $x_{A_1}^{t^*=1} = 0$ and $y_{A_1}^{t^*=1} = 3$ (downward).

So far, the expressions used to derive $\bar{t}_{ij(t_*)}$ have referred to the consequences of *habitat fragmentation* (or disruption of gene-flow) on

average coalescence times. To include *habitat loss*, corresponding to the reduction in available habitat within each fragment, we need to model $\bar{t}_0$ with an expression at non-equilibrium ($\tilde{t}_0^{t_*}$). We used an approach similar to the one used in Maruyama, (1970) on the probability of identity by descent (Eq. 5-3) and to the expressions in Austerlitz et al., (1997) on the distribution of mean coalescence time in a population undergoing range expansion (Eq. 20). Consider a toroidal stepping-stone model composed by d^2 demes, where, at each generation, individuals have a probability $1 - m_{2D}$ to remain in the same deme and a probability $m_{2D}/4$ to disperse to each of the four adjacent demes ($^{m_{2D}}/_{4}$ x 4 demes = m_{2D}). Since a torus corresponds to the product of a one-dimensional circular stepping-stone, in the x-direction, with another one-dimensional circular stepping-stone, in the y-direction, it is useful to describe the probability distribution of the migration patterns of two sampled lineages in a one-dimensional circular stepping-stone. Such probability distribution is equivalent for both x- and y-direction, assuming that dispersal rate is equal in the two directions. Thus, define a dxd circular matrix $\boldsymbol{M}$ with elements $M_{i,j}$, for $i,j = 0, \ldots, d-1$, describing the probability that two lineages A_1 and A_2 sampled at distance i were at distance j in the previous generation. Defining $m = m_{2D}/2$, note that:

$$M_{0,0} = (1-m)(1-m) + \frac{m}{2}\frac{m}{2} + \frac{m}{2}\frac{m}{2} = (1-m)^2 + \frac{m^2}{2} \quad \text{(Eq. 8a)}$$

where $M_{0,0}$ is the probability that lineages A_1 and A_2 sampled in the same deme, $i = 0$, were also in the same deme, $j = 0$, in the previous generation, which is equivalent to the probability that neither lineage moved $(1-m)^2$, plus the probability that both moved clockwise $(m/2)(m/2)$ and plus the probability that both moved counterclockwise $(m/2)(m/2)$.

$$M_{0,1} = (1-m)\frac{m}{2} + \frac{m}{2}(1-m) = m(1-m) \qquad \text{(Eq. 8b)}$$

where $M_{0,1}$ is the probability that lineages A_1 and A_2 sampled in the same deme, $i = 0$, were at distance $j = 1$ in the previous generation. This is equivalent to the probability that one of the two lineages remained in the same deme in the previous generation $(1-m)$ times the probability that the other lineage moved clockwise $(m/2)$, plus the probability that one of the two lineages remained in the same deme $(1-m)$ times the probability that the other lineage moved counterclockwise $(m/2)$.

$$M_{0,2} = \frac{m}{2}\frac{m}{2} = \frac{m^2}{4} \qquad \text{(Eq. 8c)}$$

where $M_{0,2}$ is the probability that lineages A_1 and A_2 sampled in the same deme, $i = 0$, were at distance $j = 2$ in the previous generation. This can be expressed as the probability that one of the two lineages moved clockwise $(m/2)$ in the previous generation, while the other lineage moved counterclockwise $(m/2)$. In the case of $d = 5$, matrix $\boldsymbol{M}$ is:

$$\begin{pmatrix} (1-m)^2+\frac{m^2}{2} & m(1-m) & \frac{m^2}{4} & 0 & \frac{m^2}{4} & m(1-m) \\ m(1-m) & (1-m)^2+\frac{m^2}{2} & m(1-m) & \frac{m^2}{4} & 0 & \frac{m^2}{4} \\ \frac{m^2}{4} & m(1-m) & (1-m)^2+\frac{m^2}{2} & m(1-m) & \frac{m^2}{4} & 0 \\ 0 & \frac{m^2}{4} & m(1-m) & (1-m)^2+\frac{m^2}{2} & m(1-m) & \frac{m^2}{4} \\ \frac{m^2}{4} & 0 & \frac{m^2}{4} & m(1-m) & (1-m)^2+\frac{m^2}{2} & m(1-m) \\ m(1-m) & \frac{m^2}{4} & 0 & \frac{m^2}{4} & m(1-m) & (1-m)^2+\frac{m^2}{2} \end{pmatrix}$$

Given the matrix $\boldsymbol{M}$, it is possible to compute the mean coalescence time at the next generation $\boldsymbol{T'}$ as:

$$\boldsymbol{T'} = \boldsymbol{M}(\boldsymbol{T} - \boldsymbol{T_0})\boldsymbol{M} + \boldsymbol{U} \qquad \text{(Eq. 9)}$$

where $\boldsymbol{T}$ is a dxd matrix of mean coalescence times of lineages initially separated by a distance i and j on the x- and y-direction, respectively; $\boldsymbol{T_0}$ is a dxd matrix with $(\boldsymbol{T_0})_{0,0} = \bar{t}_{0,0}/2N_e$ and 0 elsewhere; and $\boldsymbol{U}$ is a dxd matrix whose elements are 1s. The expression $\boldsymbol{T} - \boldsymbol{T_0}$ takes into account that if two lineages were in the same deme in the previous generation, then the contribution of $\bar{t}_{0,0}$ needs to be weighted by the probability of not coalescing $(1 - 1/2N)$. In fact, at $i = 0$ and $j = 0$, $(\boldsymbol{T})_{0,0} - (\boldsymbol{T_0})_{0,0}$ takes the form $(1 - 1/2N)\bar{t}_{0,0}$. Moreover, note that in *Eq. 9* the matrix $\boldsymbol{M}$ is multiplied at the left and at the right of $\boldsymbol{T} - \boldsymbol{T_0}$, since it is necessary to consider the possible migration patterns in both x- and y-direction. Ultimately, the matrix $\boldsymbol{U}$ is used to add one time unit at each generation, given that lineages did not coalesce. At any generation after HL&F, the mean coalescence time for pairs of lineages at distance i and j can then be calculated by iterating *Eq. 9*.

There are now all the necessary elements to re-rewrite *Eq. 1, 2* for a model of population divergence in space and time, such that:

$$F_{st(i,j,t_*)} = \frac{\left[\left(\bar{t}_{00} + \bar{S}_{ij(t_*)} + t_*\right) - \tilde{t}_0^{\,t_*}\right]}{\bar{t}_{00} + \bar{S}_{ij(t_*)} + t_*} \qquad \text{(Eq. 10)}$$

For a better interpretation of our results, we define $\Delta\tilde{t}_0^{\,t_*} \equiv \bar{t}_{00} + t_* - \tilde{t}_0^{\,t_*}$ as the *drift-isolation term*, which determine the rate of 'independent' evolution of each habitat fragment due to *habitat loss* since HL&F at time t_*. In particular, $\bar{t}_{00} - \tilde{t}_0^{\,t_*}$ determines the rate of loss in genetic diversity within subpopulations, while t_* considers the mutations that accumulate in each lineage in the two habitat fragments. To put it in another way, t_* corresponds also to the time is necessary to wait such that two alleles have the possibility to coalesce. This is because two lineages cannot coalesce until they occupy the same subpopulation, condition that requires a continuous habitat where

gene-flow is possible between the two sampled alleles at *i* and *j* steps apart. Thus, we can re-arrange $F_{st(i,j,t_*)}$ as:

$$F_{st(i,j,t_*)} = \frac{\bar{S}_{ij(t_*)} + \Delta\tilde{t}_0^{\,t_*}}{\bar{t}_{00} + \tilde{S}_{ij(t_*)} + t_*} \qquad \text{(Eq. 11)}$$

This last equation will turn out to be useful when we discuss the effect of HL&F on IBD pattern over time in the *Discussion* section.

3.4 Results

3.4.1 Instantaneous HL&F

In the present study, we investigated the effect of landscape-scale habitat loss and fragmentation on 'isolation-by-distance' patterns (IBD). Overall, our results suggest that significant IBD signatures can be detected for surprisingly long periods (on the order of thousands of generations) after HL&F. We showed that dispersal rate (0.02 – 0.1) and sampling strategy (*local* vs *random*) influence the time of loss of IBD (T_{IBD}) under the *instantaneous HL&F* scenario (Fig. 1b). On average, 'isolation-by-distance' pattern persists for longer time at lower dispersal rates (m = 0.02, $T_{IBD-local}$ ≈ 700; m = 0.1, $T_{IBD-local}$ ≈ 50; Fig. 1b, S2a). The influence of dispersal rate on T_{IBD} becomes more noticeable when alleles are randomly sampled within a habitat fragment (*random*) compared to a within-subpopulation (*local*) sampling scheme (m = 0.02, $T_{IBD-random}$ ≈ 1500; m = 0.1, $T_{IBD-random}$ ≈ 200). Furthermore, this suggests that sampling strategy (*local* vs *random*) can significantly influence T_{IBD} (Fig. 1b, S2a). Larger absolute differences between $T_{IBD-local}$ and $T_{IBD-random}$ were identifiable at lower dispersal rate. For instance, at m = 0.02, we estimated an average $T_{IBD-local}$ of 700 and a $T_{IBD-random}$ of 1500 generations (m = 0.02; U = 1974, p-val<0.001), whereas at m = 0.1, $T_{IBD-local}$ was detected at 50 and $T_{IBD-random}$ at 200 generations (m = 0.1; U = 2969.5, p-val<0.001). These results were consistent when IBD patterns were estimated using Wright's F_{st} instead of Slatkin's R_{st} (Fig. S3).

Carrying capacity and mutation rate appear to influence T_{IBD} much less than dispersal rate, for both *local* and *random* sampling schemes (Fig. 1c, d). The distribution of T_{IBD} for each independent simulation (Fig. 1c, S2b) and average IBD p-value (Fig. 1c) are nearly indistinguishable among the three values of carrying capacity herein tested (50 – 200). Small differences were observed in R_{st}-based IBD pattern between $K = 200$ and $K = 50/100$, but they were nearly absent when IBD was estimated from F_{st} (Fig. S4).

Accordingly, average R_{st} (or F_{st}) behaved similarly in relation to geographical distance, regardless of the simulated carrying capacity (Fig. S2e and Fig. S4d). With some exceptions, we did not observe differences in T_{IBD} in relation to mutation rate (Fig. 1d). In fact, for mutation rates between 10^{-2} and 10^{-5} per locus per generation, our results show no difference in the time at which IBD tends to be lost. Different is the case for mutation rate equal to 5 x 10^{-6}, since IBD tends to be lost earlier than the other tested values. The independence of T_{IBD} from mutation rate is further supported by the lack of differences in the temporal dynamic of IBD slope among simulations with different values of mutation rate (Fig. S2c, f).

The number of genotyped markers has a major role in determining the time of IBD loss (Fig. S6). As an example, when we simulated an *instantaneous HL&F* scenario ($K = 50$; $m = 0.04$; $\mu = 5 \times 10^{-4}$) and sampled either 15, 30, 100 or 400 microsatellite loci, we found that the estimated time before the loss of a significant IBD signature (T_{IBD}) correlated with the number of genotyped markers. We did not observe major differences in T_{IBD} for 15 and 30 loci ($U_{15\text{-}30}$ = 193.5, p-value = 0.003), but strong discordance between T_{IBD} for 15/30 loci and 100/400 loci ($U_{15\text{-}100}$ = 597.5, p-value<0.001; $U_{30\text{-}400}$ = 105.5, p-value<0.001). Similar results were obtained for IBD patterns measured using Wright's F_{st}; however, in this case, differences of T_{IBD} were less strong ($U_{15\text{-}100}$ = 797, p-value<0.001), but still important (Fig. S7). A qualitative comparison with the previous results suggests that the number of genotyped loci can have a greater impact on T_{IBD} than dispersal rate (e.g., Fig. S3 *vs* Fig. S7).

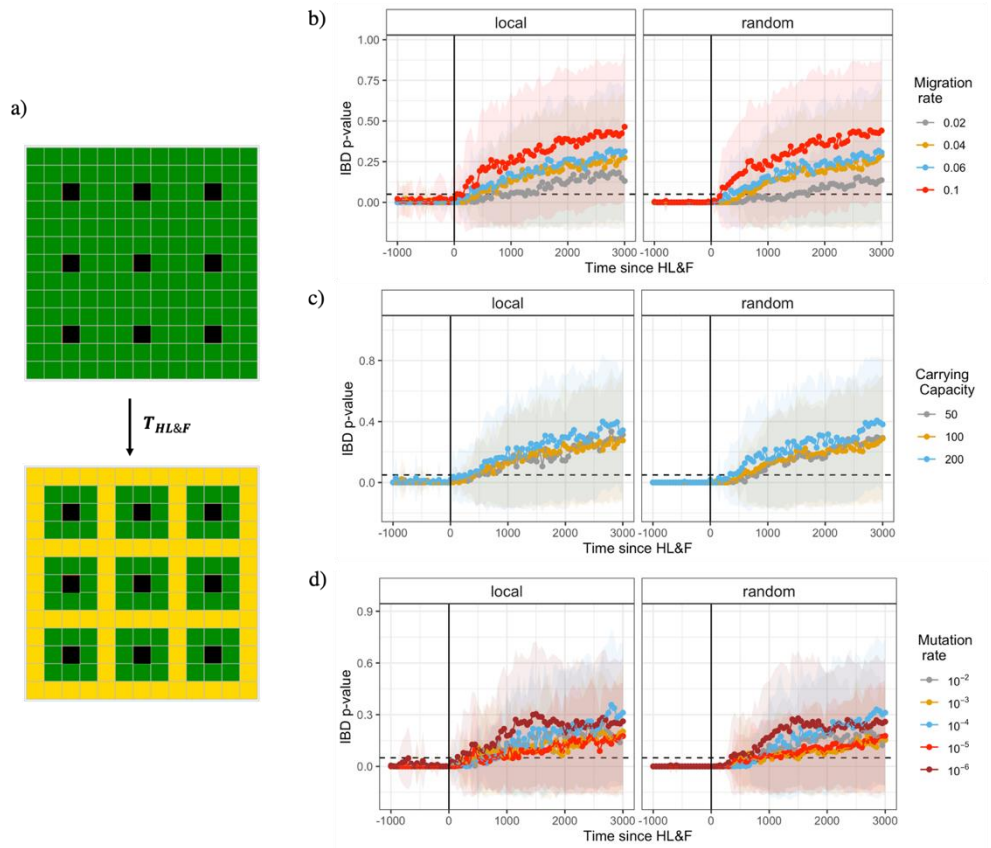


Figure 1. The effect of habitat loss and fragmentation on R_{st}-based isolation-by-distance (IBD) in relation to dispersal rate, population size and mutation rate. a) Simulated 13 x 13 square grid of demes ($d_{tot} = 169$) undergoing HL&F at time $T_{HL\&F}$. Black cells refer to the local sampling scheme, green and yellow cells refer to suitable and unsuitable habitat, respectively. Time at which IBD is no longer significant (T_{IBD}) as a function of **b)** dispersal rate ($m = 0.02–0.01$, $K = 100$, $\mu = 5 \times 10^{-4}$), **c)** population size ($K = 50 - 200$, $m = 0.04$, $\mu = 5 \times 10^{-4}$) and mutation rate ($\mu = 10^{-2}$-10^{-6}, $K = 100$, $m = 0.02$). Mean (dots) and variance (shaded area) of IBD p-value across simulation replicates. Dashed line represents p-value = 0.05. Local sampling: patch-sampled individuals come from the same deme; Random sampling: patch-sampled individuals are randomly collected across demes within the patch. The results suggest i) that IBD is lost quicker at high than low dispersal rate, ii) that population size has little effect on T_{IBD}, and iii) that mutation rate does not influence T_{IBD}. Note, however, a large variance in T_{IBD} across all simulated parameter combinations.

3.4.2 Range expansion prior to instantaneous HL&F

We also considered a more complex scenario where *instantaneous HL&F* occurs after a range expansion, i.e., in a non-equilibrium population (Fig. 2a). Thus, two evolutionary processes are expected to contribute to the observed spatial genetic pattern: *range expansion* and *HL&F*. Several *range expansion + instantaneous HL&F* scenarios were simulated, where HL&F occurred at different time points since the onset of the range expansion ($T_{exp} = 600, 750, 900, 1050 \text{ and } 1500$). We observed a significant increase in T_{IBD} for the *range expansion + instantaneous HL&F* compared to the *instantaneous HL&F* scenario (Fig. 2a, Fig. S8a, c). Such differences were observable with both *local* and *random* sampling, but more noticeable with the *random* sampling (*equilibrium* vs $T_{exp} = 600$: U_{local} = 1740.5, p-value = 0.01; U_{random} = 1136, p-value < 0.001). Differences between $T_{IBD-equilibrium}$ and $T_{IBD-expansion}$ were larger for recent than for ancient range expansions (*equilibrium* vs $T_{exp} = 900$: U_{random} = 1200.5, p-value = 0.027), converging to the behavior of the *instantaneous HL&F* for more ancient range expansions (*equilibrium* vs $T_{exp} = 1500$: U_{random} = 1244.5, p-value = 0.8774). Similar results were obtained for IBD-patterns estimated using Wright 's F_{st} (Fig. S9). The temporal dynamics of pairwise R_{st} indicates that differences between $T_{IBD-equilibrium}$ and $T_{IBD-expansion}$ are related to the larger effect of geographical distance on pairwise R_{st} in the *range expansion + instantaneous HL&F* compared to the *instantaneous HL&F* scenario. For instance, using relative weights analysis, we estimated a relative rescaled weight for geography on pairwise R_{st} ($RRW_{geography}$) of ≈ 1% for *equilibrium* and ≈ 7% for T_{exp}= 600 scenarios (Fig. S11a, b). The results show that *range expansion* has also the effect of increasing the variance around the mean R_{st} (shaded area in Fig. S11a). As shown by the behavior of $T_{IBD-expansion}$, discrepancies in pairwise R_{st} (or F_{st}) between the *range expansion + instantaneous HL&F* and the *instantaneous HL&F* scenario decrease with the age of the range

expansion ($RRW_{geography} \approx$ 1% for *equilibrium*; $RRW_{geography} \approx$ 1.5% for T_{exp}= 1500; Fig. S11a).

3.4.3 Two-steps HL&F

We assessed the behavior of the time of loss of a significant IBD signature (T_{IBD}) in a scenario where two consecutive HL&F events generated the same HL&F landscape configuration as in the *instantaneous HL&F* (Fig. 2b). We tested the effect of the time since the first HL&F event (T_{split}) on T_{IBD} and compared it to the T_{IBD} in an equivalent *instantaneous HL&F* scenario ($T_{equilibrium}$). We observed that T_{IBD} increased with T_{split}. For instance, we observed a difference in T_{IBD} of ~ 400 generations between $T_{equilibrium}$ and $T_{split} = 500$ scenarios and a difference of ~ 2000 generations between $T_{equilibrium}$ and $T_{split} = 1500$. Similar results are reported also for Wright 's F_{st} – IBD pattern (Fig. S10). As for the other investigated scenarios, the effect of *Two-steps HL&F* is even more marked for *random* than *local* sampling. Pairwise R_{st} averaged across simulations or pairwise R_{st} for single simulation replicates (Fig. S11c, d) shows that the longer persistence of IBD in the *Two-steps HL&F* scenario, compared to the *instantaneous HL&F* scenario (*equilibrium*), is related to the larger effect of geographic distance on R_{st} ($RRW_{geography} \approx$ 1% for *equilibrium*; $RRW_{geography} \approx$ 8% for T_{split}= 1500).

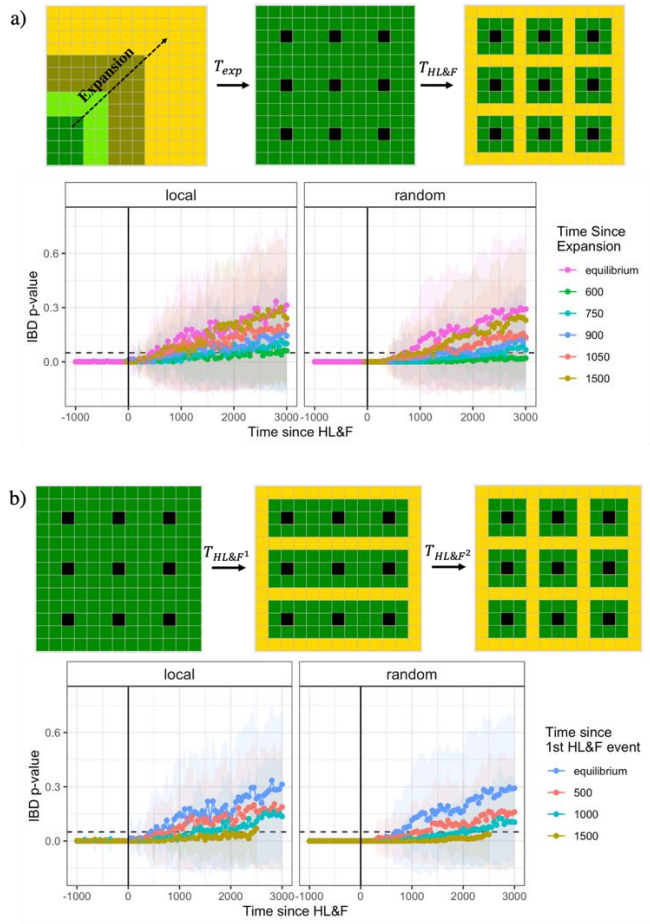


Figure 2. The effect of ancient range expansion followed by HL&F or spatio-temporal variation in HL&F on loss of R_{st}-based isolation-by-distance. a) Simulated 13 x 13 square grid of demes, where T_{exp} is the time since the onset of the range expansion and $T_{HL\&F}$ is the time of instantaneous HL&F. Black cells refer to the local sampling scheme, green and yellow cells refer to suitable and unsuitable habitat, respectively. Range expansion was simulated for a scenario with $K = 50$, $m = 0.04$ and $\mu = 5 \times 10^{-4}$. The results show that the genetic drift originated during a recent range expansion can contribute to a longer persistence of IBD following a HL&F event. **b)** Simulated 13 x 13 square grid of demes, where $T_{HL\&F^1}$ and $T_{HL\&F^2}$ indicate, respectively, the time of the 1st and 2nd HL&F event. Piecewise HL&F was simulated for a scenario with $K = 50$, $m = 0.04$ and $\mu = 5 \times 10^{-4}$. The results show that spatio-temporal variation in HL&F can have a strong effect on T_{IBD}, which is exacerbated with increasing $\Delta T_{HL\&F}$. Mean (dots) and variance (shaded area) of IBD p-value across simulation replicates. Dashed line represents p-value = 0.05. Local sampling: patch-sampled individuals come from the same deme; Random sampling: patch-sampled individuals are randomly collected across demes within the patch.

3.4.4 Population divergence in space and time

F_{st} between pairs of populations can be expressed in terms of expected coalescence time and can be calculated for various types of structured model, such as a 2D-stepping stone model (Slatkin, 1991). Using the approach developed in Duforet-Frebourg and Slatkin, (2016), we proposed an analytical model of isolation-by-distance in time in the context of habitat loss and fragmentation. Following Slatkin, (1991), we assumed a two-dimensional grid of demes distributed on a torus. The advantage of using a torus-like space is that many of the analytical results in a continuous space are known and easy to compute.

The analytical model considers an initial continuous two-dimensional habitat (e.g., forest) of size d_x x d_y, at mutation-dispersal-drift equilibrium. At a given time $T_{HL\&F}$ (forward-in time), the continuous habitat breaks apart, forming two habitat fragments (A and B) of size, respectively, d_{A_x} x d_{A_y}, and d_{B_x} x d_{B_y}, among which there is no gene-flow. The question being asked is 'how does genetic differentiation accumulate over time between two spatially-distributed subpopulations in A and B after HL&F?' Does geography contribute to the level of genetic differentiation measured between A and B after HL&F for long time? Thinking backward in time, we assume that two alleles are sampled in two isolated habitat fragments (A and B) at the present time ($t = 0$), and that an event of HL&F between the two fragments has occurred t_* generations ago (see Fig. S1 for a graphical representation of the model). As outlined in *Material and Methods* section (*Eq. 11*), the $F_{st(i,j,\,t_*)}$ over time for two isolated habitat fragments can be expressed as: $F_{st(i,j)} = \left[\bar{S}_{ij(t_*)} + \Delta\tilde{t}_0^{\,t_*}\right] / \left[\bar{t}_{00} + \bar{S}_{ij(t_*)} + t_*\right]$, where $\Delta\tilde{t}_0^{\,t_*} \equiv \bar{t}_{00} + t_* - \tilde{t}_0^{\,t_*}$. For the sake of clarity, we will refer to $\Delta\tilde{t}_0^{\,t_*}$ as the *drift-isolation term* and to $\bar{S}_{ij(t_*)}$ as the *spatial term*. In Fig. 3 and 4, we represent the behavior of $\bar{S}_{ij(t_*)}$ and $\tilde{t}_0^{\,t_*}$ in relation to dispersal rate. In particular, Fig. 3a shows the impact of dispersal rate on the relative distribution of S_{ij} values within habitat fragment ($S_{ij} - \bar{S}_{ij}$),

highlighting larger differences in S_{ij} among subpopulations when dispersal rate is low (scaled $\sigma_{m=0.02} = 8{,}715$ *vs* $\sigma_{m=0.1} = 1{,}742$). In Fig. 3b and Fig. 3c, we show the transition probabilities $P(\boldsymbol{A}_1^{t_*}|\boldsymbol{A}_1^{t_0})$ for a given gene sampled at time $t = 0$ in the center of a 9 x 9 habitat fragment ($x_A = 4$, $y_A = 4$) to be at any other position within the fragment, respectively, at $t_* = 50$ and $t_* = 1000$. As expected, higher is dispersal rate, higher is the probability of having sampled an allele in (4, 4) at time $t = 0$ that was originally further apart at the time of HL&F (Fig. 3b). Similarly, older is the HL&F event, higher is the probability of having sampled a gene in (4, 4) at time $t = 0$ that was originally further apart at the time of HL&F ($t_* = 50$, Fig. 3b *versus* $t_* = 1000$, Fig. 3c). Fig. 4a show the dependence of $\bar{t}_{ij(t_*)} = \bar{t}_{00} + \bar{S}_{ij(t_*)} + t_*$ on geographical distance, time since HL&F (t_*) and dispersal rate. We found that lower is the dispersal rate, lower is the relative impact of time since HL&F compared to geography on pairwise $F_{st(i,j,t_*)}$ ($m = 0.02$, $RRW_{geography} \approx 99\%$, $RRW_{t_*} \approx 1\%$; $m = 0.1$, $RRW_{geography} \approx 80\%$, $RRW_{t_*} \approx 20\%$).

In Fig. 4b we show the decay in $\tilde{t}_0^{t_*}$ after HL&F as a function of habitat fragment size and dispersal rate. As expected, smaller is the habitat fragment size after HL&F, steeper is the decay in expected coalescence time. Moreover, lower is dispersal rate, slower is the decay in expected coalescence time. Using the expressions describing the decay in $\tilde{t}_0^{t_*}$ and the increase in $\bar{t}_{ij(t_*)}$, we can numerically compute $F_{st(i,j,t_*)}$. The results are displayed in Fig. 5 and 6, which shows how $F_{st(i,j,t_*)}$ changes over time, as function of geographical distance, dispersal rate, deme population size and habitat loss. For the sake of clarity, we decided to represent only two extreme cases among those here considered, namely a HL&F event that break a 78 x 78 torus-like continuous space in fragments of either 24 x 24 or 4x4. Regarding dispersal rate, relative weights analyses showed that the relative contribution of geography on $F_{st(i,j,t_*)}$ decreases with increasing dispersal rate. For instance, in the case of habitat fragment size 24 x 24, we quantified

a $RRW_{geography} \approx 81\%$ *versus* $RRW_{t_*} \approx 19\%$ for the lower dispersal rate ($m = 0.02$) and a $RRW_{geography} \approx 4\%$ *versus* $RRW_{t_*} \approx 96\%$ for the higher dispersal rate ($m = 0.1$). Concerning deme population size, the relative contribution of geography on $F_{st(i,j,t_*)}$ remains nearly constant with increasing deme population size. For habitat fragment size 24 x 24, we quantified a $RRW_{geography} \approx 81\%$ *versus* $RRW_{t_*} \approx 19\%$ for the lower deme population size ($N_e = 50$) and a $RRW_{geography} \approx 80\%$ *versus* $RRW_{t_*} \approx 20\%$ for the higher deme population size ($N_e = 200$). In Fig. 5 and 6, we also show that the relative contribution of geography on $F_{st(i,j,t_*)}$ strongly depends on habitat loss. Accordingly, relative weights analysis quantified, for a smaller habitat fragment size (4 x 4), a $RRW_{geography} \approx 0.1\%$ *versus* $RRW_{t_*} \approx 99.9\%$ at low dispersal rate ($m = 0.02$) and a $RRW_{geography} \approx 0.006\%$ *versus* $RRW_{t_*} \approx 99.994\%$ at high dispersal rate ($m = 0.1$). A wider exploration of the parameter space is shown in Fig. 7, where $RRW_{geography}$ is displayed for any combination of deme population size ($N_e = 25 - 300$), dispersal rate ($m = 0.002 - 0.1$) and habitat loss (24x24, 20x20, 16x16, 12x12, 8x8, 4x4). The results show that dispersal rate and habitat loss are the factors that consistently determine the contribution of geography on pairwise F_{st}.

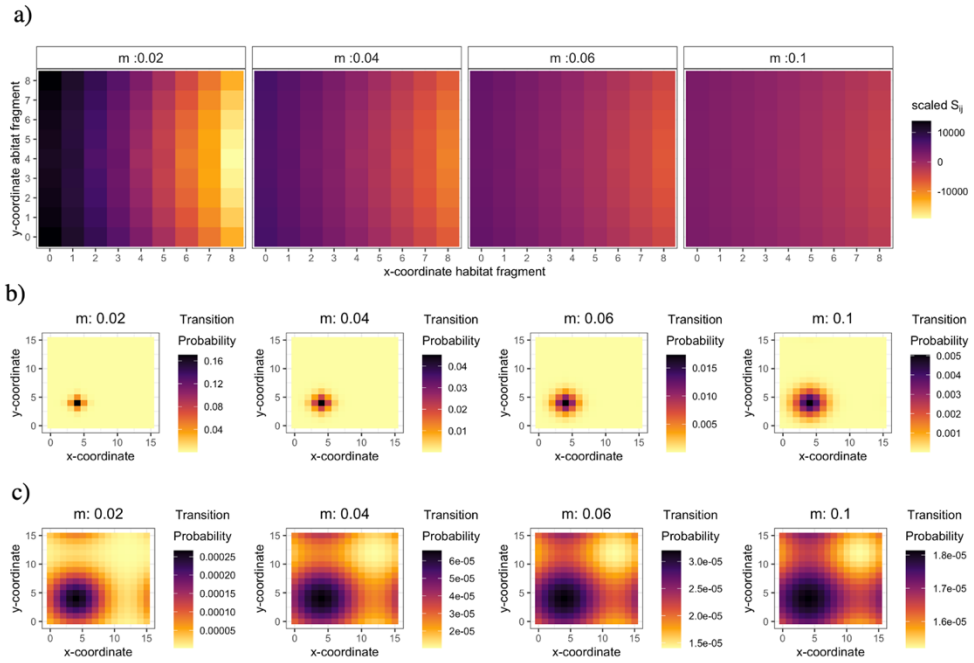


Figure 3. Graphical representation of the behaviour of the spatial term $\overline{S}_{ij(t_*)}$ as function of geographical position and time since HL&F (t_*). a) Relative distribution of S_{ij} for all x_A and y_A within fragment A (d_{A_x}= 9, d_{A_y} = 9) by fixing the sampled location of a random allele in fragment B (d_{B_x}= 9, d_{B_y} = 9) at x_B= 4 and y_B= 4. Absolute S_{ij} values for each dispersal rate were scaled with respect to its mean S_{ij} across the fragment, e.g., $S_{ij(m:0.02)} - mean(S_{ij(m:0.02)})$. Transition probabilities $P(\boldsymbol{A}_1^{t_*} \mid \boldsymbol{A}_1^{t_0})$ for a given allele sampled at time $t = 0$ in the center of fragment A (x_A= 4 and y_A= 4) to be at any other position within the fragment A, respectively, at **b)** $t_* = 50$ and **c)** $t_* = 1000$, as function of dispersal rate.

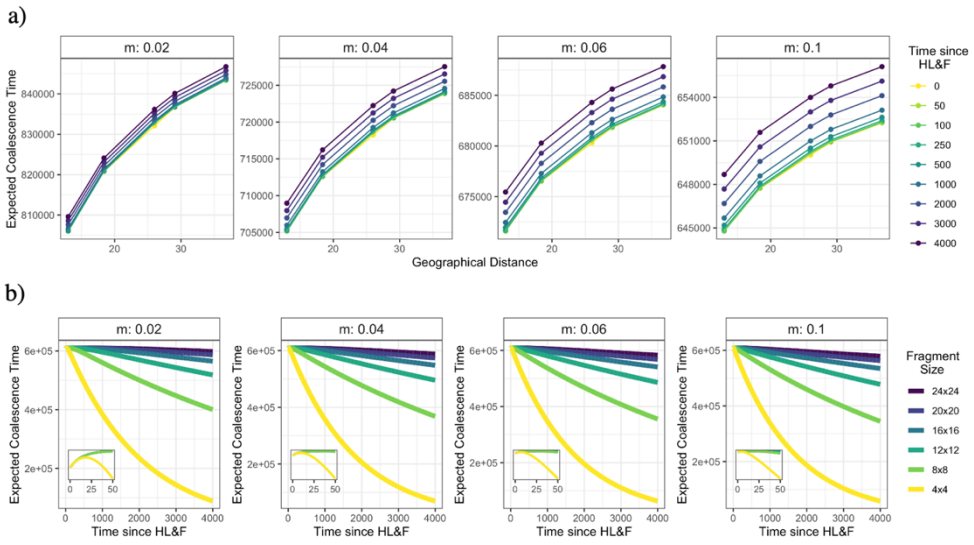


Figure 4. Non-equilibrium mean coalescence time between and within habitat fragments. **a)** Change in mean coalescence time between two alleles sampled at the centre of two isolated habitat fragments as function of dispersal rate and geographical distance ($d_x = 78$, $d_y = 78$, $N_e = 50$) (*Eq. 5*). **b)** Decay in mean coalescence time for two alleles randomly sampled within the same deme in a 78 x 78 torus-like space undergoing habitat contraction. The results show i) a stronger effect of time since HL&F on between-habitat fragments mean coalescence time at high than low dispersal rate, and ii) a slower decay in within-deme mean coalescence time at low than high dispersal rate

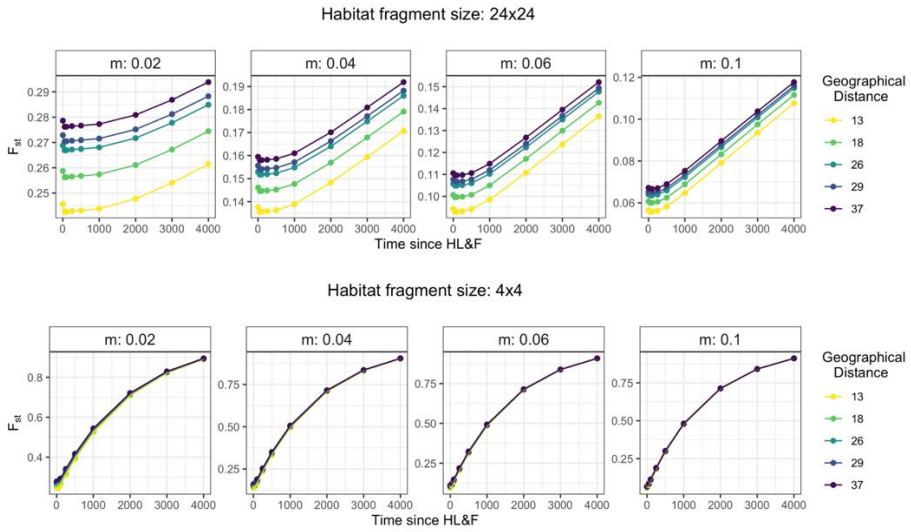


Figure 5. The effect of dispersal rate on expected pairwise F_{st} in a torus-like space undergoing HL&F. First and second rows refer to results for habitat fragments of size 24x24 and 4x4, respectively. Expected pairwise F_{st} were computed from *Eq. 10*. The analytical model suggests that time since HL&F has a stronger influence on pairwise F_{st} at high dispersal rate and/or when habitat undergoes severe habitat loss (4 x 4).

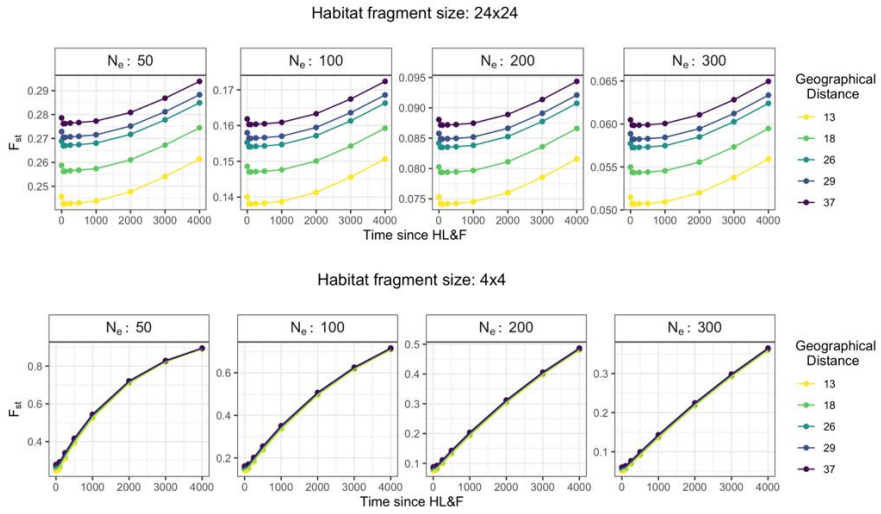


Figure 6. The effect of deme population size on expected pairwise F_{st} in a torus-like space undergoing HL&F. First and second rows refer to results for habitat fragments of size 24 x 24 and 4 x 4, respectively. Expected pairwise F_{st} were computed from Eq. 10. The analytical model suggests that the relative contribution of geography on pairwise F_{st} is little affected by deme population size.

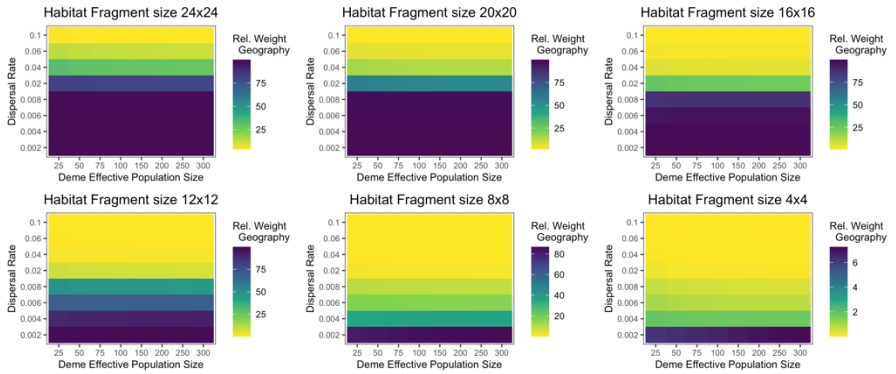


Figure 7. The effect of dispersal rate, deme population size and habitat loss on the relative contribution of geography on expected pairwise F_{st} in a torus-like space undergoing HL&F. Each panel refers to results obtained in a torus-like space of size 78x78 undergoing HL&F, leading to the formation of nine habitat fragments of size 24x24, 20x20, 16x16, 12x12, 8x8 or 4x4. Expected pairwise F_{st} was numerically computed from *Eq. 10* for different time points after HL&F: 50, 100, 250, 500, 1000, 2000, 3000, 4000 generations. Rel. Weight Geography: relative contribution of geography computed using the relative weight analysis (see *Material and Methods*). Overall, the results suggest that dispersal rate and habitat loss are the factors that determine the most the relative contribution of geography on pairwise F_{st}.

3.5 Discussion

3.5.1 Overview of the study

We used individual-based spatially-explicit simulations to explore non-equilibrium isolation-by-distance patterns (IBD) during habitat loss and fragmentation (HL&F). Classical IBD models relies on the important assumption that subpopulations are at mutation–dispersal–drift equilibrium and connected by ongoing gene-flow. In the present study, we were interested in understanding why genetic data from species occurring in currently fragmented landscapes (without ongoing gene-flow and, supposedly, not at equilibrium) can reveal significant IBD patterns (e.g., Schneider et al., 2010; Aleixo-Pais et

al., 2019), although violating one of the main assumptions of the IBD theoretical model (i.e., continuous habitat). We explored three scenarios of HL&F, aiming at characterizing alternative models that are likely to occur in natural populations. In particular, we investigated two scenarios in which populations were at mutation–dispersal–drift equilibrium prior to HL&F (*instantaneous HL&F* and *Two-steps HL&F*) and one scenario were HL&F occur in a non-equilibrium population (*range expansion prior to HL&F*).

We start considering the case of an *instantaneous HL&F*, assessing the influence of local population size, dispersal and mutation rates on the time of IBD loss (T_{IBD}). Since, natural populations may not necessarily be at mutation–dispersal–drift equilibrium, we also investigated a scenario in which an *instantaneous HL&F* occur after a *range expansion*. The last explored scenario focuses on the effect of *Two-steps HL&F* on IBD patterns. In fact, it is likely more realistic to consider that landscape-level HL&F occurs gradually over time following fine-scale spatial-temporal processes (Mandelbrot, 1982; Krummel et al., 1987; Milne, 1988; Solé and Manrubia, 1995).

The aim of the present study is to provide some insights on why significant IBD patterns can still be observed in fragmented natural populations. The demographic history of species is certainly complex and the simple scenarios explored in this work do not represent a comprehensive understanding of non-equilibrium IBD pattern in the context of HL&F. However, we attempted to consider scenarios that are likely common in real populations and thus, may contribute to a better interpretation of non-equilibrium IBD pattern from wild populations. This is the case, for instance, for *range expansion + instantaneous HL&F* scenario. In fact, observed present-day genetic data suggest that range expansion has shaped the genetic diversity of many taxa across the world. A well-known example is given by

species from Eurasia that have undergone expansion following the retreat of ice sheets at the end of the Last Glacial Period (LGM), about 20 thousand years ago (Hewitt, 2000, 1999; Lattin, 1967; Taberlet et al., 1998; Juřičková et al., 2014).

We conclude the present work proposing a theoretical model of *isolation-by-distance in time* in the context of HL&F, which could be interpreted also as a model of *population divergence in space and time*. In *Material and Methods*, we presented the details of the model by first summarizing the original theoretical results on a toroidal 2D stepping-stone model and then extending the theory in the context of habitat loss and fragmentation. Our analytical model shows that non-equilibrium F_{st} can be expressed as $F_{st(i,j,t_*)} = \left[\bar{S}_{ij(t_*)} + \bar{t}_{00} + t_* - \tilde{t}_0^{t_*}\right] / \left[\bar{t}_{00} + \bar{S}_{ij(t_*)} + t_*\right]$. We have shown that the dependence of $F_{st(i,j,t_*)}$ on geographical distance relies on the relative contribution of the *spatial term* ($\bar{S}_{ij(t_*)}$) compared to the *drift-isolation term* ($\Delta\tilde{t}_0^{t_*} \equiv \bar{t}_{00} + t_* - \tilde{t}_0^{t_*}$). In the *spatial term*, the effect of geographical distance relies on how dispersal rate influences the *i*) expected coalescence time between two genes at *i* and *j* steps apart, in the continuous torus-like space, before HL&F; and *ii*) the probability that genes sampled at time $t = 0$, respectively, in fragment A and B were in other neighboring subpopulations within each fragment at the time of HL&F (*Eq. 5*). In the *drift-isolation term*, geographical distance (related to habitat fragment size) and dispersal rate influence the decay in $\tilde{t}_0^{t_*}$, which occurs at a slower rate i) for low compared to high dispersal rate (Fig. 4b) and ii) for large compared to small fragment size (Fig. 5 and 6). To put it briefly, the numerical results have shown that dispersal rate and habitat fragment size are the two most important factors influencing the relative contribution of geography on pairwise F_{st} (Fig. 7). This short overview of the analytical model is

going to be useful for discussing the simulation results presented below.

3.5.2 Instantaneous HL&F

Overall, our simulation results have shown that IBD generated before HL&F can be maintained for thousands of generations after HL&F. We started investigating the contribution of deme population size, dispersal or mutation rate on the persistence of IBD during HL&F. We found that the time of IBD loss (T_{IBD}) is mainly dependent on dispersal rate, and little affected by population size and mutation rate. In particular, we showed that at low dispersal rate IBD is maintained for a larger number of generations after HL&F compared to an equivalent scenario at higher dispersal rate (Fig. 1a). Conversely, when testing different values of either population size or mutation rate, T_{IBD} remained nearly unchanged across the different parameters combinations (Fig. 1b, c).

We interpret these results using coalescence theory and the relationship between average coalescence times and pairwise F_{st} (*Eq. 11*). In our simulation results (Fig. S6-S7), it is shown that, following HL&F, IBD patterns are lost due to the fact that pairwise $F_{st(i,j)}$ are no longer distinguishable in relation to geographical distance. The analytical model suggests that this condition is satisfied when the relative contribution of $\bar{S}_{ij(t_*)}$ is much smaller than $\Delta\tilde{t}_0^{t_*}$, which can be due to either *i*) a very ancient HL&F (t_*), or *ii*) to a large reduction in total population size within habitat fragment ($\tilde{t}_0^{t_*}$); *iii*) or to both phenomena. The decrease in T_{IBD} with increasing dispersal rate observed in the simulation results can be understood from the relationship between S_{ij} and dispersal rate (Fig. 3a). In fact, higher is dispersal rate, lower is S_{ij}, and thus a smaller amount of time (since

HL&F) is required for the *spatial term* to become negligible compared to $\Delta\tilde{t}_0^{t_*}$ (Fig. 4a and 7).

The analytical model also explains why T_{IBD} is mainly influenced by dispersal rate, and little by mutation rate and population size. In fact, as shown by Slatkin, (1991), one of the advantages of expressing F_{st} in terms of coalescence times is that it does not depend on mutation rate (μ). This is analytically derived under the assumption that $\mu \ll 1$. Accordingly, our simulations showed that mutation rate do not influence T_{IBD}. An exception was found for $\mu = 5\,x\,10^{-6}$, which can be explained by the fact that the simulated population did not have enough time before HL&F (10^4 generations) to reach mutation-dispersal-drift equilibrium. In fact, it has been shown that in a 2D stepping-stone model, pairwise F_{st} approaches equilibrium over a timescale on the order of μ^{-1}, which in this case would be 10^6 generations (Slatkin and Barton, 1989; Barton et al., 2013). Both simulation and numerical results have shown that deme populations size has little effect on T_{IBD} (Fig. 1b) and on the relative contribution of geography on pairwise F_{st} (Fig. 7). This can be understood by the fact that i) the *spatial* term ($\bar{S}_{ij(t_*)}$) is independent from deme population size (*Eq. 4*), ii) the *drift-isolation* term ($\Delta\tilde{t}_0^{t_*}$) is smaller than the *spatial* term for any deme population size values, except under strong habitat loss and long time since HL&F (Fig. S12), and iii) the absolute decrease in $\tilde{t}_0^{t_*}$ is nearly identical among different values of deme population size values, except under strong habitat loss (Fig. S12).

Our simulation results have also shown that sampling scheme can contribute significantly to IBD and thus to influence T_{IBD}, when dispersal rate is relatively low (Fig. 1 and 2). In particular, we show that a scattered sampling within habitat fragments (*random*) contribute in maintaining IBD for longer after HL&F than a localized sampling

(*local*). The effect of sampling scheme on IBD was consistent among all HL&F scenarios here simulated, thus we give an interpretation of this result in this section, which will be valid for the other HL&F scenarios. In particular, we suggest that the longer T_{IBD} in *random* sampling is due to the higher expected coalescence times for *random* than *local* sampling scheme, as shown by measuring expected heterozygosity over time (Fig. S13). Referring to *Eq. 10*, this means that $\tilde{t}_0^{t_*}$ will be higher in *random* than *local* sampling, resulting in a lower contribution of the *drift-isolation term* on pairwise F_{st}. Instead, S_{ij} is expected to remain nearly unchanged between the two sampling schemes, given that *random* sampling collects genetic data with equal probability from the available subpopulations within habitat fragment. Therefore, changes in the temporal dynamic of $F_{st(i,j)}$ for the two-sampling scheme are mostly related to differences in the expected coalescence time for the alleles sampled within fragments.

3.5.3 Range expansion priori to instantaneous HL&F

Range expansion is a common process in nature (Hewitt, 1999; Lawson Handley et al., 2011), which has shaped spatial genetic patterns in several taxa across the world. It is characterized by successive events of colonization that reduce genetic diversity towards the direction of expansion (e.g., Austerlitz et al., 1997; Edmonds et al., 2004; Hallatschek and Nelson, 2008). A mechanism that can explain such decay in genetic diversity is called 'gene-surfing', which refers to the random spread of rare alleles that, by chance, can reach very high frequencies at the front of the expansion (Klopfstein et al., 2006). The non-homogenous spatial distribution of allele frequency due to a recent range expansion can generate large differences in allele frequency between the origin and the front-wave of the spatial expansion. This can in turn lead to higher genetic

differentiation (e.g., F_{st}) compared to an equivalent model at mutation-dispersal-drift equilibrium (Hallatschek et al., 2007). The consequences of range expansion on the spatial distribution of allele frequencies point out that both geographically-limited dispersal and range expansion can generate isolation-by-distance patterns. For instance, Meirmans, (2012) found through spatially-explicit simulations that in a colonization scenario from a single refugium, the percentage of significant tests in IBD were higher (44%) than in an equilibrium scenario (24%).

These series of studies led us to consider that range expansion may probably contribute to the persistence of IBD pattern after HL&F. Earlier simulation results by Mona et al., (2014) have, in fact, shown that under a scenario of *range expansion in a fragmented habitat*, F_{st} average and variance are higher than i) a non-fragmented habitat, and ii) an equivalent fragmented population that has not undergone range expansion. As expected, the strength of such discrepancy between equilibrium (*eq*) and expanded (*exp*) populations decreases with the age of the range expansion ($T_{since\ expansion}$ = 1000: $F_{st-exp} - F_{st-eq} \approx 0.05$ and $\sigma^2_{F_{st-exp}} - \sigma^2_{F_{st-eq}} \approx 0.30$; $T_{since\ expansion}$ = 3000: $F_{st-exp} - F_{st-eq} \approx 0.01$ and $\sigma^2_{F_{st-exp}} - \sigma^2_{F_{st-eq}} \approx 0.11$). In the present study we examined a different scenario, in which *habitat loss and fragmentation occurs after range expansion* (Fig. 2a). Similar to Mona et al., (2014), we have studied the impact of the age of the range expansion on the persistence of IBD patterns after HL&F. Our results suggest a significant effect of recent range expansion on IBD patterns, for instance, by increasing T_{IBD} of about 1000 generations in T_{exp} = 600 compared to an equilibrium model (Fig. 2a, Fig. S8a, c, Fig. S9). However, the influence of range expansion on T_{IBD} decreases with the age of the

onset of the expansion, converging to the values of an equivalent *instantaneous HL&F*.

As for the instantaneous HL&F scenario, we interpret the simulation results in terms of expected coalescence time. Following the theoretical results derived in Austerlitz et al., (1997) on the impact of colonization on coalescence times, intra-subpopulation coalescence times are expected to be lower in a colonizing population than in an equivalent population at equilibrium. Consequently, the average coalescence time for two alleles sampled within the same subpopulation ($\bar{t}_{00}$) is expected to be lower than an equilibrium population. Accordingly, also $\bar{t}_0$ will be lower than expected. However, in this case $\bar{t}_0$ is not the same across the whole population, but rather dependent on the location of the sampled subpopulation with respect to the origin of expansion. This is because the effect of colonization on intra-subpopulation coalescence times becomes stronger with the distance from the origin of colonization (see Fig. 6, 7 in Austerlitz et al., 1997). If we assume that colonization has no strong effect on S_{ij}, and considering that $\bar{t}_0$ decreases towards the direction of expansion, then from the expression of $F_{st(i,j)}$ (*Eq. 3*) we note that for $i \neq 0$ and $j \neq 0$, $\bar{t}_{00} - \bar{t}_0$ is greater than zero, increasing with the distance from the origin of colonization. Conversely, in an equivalent equilibrium model, $\bar{t}_{00} - \bar{t}_0$ is always equal to zero. This might explain why, before HL&F, the range expansion simulation results show a larger effect of geographical distance on pairwise R_{st} (or F_{st}) than an equilibrium model (Fig. S11a, b). We can then conclude that in the range expansion scenario the influence of geography on pairwise $F_{st(i,j)}$ is given by i) the relative contribution of S_{ij} on $F_{st(i,j)}$, and ii) the spatial gradient in $\bar{t}_0$ towards the direction of expansion, which, for demes far apart, results into pairwise $F_{st(i,j)}$ that are higher than expected under an equilibrium model.

After full colonization of the habitat (before HL&F), $\bar{t}_0$ within each deme tends to the same equilibrium value over time, which is determined by the total population size of the continuous habitat (Austerlitz et al., 1997). Similarly, $\bar{t}_{00}$ approaches to the value of a population at mutation-dispersal-drift equilibrium, such that $\bar{t}_{00} = \bar{t}_0$. This means that differences in $F_{st(i,j)}$ between equilibrium and range expansions scenario decrease significantly with the time since the onset of the expansion.

After HL&F, deme $\tilde{t}_0^{t_*}$ within each habitat patch tends to the same equilibrium value, given by the total population size of the habitat patch, which in our simulations is the same across patches. In the range expansion scenario, IBD is then expected to be lost once i) $\tilde{t}_0^{t_*}$ has reached the same value across habitat patches, regardless of their positions on the expansion gradient, and ii) once the relative contribution of the *spatial term* is much smaller than the *drift-isolation term*, as for the *instantaneous HL&F* scenario. Therefore, differences in T_{IBD} between *range expansion + instantaneous HL&F* and *instantaneous HL&F* scenarios could be explained by the fact that T_{IBD} in the *instantaneous HL&F* scenario only depends on the time required for the *spatial term* to become negligible compared to the *drift-isolation term*.

3.5.4 Two-steps HL&F

It is probably more realistic to consider that habitat loss and fragmentation is a gradual spatial-temporal process. Therefore, in the present study we also compared the emerging IBD patterns in the case of an *instantaneous HL&F versus* a *Two-steps HL&F*. The scenario that we simulated consider a hierarchical HL&F process, where a first HL&F event generates three clusters of forest fragments which then undergoes a second HL&F event (Fig. 2b). The variable of

interest in this case is the amount of time spent in the *intermediate state* between the first and the second HL&F events (T_{split}). Our simulation results have shown that *Two-steps HL&F* can significantly influence the temporal dynamic of IBD patterns (Fig. 2b, Fig. S8b, d, Fig. S10). As expected, the effect of *Two-steps HL&F* on IBD becomes stronger with increasing T_{split}, since it takes more time for IBD to be lost for higher T_{split}. Overall, this can be explained by the longer-in-time relative contribution of geography (S_{ij}) on pairwise R_{st}, the shorter t_* for habitat patches that were in the same cluster before the second HL&F event and by the slower decay in average coalescence time ($\tilde{t}_0^{t_*}$). Note that, after $T_{1^{st}\,HL\&F}$, both connected (within the same cluster) and fragmented (different clusters) subpopulation pairs show an increase in pairwise R_{st}. However, pairwise R_{st} between subpopulations that are within the same cluster after $T_{1^{st}\,HL\&F}$ does not increase as much as subpopulations that are at the same geographical distance but on different clusters (Fig. S11c, d). It is therefore useful to first distinguish the processes that influence pairwise F_{st} for subpopulations within or between clusters.

Increase in pairwise R_{st} for comparisons within clusters can be deduced from the expression of F_{st} in a continuous habitat (*Eq. 2*), which can be re-written as $F_{st(ij)} = 1 - \tilde{t}_0^{t_*}/(\tilde{t}_{00}^{t_*} + \tilde{S}_{ij}^{t_*})$. Since the considered subpopulations lie on the same cluster, $\tilde{t}_0^{t_*}$ is equal to $\tilde{t}_{00}^{t_*}$, and decrease at the same rate after $T_{1^{st}\,HL\&F}$. Also $\tilde{S}_{ij}^{t_*}$ decreases after $T_{1^{st}\,HL\&F}$, given that the amount of time required for two genes to be in the same deme (S_{ij}; *Eq. 4*) is positively correlated to the total number of demes in the habitat. According to this *rationale*, an increase in F_{st} could then be explained by a slower decrease in $\tilde{S}_{ij}^{t_*}$ compared to $\tilde{t}_0^{t_*} \approx \tilde{t}_{00}^{t_*}$.

Increase in pairwise R_{st} for comparisons between clusters can be explained following the arguments given for the *instantaneous HL&F*. In fact, under this condition, pairwise F_{st} evolve according to *Eq. 10*. However, we show that pairwise F_{st} between clusters increases slower than an equivalent comparison in the *instantaneous HL&F* (Fig. S11c, d). This is mainly due to the slower decrease in $\tilde{t}_0^{t_*}$ in the *Two-steps HL&F*, given that cluster size is about twice larger than patch size in the *instantaneous HL&F*.

The increase in T_{IBD} with T_{split} can thus be explained, compared to an equivalent *instantaneous HL&F* scenario, i) by the slower increase in pairwise R_{st} for both connected and fragmented subpopulations and ii) by the steeper IBD slope that is generated by the lower pairwise R_{st} of geographically close subpopulations, given that they were connected until $T_{2^{nd}\ HL\&F}$.

3.6 Conclusions

The present study aimed at clarifying some aspects on the influence of HL&F on spatial genetic structure, taking into account both space and time. To our knowledge, few studies have theoretically addressed this question with spatially-explicit simulations (Ezard and Travis, 2006; Cushman et al., 2012; Mona et al., 2014; Jackson and Fahrig, 2016). More importantly, currently there are no analytical models in population genetics which jointly model space and time in the context of HL&F.

A study that is often cited is Hutchison and Templeton, (1999), where the authors proposed four hypothetical relationships between F_{st} and geographical distance, which had become, in many instances in literature, a reference for interpreting isolation-by-distance pattern from natural populations. Hutchison and Templeton, (1999) proposed four ‘verbal’ models for which they found support by analyzing genetic

data from eastern collared lizards collected in four geographical regions in USA. This study has largely contributed to qualitatively classify the possible interactions between drift and gene-flow in species with geographically limited dispersal; however, it lacked a quantitative framework on which these hypotheses could rely on and being tested. With our study, we aimed to contribute and extend previous knowledge on the behavior of IBD in time and space, and provide an analytical framework to be explored according to the study case (e.g., landscape configuration).

Two of the theoretical study we cited Cushman et al., (2012) and Jackson and Fahrig, (2016) have investigated the effect of habitat amount or fragmentation (called 'configuration') on genetic structure through spatially-explicit simulations. The questions addressed in both studies aimed at identify if either habitat amount or habitat fragmentation would be better predictor of genetic structure. To do so, they performed individual-based simulations on landscapes with various degree of habitat amount and fragmentation, and correlated genetic structure (Mantel *r* values) with a range of fragmentation metrics. The two studies reached different conclusions. Cushman et al., (2012) found that habitat fragmentation has a higher effect on genetic structure than habitat area, whereas Jackson and Fahrig, (2016) found the opposite results. Jackson and Fahrig, (2016) discussed this discordance between the two studies considering, as also Cushman et al., (2012) point out in their work, that the fragmentation metrics used in Cushman et al., (2012) are strongly correlated with habitat amount. However, even Jackson and Fahrig, (2016) noted that although, overall, their results were pointing out to a minor role of fragmentation in shaping genetic structure, this was not correct for species with low dispersal rate. The analytical framework that we propose may help to such debate. Habitat fragmentation, defined as the breaking apart of a continuous landscape in habitat

fragments, does consequently reduce habitat amount. That is, once a complete barrier to gene-flow is formed (fragmentation) the only available habitat for a given population would be within the habitat fragment. The consequence is that the total population size of that population will be reduced ($\tilde{t}_0^{t_*}$), depending on the fragmentation configuration. Nevertheless, the total amount of habitat in the landscape do not necessarily change. In fact, the same reduction in total population size (and thus $\tilde{t}_0^{t_*}$) occurs even if that landscape is fragmented by sufficiently small barriers, such that landscape-scale habitat amount does not change. According to our model, an increase in pairwise R_{st} after habitat fragmentation is then mostly generated by the decay in average coalescence time within habitat fragments, suggesting that indeed total population size (within fragment) will play a major role in shaping genetic structure. However, this conclusion does not mean that habitat fragmentation has less importance than habitat amount, but rather that the two mechanisms are necessarily and often confounded.

3.6.1 Limitations of the study and Future Directions

Our simulation results have shown that IBD patterns can persist for long time after HL&F. The demographic history of populations, prior to HL&F, can influence the time of IBD loss. This was shown with two examples: *range expansion before HL&F* and *spatio-temporal variation in HL&F*. In the present study, we did not fully explore the interaction between several possible parameters in the *range expansion* or *spatio-temporal variation in HL&F* scenarios, but rather provide a somehow proof-of-concept and intuition on how and to which extent the past of a given population can influence spatial

genetic patterns. This question is not new (e.g., (Nichols, 1996; Hewitt, 1999)), but it was not yet explored in the context of HL&F.

The results from more complex scenarios of HL&F have shown that, under certain conditions, an *instantaneous HL&F* model might not be appropriate. Our analytical model for *isolation-by-distance in time* assume the same mode of *instantaneous HL&F*. Still, the simplifying assumptions of the simulation and analytical model have helped providing a first and general understanding of how space and time jointly contribute to spatial genetic patterns. Moreover, it is not necessarily unrealistic to consider an *instantaneous HL&F*, if we consider the current rate and mode of HL&F in natural habitats due to anthropogenic activities (Silva Junior et al., 2020; Vancutsem et al., 2021). Nevertheless, our analytical model may result sufficiently accurate if used to interpret pairwise F_{st} between geographically close habitat fragments, since those are likely to behave as an instantaneous HL&F scenario.

The model could be further extended by considering for instance the work of Austerlitz et al., (1997) on average coalescence time in a population under range expansion, and derive a framework that include both the impact of range expansion and HL&F on coalescence times and thus on pairwise F_{st}. Other improvements could consist on overcoming the assumption of homogenous subpopulation size and dispersal rate across the landscape, or complete isolation among habitat fragments by considering permeable barriers to gene-flow (Ringbauer et al., 2018). In addition, the development of an inferential framework aiming at estimating time since HL&F (t_*) based on, for instance, the method of moments or maximum-likelihood could be thought. Further extensions include also the use of the full distribution of coalescence times, instead of the expected coalescence times as done in the present study, which would allow then to explore more sophisticated information in the

genome such as the distribution of genome block sharing (identity-by-descent blocks; IBD blocks) (e.g., Ringbauer et al., 2017). Currently, the approach of IBD blocks is constrained to the availability of dense SNP or whole-genome data (reviewed in Leitwein et al., 2020), which at-present are not the norm for the great majority of non-model organisms, although in the next future this will likely change.

3.7 References

Aguillon, S.M., Fitzpatrick, J.W., Bowman, R., Schoech, S.J., Clark, A.G., Coop, G., Chen, N., 2017. Deconstructing isolation-by-distance: The genomic consequences of limited dispersal. PLOS Genetics 13, e1006911.

Al-Asadi, H., Petkova, D., Stephens, M., Novembre, J., 2019. Estimating recent migration and population-size surfaces. PLOS Genetics 15, e1007908.

Aleixo-Pais, I., Salmona, J., Sgarlata, G.M., Rakotonanahary, A., Sousa, A.P., Parreira, B., Kun-Rodrigues, C., Ralantoharijaona, T., Jan, F., Rasolondraibe, E., Minhós, T., Zaonarivelo, J.R., Andriaholinirina, N.V., Chikhi, L., 2019. The genetic structure of a mouse lemur living in a fragmented habitat in Northern Madagascar. Conserv Genet 20, 229–243.

Arbogast, B.S., Edwards, S.V., Wakeley, J., Beerli, P., Slowinski, J.B., 2002. Estimating Divergence Times from Molecular Data on Phylogenetic and Population Genetic Timescales. Annual Review of Ecology and Systematics 33, 707–740.

Archer, F.I., Adams, P.E., Schneiders, B.B., 2017. stratag: An r package for manipulating, summarizing and analysing population genetic data. Molecular Ecology Resources 17, 5–11.

Arredondo, A., Mourato, B., Nguyen, K., Boitard, S., Rodríguez, W., Noûs, C., Mazet, O., Chikhi, L., 2021. Inferring number of populations

and changes in connectivity under the n-island model. Heredity 126, 896–912.

Austerlitz, F., Jung-Muller, B., Godelle, B., Gouyon, P.-H., 1997. Evolution of Coalescence Times, Genetic Diversity and Structure during Colonization. Theoretical Population Biology 51, 148–164.

Baharian, S., Barakatt, M., Gignoux, C.R., Shringarpure, S., Errington, J., Blot, W.J., Bustamante, C.D., Kenny, E.E., Williams, S.M., Aldrich, M.C., Gravel, S., 2016. The Great Migration and African-American Genomic Diversity. PLOS Genetics 12, e1006059.

Bailey, S.F., Bataillon, T., 2016. Can the experimental evolution programme help us elucidate the genetic basis of adaptation in nature? Molecular Ecology 25, 203–218.

Barak, R.S., Hipp, A.L., Cavender-Bares, J., Pearse, W.D., Hotchkiss, S.C., Lynch, E.A., Callaway, J.C., Calcote, R., Larkin, D.J., 2016. Taking the Long View: Integrating Recorded, Archeological, Paleoecological, and Evolutionary Data into Ecological Restoration. International Journal of Plant Sciences 177, 90–102.

Barton, Kelleher, J., Etheridge, A.M., 2010. A new model for extinction and recolonization in two dimensions: quantifying phylogeography. Evolution 64, 2701–2715.

Barton, N.H., 2008. The effect of a barrier to gene flow on patterns of geographic variation. Genetics Research 90, 139–149.

Barton, N.H., Depaulis, F., Etheridge, A.M., 2002. Neutral Evolution in Spatially Continuous Populations. Theoretical Population Biology 61, 31-48.

Barton, N.H., Etheridge, A.M., Kelleher, J., Véber, A., 2013. Inference in two dimensions: Allele frequencies versus lengths of shared sequence blocks. Theoretical Population Biology, Coalescent Theory 87, 105–119.

Barton, N.H., Wilson, I., 1995. Genealogies and geography. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 349, 49–59.

Battey, C.J., Ralph, P.L., Kern, A.D., 2020. Space is the Place: Effects of Continuous Spatial Structure on Analysis of Population Genetic Data. Genetics 215, 193–214.

Beaumont, M.A., 1999. Detecting population expansion and decline using microsatellites. Genetics 153, 2013–2029.

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian Computation in Population Genetics. Genetics 162, 2025–2035.

Beichman, A.C., Huerta-Sanchez, E., Lohmueller, K.E., 2018. Using Genomic Data to Infer Historic Population Dynamics of Nonmodel Organisms. Annual Review of Ecology, Evolution, and Systematics 49, 433–456.

Bradburd, G.S., Coop, G.M., Ralph, P.L., 2018. Inferring Continuous and Discrete Population Genetic Structure Across Space. Genetics 210, 33–52.

Bradburd, G.S., Ralph, P.L., 2019. Spatial Population Genetics: It's About Time. Annual Review of Ecology, Evolution, and Systematics 50, 427–449.

Cayuela, H., Rougemont, Q., Prunier, J.G., Moore, J.-S., Clobert, J., Besnard, A., Bernatchez, L., 2018. Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review. Molecular Ecology 27, 3976–4010.

Ceballos, G., Ehrlich, P.R., Dirzo, R., 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. PNAS 114, E6089–E6096.

Chan, M., 2022. rwa: Perform a Relative Weights Analysis.

Chikhi, L., Rodríguez, W., Grusea, S., Santos, P., Boitard, S., Mazet, O., 2018. The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: insights into demographic inference and model choice. Heredity 120, 13–24.

Chikhi, L., Sousa, V.C., Luisi, P., Goossens, B., Beaumont, M.A., 2010. The Confounding Effects of Population Structure, Genetic Diversity and the Sampling Scheme on the Detection and Quantification of Population Size Changes. Genetics 186, 983–995.

Clark, L.V., Jasieniuk, M., 2011. POLYSAT: an R package for polyploid microsatellite analysis. Mol Ecol Resour 11, 562–566.

Cooper, A., Turney, C.S.M., Palmer, J., Hogg, A., McGlone, M., Wilmshurst, J., et al., 2021. A global environmental crisis 42,000 years ago. Science 371, 811–818.

Costa, R.J., Wilkinson-Herbots, H., 2017. Inference of Gene Flow in the Process of Speciation: An Efficient Maximum-Likelihood Method for the Isolation-with-Initial-Migration Model. Genetics 205,1597-1618.

Cox, J.T., Durrett, R., 2002. The stepping stone model: New formulas expose old myths. The Annals of Applied Probability 12, 1348–1377.

Currat, M., Ray, N., Excoffier, L., 2004. splatche: a program to simulate genetic diversity taking into account environmental heterogeneity. Molecular Ecology Notes 4, 139–142.

Cushman, S.A., Shirk, A., Landguth, E.L., 2012. Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex landscapes. Landscape Ecol 27, 369–380.

Duforet-Frebourg, N., Slatkin, M., 2016. Isolation-by-distance-and-time in a stepping-stone model. Theoretical Population Biology 108, 24-35.

Eckert, C.G., Samis, K.E., Lougheed, S.C., 2008. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. Molecular Ecology 17, 1170–1188.

Edmonds, C.A., Lillie, A.S., Cavalli-Sforza, L.L., 2004. Mutations arising in the wave front of an expanding population. PNAS 101, 975-979.

Epps, C.W., Keyghobadi, N., 2015. Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. Molecular Ecology 24, 6021–6040.

Excoffier, L., Foll, M., 2011. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics 27, 1332–1334.

Excoffier, L., Ray, N., 2008. Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol Evol 23, 347–351.

Exposito-Alonso, M., Rodríguez, R.G., Barragán, C., Capovilla, G., Chae, E., et al., 2018. A rainfall-manipulation experiment with 517 Arabidopsis thaliana accessions. bioRxiv 186767.

Ezard, T.H.G., Travis, J.M.J., 2006. The Impact of Habitat Loss and Fragmentation on Genetic Drift and Fixation Time.Oikos 114,367-375.

Fordham, D.A., Akçakaya, H.R., Alroy, J., Saltré, F., Wigley, T.M.L., Brook, B.W., 2016. Predicting and mitigating future biodiversity loss using long-term ecological proxies.Nature Climate Change 6,909-916.

Grace, M., Akçakaya, H.R., Bennett, E., Hilton-Taylor, C., Long, B., Milner-Gulland, E.J., Young, R., Hoffmann, M., 2019. Using historical and palaeoecological data to inform ambitious species recovery targets. Philosophical Transactions of the Royal Society B: Biological Sciences 374, 20190297.

Grosser, S., Abdelkrim, J., Wing, J., Robertson, B.C., Gemmell, N.J., 2017. Strong isolation by distance argues for separate population management of endangered blue duck (Hymenolaimus malacorhynchos). Conserv Genet 18, 327–341.

Guindon, S., Guo, H., Welch, D., 2016. Demographic inference under the coalescent in a spatial continuum. Theoretical Population Biology 111, 43–50.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., Bustamante, C.D., 2009. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. PLOS Genetics 5, e1000695.

Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D., Cook, W.M., Damschen, E.I., Ewers, R.M., Foster, B.L., Jenkins, C.N., King, A.J., Laurance, W.F., Levey, D.J., Margules, C.R., Melbourne, B.A., Nicholls, A.O., Orrock, J.L., Song, D.-X., Townshend, J.R., 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. Science Advances 1, e1500052.

Hallatschek, O., Hersen, P., Ramanathan, S., Nelson, D.R., 2007. Genetic drift at expanding frontiers promotes gene segregation. PNAS 104, 19926–19930.

Hallatschek, O., Nelson, D.R., 2008. Gene surfing in expanding populations. Theoretical Population Biology 73, 158–170.

Hardy, O.J., Vekemans, X., 1999. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity 83, 145–154.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907–913.

Hewitt, G.M., 1999. Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society 68, 87–112.

Hewitt, G.M., 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society 58, 247–276.

Hey, J., 1991. A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. Theoretical Population Biology 39, 30–48.

Hey, J., Nielsen, R., 2004. Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, With Applications to the Divergence of Drosophila pseudoobscura and D. persimilis. Genetics 167, 747–760.

Hudson, R.R., Slatkin, M., Maddison, W.P., 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132, 583–589.

Hutchison, D.W., Templeton, A.R., 1999. Correlation of Pairwise Genetic and Geographic Distance Measures: Inferring the Relative Influences of Gene Flow and Drift on the Distribution of Genetic Variability. Evolution 53, 1898–1914.

Ibrahim, K.M., Nichols, R.A., Hewitt, G.M., 1996. Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. Heredity 77, 282–291.

Jackson, N.D., Fahrig, L., 2016. Habitat amount, not habitat configuration, best predicts population genetic structure in fragmented landscapes. Landscape Ecol 31, 951–968.

Juřičková, L., Horáčková, J., Ložek, V., 2014. Direct evidence of central European forest refugia during the last glacial period based on mollusc fossils. Quaternary Research 82, 222–228.

Kawecki, T.J., Lenski, R.E., Ebert, D., Hollis, B., Olivieri, I., Whitlock, M.C., 2012. Experimental evolution. Trends in Ecology & Evolution 27, 547–560.

Keyghobadi, N., Roland, J., Matter, S.F., Strobeck, C., 2005. Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. Proceedings of the Royal Society B: Biological Sciences 272,553-560.

Killick, R., Eckley, I.A., 2014. changepoint: An R Package for Changepoint Analysis. Journal of Statistical Software 58, 1–19.

Killick, R., Fearnhead, P., Eckley, I.A., 2011. Optimal detection of changepoints with a linear computational cost. arXiv e-prints 1101, arXiv:1101.1438.

Kimura, M., Weiss, G.H., 1964. The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. Genetics 49, 561–576.

Klopfstein, S., Currat, M., Excoffier, L., 2006. The fate of mutations surfing on the wave of a range expansion. Mol Biol Evol 23, 482–490.

Krummel, J.R., Gardner, R.H., Sugihara, G., O'Neill, R.V., Coleman, P.R., 1987. Landscape Patterns in a Disturbed Environment. Oikos 48, 321–324.

Landguth, E.L., Cushman, S.A., Schwartz, M.K., McKelvey, K.S., Murphy, M., Luikart, G., 2010. Quantifying the lag time to detect barriers in landscape genetics. Mol. Ecol. 19, 4179–4191.

Lattin, G. de, 1967. Grundriss der Zoogeographie. VEB Gustav Fischer Verlag, Jena.

Lawson Handley, L.-J., Estoup, A., Evans, D.M., Thomas, C.E., Lombaert, E., Facon, B., Aebi, A., Roy, H.E., 2011. Ecological genetics of invasive alien species. BioControl 56, 409–428.

Leitwein, M., Duranton, M., Rougemont, Q., Gagnaire, P.-A., Bernatchez, L., 2020. Using Haplotype Information for Conservation Genomics. Trends in Ecology & Evolution 35, 245–258.

Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. Nature 475, 493–496.

Liu, X., Fu, Y.-X., 2015. Exploring population size changes using SNP frequency spectra. Nat. Genet. 47, 555–559.

Loog, L., 2021. Sometimes hidden but always there: the assumptions underlying genetic inference of demographic histories. Philosophical

Transactions of the Royal Society B: Biological Sciences 376, 20190719.

Louys, J., Braje, T.J., Chang, C.-H., Cosgrove, R., Fitzpatrick, S.M., Fujita, M., et al., 2021. No evidence for widespread island extinctions after Pleistocene hominin arrival. PNAS 118.

Lyman, R.L., 2017. Paleoenvironmental Reconstruction from Faunal Remains: Ecological Basics and Analytical Assumptions. J Archaeol Res 25, 315–371.

Malécot, G., 1948. Les Mathématiques de L'hérédité. Masson.

Mandelbrot, B.B., 1982. The Fractal Geometry of Nature, 2nd prt. edition. ed. Times Books, San Francisco.

Maruyama, T., 1970. Effective number of alleles in a subdivided population. Theor Popul Biol 1, 273–306.

Maynard-Smith, J., Slatkin, M., 1973. The Stability of Predator-Prey Systems. Ecology 54, 384–391.

Mazet, O., Rodríguez, W., Grusea, S., Boitard, S., Chikhi, L., 2016. On the importance of being structured: instantaneous coalescence rates and human evolution--lessons for ancestral population size inference? Heredity (Edinb) 116, 362–371.

McRae, B.H., 2006. Isolation by Resistance. Evolution 60, 1551–1561

Meirmans, P.G., 2012. The trouble with isolation by distance. Molecular Ecology 21, 2839–2846.

Milne, B.T., 1988. Measuring the fractal geometry of landscapes. Applied Mathematics and Computation 27, 67–79.

Mitchell, K.J., Rawlence, N.J., 2021. Examining Natural History through the Lens of Palaeogenomics. Trends in Ecology & Evolution 36, 258–267.

Mona, S., Ray, N., Arenas, M., Excoffier, L., 2014. Genetic consequences of habitat fragmentation during a range expansion. Heredity 112, 291–299.

Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Lysenko, I., Senior, R.A., Börger, L., et al., 2015. Global effects of land use on local terrestrial biodiversity. Nature 520, 45–50.

Nichols, R., 1996. Is it ancient or modern history that we can read in the genes? pp. 69–87.

Nielsen, R., Beaumont, M.A., 2009. Statistical inferences in phylogeography. Mol Ecol 18, 1034–1047.

Nielsen, R., Wakeley, J., 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics 158, 885–896.

Nikolic, N., Chevalet, C., 2014. Detecting past changes of effective population size. Evolutionary Applications 7, 663–681.

Ohta, T., Kimura, M., 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet Res 22, 201–204.

Ortega-Del Vecchyo, D., Slatkin, M., 2019. F ST between archaic and present-day samples. Heredity 122, 711–718.

Pauls, S.U., Nowak, C., Bálint, M., Pfenninger, M., 2013. The impact of global climate change on genetic diversity within populations and species. Molecular Ecology 22, 925–946.

Pertoldi, C., Ruiz-Gonzalez, A., Bahrndorff, S., Renee Lauridsen, N., Nisbeth Henriksen, T., Eskildsen, A., Høye, T.T., 2021. Strong isolation by distance among local populations of an endangered butterfly species (Euphydryas aurinia). Ecology and Evolution 11, 12790–12800.

Petkova, D., Novembre, J., Stephens, M., 2016. Visualizing spatial population structure with estimated effective migration surfaces. Nature Genetics 48, 94–100.

Racimo, F., Woodbridge, J., Fyfe, R.M., Sikora, M., Sjögren, K.-G., Kristiansen, K., Linden, M.V., 2020. The spatiotemporal spread of human migrations during the European Holocene. PNAS 117, 8989–9000.

Ralph, P., Coop, G., 2013. The Geography of Recent Genetic Ancestry across Europe. PLOS Biology 11, e1001555.

Rasteiro, R., Bouttier, P.-A., Sousa, V.C., Chikhi, L., 2012. Investigating sex-biased migration during the Neolithic transition in Europe, using an explicit spatial simulation framework. Proceedings of the Royal Society B: Biological Sciences 279, 2409–2416.

Ringbauer, H., Coop, G., Barton, N.H., 2017. Inferring Recent Demography from Isolation by Distance of Long Shared Sequence Blocks. Genetics 205, 1335–1351.

Ringbauer, H., Kolesnikov, A., Field, D.L., Barton, N.H., 2018. Estimating Barriers to Gene Flow from Distorted Isolation-by-Distance Patterns. Genetics 208, 1231–1245.

Robledo-Arnuncio, J.J., Rousset, F., 2010. Isolation by distance in a continuous population under stochastic demographic fluctuations. Journal of Evolutionary Biology 23, 53–71.

Román-Palacios, C., Wiens, J.J., 2020. Recent responses to climate change reveal the drivers of species extinction and survival. PNAS 117, 4211–4217.

Roser, L.G., Ferreyra, L.I., Ewens, M., Vilardi, J.C., Saidman, B.O., 2017. Isolation by distance and stochastic migration in a fragmented population of Prosopis alba. American Journal of Botany 104, 313–321.

Rousset, 2000. Genetic differentiation between individuals. Journal of Evolutionary Biology 13, 58–62.

Rousset, F., 1997. Genetic Differentiation and Estimation of Gene Flow from F-Statistics Under Isolation by Distance. Genetics 145, 1219–1228.

Rousset, F., Leblois, R., 2012. Likelihood-Based Inferences under Isolation by Distance: Two-Dimensional Habitats and Confidence Intervals. Molecular Biology and Evolution 29, 957–973.

Saltré, F., Rodríguez-Rey, M., Brook, B.W., Johnson, C.N., Turney, C.S.M., Alroy, J.,et al., 2016. Climate change not to blame for late Quaternary megafauna extinctions in Australia. Nature Communications 7, 10511.

Sandom, C., Faurby, S., Sandel, B., Svenning, J.-C., 2014. Global late Quaternary megafauna extinctions linked to humans, not climate change. Proceedings of the Royal Society B: Biological Sciences 281, 20133254.

Schneider, N., Chikhi, L., Currat, M., Radespiel, U., 2010. Signals of recent spatial expansions in the grey mouse lemur (Microcebus murinus). BMC Evolutionary Biology 10, 105.

Sexton, J.P., Hangartner, S.B., Hoffmann, A.A., 2014. Genetic Isolation by Environment or Distance: Which Pattern of Gene Flow Is Most Common? Evolution 68, 1–15.

Silva Junior, C.H.L., Aragão, L.E.O.C., Anderson, L.O., Fonseca, M.G., Shimabukuro, Y.E., Vancutsem, C., Achard, F., Beuchle, R., Numata, I., Silva, C.A., Maeda, E.E., Longo, M., Saatchi, S.S., 2020. Persistent collapse of biomass in Amazonian forest edges following deforestation leads to unaccounted carbon losses. Science Advances 6, eaaz8360.

Slatkin, M., 2005. Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. Molecular Ecology 14, 67–73.

Slatkin, M., 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139, 457–462.

Slatkin, M., 1993. Isolation by Distance in Equilibrium and Non-Equilibrium Populations. Evolution 47, 264–279.

Slatkin, M., 1991. Inbreeding coefficients and coalescence times. Genetics Research 58, 167–175.

Slatkin, M., Barton, N.H., 1989. A COMPARISON OF THREE INDIRECT METHODS FOR ESTIMATING AVERAGE LEVELS OF GENE FLOW. Evolution 43, 1349–1368.

Slatkin, M., Hudson, R.R., 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129, 555–562.

Solé, R.V., Manrubia, S.C., 1995. Are rainforests self-organized in a critical state? Journal of Theoretical Biology 173, 31–40.

Städler, T., Haubold, B., Merino, C., Stephan, W., Pfaffelhuber, P., 2009. The Impact of Sampling Schemes on the Site Frequency Spectrum in Nonequilibrium Subdivided Populations. Genetics 182, 205–216.

Stewart, M., Carleton, W.C., Groucutt, H.S., 2021. Climate change, not human population growth, correlates with Late Quaternary megafauna declines in North America. Nat Commun 12, 965.

Strobeck, C., 1987. Average Number of Nucleotide Differences in a Sample from a Single Subpopulation: A Test for Population Subdivision. Genetics 117, 149–153.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G., Cosson, J.-F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology 7, 453–464.

Tonidandel, S., Lebreton, J.M., Johnson, J.W., 2009. Determining the statistical significance of relative weights. Psychol Methods 14, 387–399.

Tóth, E.G., Tremblay, F., Housset, J.M., Bergeron, Y., Carcaillet, C., 2019. Geographic isolation and climatic variability contribute to genetic differentiation in fragmented populations of the long-lived subalpine conifer Pinus cembra L. in the western Alps. BMC Evolutionary Biology 19, 190.

Vancutsem, C., Achard, F., Pekel, J.-F., Vieilledent, G., Carboni, S., Simonetti, D., Gallego, J., Aragão, L.E.O.C., Nasi, R., 2021. Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. Science Advances 7, eabe1603.

Vekemans, X., Hardy, O.J., 2004. New insights from fine-scale spatial genetic structure analyses in plant populations. Molecular Ecology 13, 921–935.

Wakeley, J., 2000. The Effects of Subdivision on the Genetic Divergence of Populations and Species. Evolution 54, 1092–1101.

Wiens, J.J., Ackerly, D.D., Allen, A.P., Anacker, B.L., Buckley, L.B., Cornell, H.V., Damschen, E.I., Davies, T.J., Grytnes, J.-A., Harrison, S.P., Hawkins, B.A., Holt, R.D., McCain, C.M., Stephens, P.R., 2010. Niche conservatism as an emerging principle in ecology and conservation biology. Ecology Letters 13, 1310–1324.

Wilkins, J.F., 2004. A Separation-of-Timescales Approach to the Coalescent in a Continuous Population. Genetics 168, 2227–2244.

Wilkins, J.F., Wakeley, J., 2002. The Coalescent in a Continuous, Finite, Linear Population. Genetics 161, 873–888.

Wilkinson-Herbots, H.M., 1998. Genealogy and subpopulation differentiation under various models of population structure. J Math Biol 37, 535–585.

Wright, S., 1946. Isolation by Distance under Diverse Systems of Mating. Genetics 31, 39–59.

Wright, S., 1943. Isolation by Distance. Genetics 28, 114–138.

Wright, S., 1940. Breeding Structure of Populations in Relation to Speciation. The American Naturalist 74, 232–248.

Zhao, Y., Vrieling, K., Liao, H., Xiao, M., Zhu, Y., Rong, J., Zhang, W., Wang, Y., Yang, J., Chen, J., Song, Z., 2013. Are habitat fragmentation, local adaptation and isolation-by-distance driving population divergence in wild rice Oryza rufipogon? Mol Ecol 22, 5531–5547.

3.8 Supplementary Figures

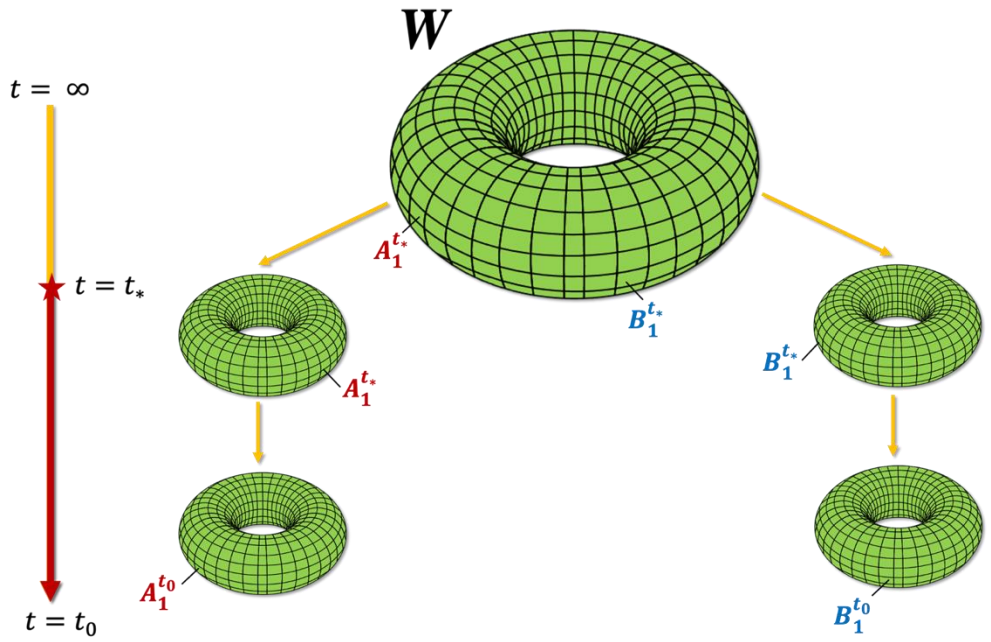


Figure S1. Illustration of the isolation-by-distance in time model in the context of habitat loss and fragmentation. $\boldsymbol{A}_1^{t_0}$and $\boldsymbol{B}_1^{t_0}$ indicate the location of individuals sampled in habitat fragments A and B, respectively, at time $t = t_0$. $\boldsymbol{A}_1^{t_*}$ and $\boldsymbol{B}_1^{t_*}$ indicate one possible location of the ancestral lineages at the time of HL&F ($t = t_*$) for individuals sampled in habitat fragments A and B, respectively. W: finite two-dimensional habitat prior to HL&F. $t = \infty$ specifies that population W is at mutation-drift-dispersal equilibrium.

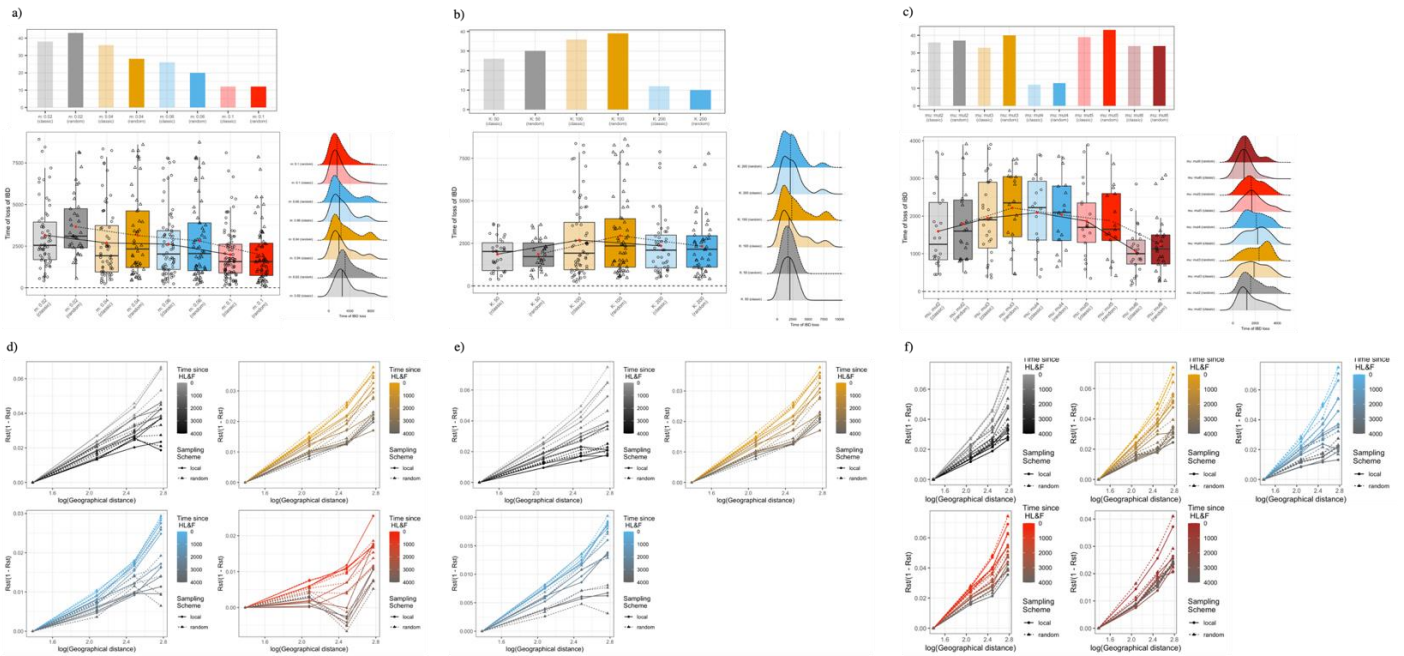


Figure S2. The effect of habitat loss and fragmentation on R_{st}-based isolation-by-distance (IBD) in relation to dispersal rate, deme population size and mutation rate. Time at which IBD is no longer significant (T_{IBD}) as a function of sampling scheme (*local* vs *random*), **a)** dispersal rate ($m = 0.02 – 0.01$, $K = 100$, $\mu = 5 \times 10^{-4}$), **b)** deme population size ($K = 50 – 200$, $m = 0.04$, $\mu = 5 \times 10^{-4}$) and **c)** mutation rate ($\mu = 10^{-2} – 10^{-6}$, $K = 50$, $m = 0.04$). Dots indicate the T_{IBD} estimated for each simulation replicate on the basis of 30 microsatellites. Density plots equivalent to box plot results are shown for visualization purposes. Upper histograms represent the number of replicates for which no T_{IBD} could be identified, because IBD was not lost within the time of the simulation. Mean pairwise R_{st} as function of log(geographic distance) over time, distinguishing random (dashed line) and local (solid line) sampling in relation to **d)** dispersal rate, **e)** deme population size and **f)** mutation rate.

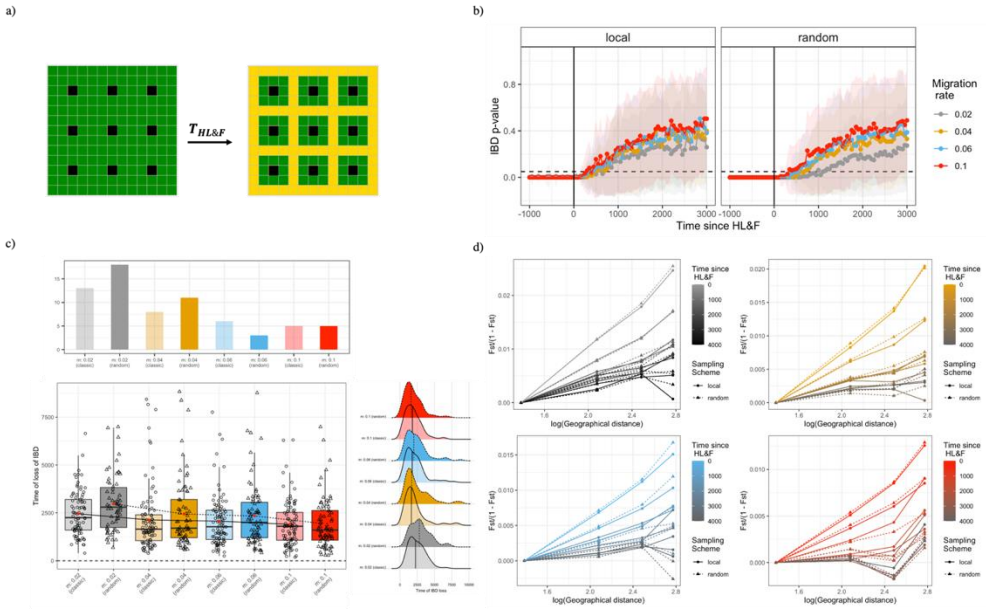


Figure S3. The effect of habitat loss and fragmentation on F_{st}-based isolation-by-distance (IBD) in relation to dispersal rate. a) Simulated 13 x 13 square grid of demes ($d_{tot} = 169$) undergoing HL&F at time $T_{HL\&F}$, sampling genetic data every 50 generations. Black cells refer to the sampled demes, green and yellow cells refer to suitable and unsuitable habitat, respectively. Dispersal rate values between 0.02 and 0.1 were tested, while fixing $K = 100$ and $\mu = 5$ x 10^{-4}. **b)** Mean (dots) and standard deviation (shaded area) of IBD p-value across generations. **c)** Time at which IBD is no longer significant for each simulation replicate as a function of dispersal rate and sampling scheme (*local* vs *random*). Density plots equivalent to box plot results are shown for visualization purposes. Upper histograms represent the number of replicates for which no T_{IBD} could be identified, because IBD was not lost within the time of the simulation. **d)** Mean pairwise F_{st} as function of *log*(geographic distance) over time, distinguishing *random* (dashed line) and *local* (solid line) sampling. The results suggest that IBD is lost quicker at high dispersal rate.

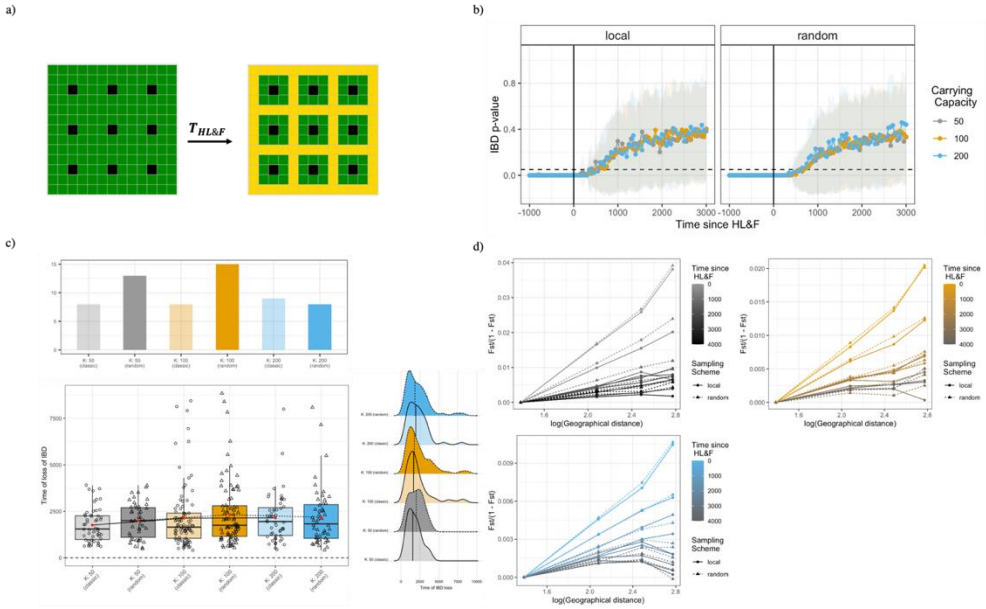


Figure S4. The effect of habitat loss and fragmentation on F_{st}-based isolation-by-distance (IBD) in relation to deme population size. a) Simulated 13 x 13 square grid of demes ($d_{tot} = 169$) undergoing HL&F at time $T_{HL\&F}$, sampling genetic data every 50 generations. Black cells refer to the sampled demes, green and yellow cells refer to suitable and unsuitable habitat, respectively. Deme population size values between 50 and 200 were tested, while fixing $m = 0.02$ and $\mu = 5 \times 10^{-4}$. **b)** Mean (dots) and standard deviation (shaded area) of IBD p-value across generations. **c)** Time at which IBD is no longer significant for each simulation replicate as a function of deme population size rate and sampling scheme (local vs random). Density plots equivalent to box plot results are shown for visualization purposes. Upper histograms represent the number of replicates for which no T_{IBD} could be identified, because IBD was not lost within the time of the simulation. **d)** Mean pairwise F_{st} as function of log(geographic distance) over time, distinguishing random (dashed line) and local (solid line) sampling. The results suggest that deme population size has little effect on the time of IBD loss.

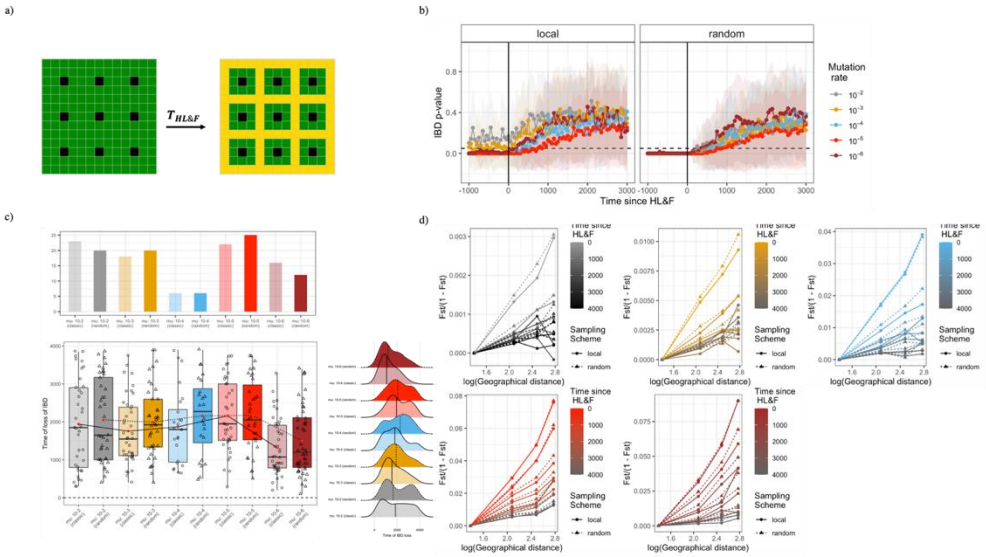


Figure S5. The effect of habitat loss and fragmentation on F_{st}-based isolation-by-distance (IBD) in relation to mutation rate. a) Simulated 13 x 13 square grid of demes ($d_{tot} = 169$) undergoing HL&F at time $T_{HL\&F}$, sampling genetic data every 50 generations. Black cells refer to the local sampling scheme, green and yellow cells refer to suitable and unsuitable habitat, respectively. Mutation rate values between 5 x 10^{-2} and 5 x 10^{-6} were tested, while fixing $K = 50$ and $m = 0.04$. **b)** Mean (dots) and standard deviation (shaded area) of IBD p-value across generations. **c)** Time at which IBD is no longer significant as a function of mutation rate and sampling scheme (local vs random). Density plots equivalent to box plot results are shown for visualization purposes. Upper histograms represent the number of replicates for which no T_{IBD} could be identified, because IBD was not lost within the time of the simulation. **d)** Mean pairwise F_{st} as function of log(geographic distance) over time, distinguishing random (dashed line) and local (solid line) sampling. The results suggest that IBD is lost quicker at high dispersal rate.

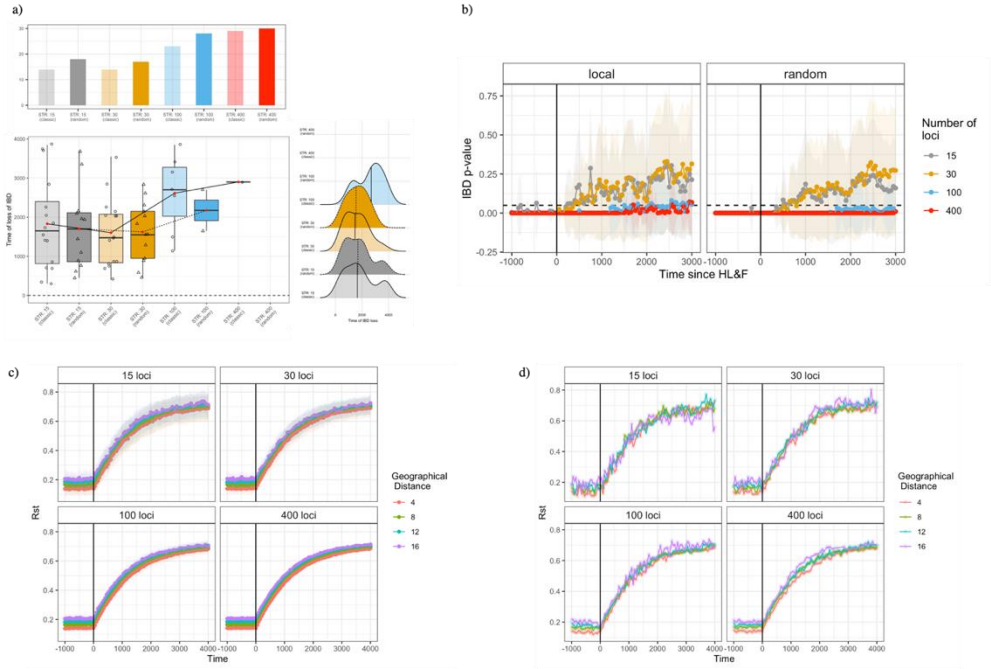


Figure S6. Influence of the number of genotyped markers on the time of IBD loss since HL&F, based on R_{st} isolation-by-distance. a) Time at which IBD is no longer significant as a function of the number of markers (15 – 400) and sampling scheme (*local* vs *random*). Density plots equivalent to box plot results are shown for visualization purposes. Upper histograms represent the number of replicates for which no T_{IBD} could be identified, because IBD was not lost within the time of the simulation. **b)** Mean (dots) and standard deviation (shaded area) of IBD p-value across generations. **c)** Pairwise R_{st} over time as function of geographical distance averaged across simulation replicates (Mean: dots, standard deviation: shade). **d)** Pairwise R_{st} over time as function of geographical distance for a single simulation replicate.

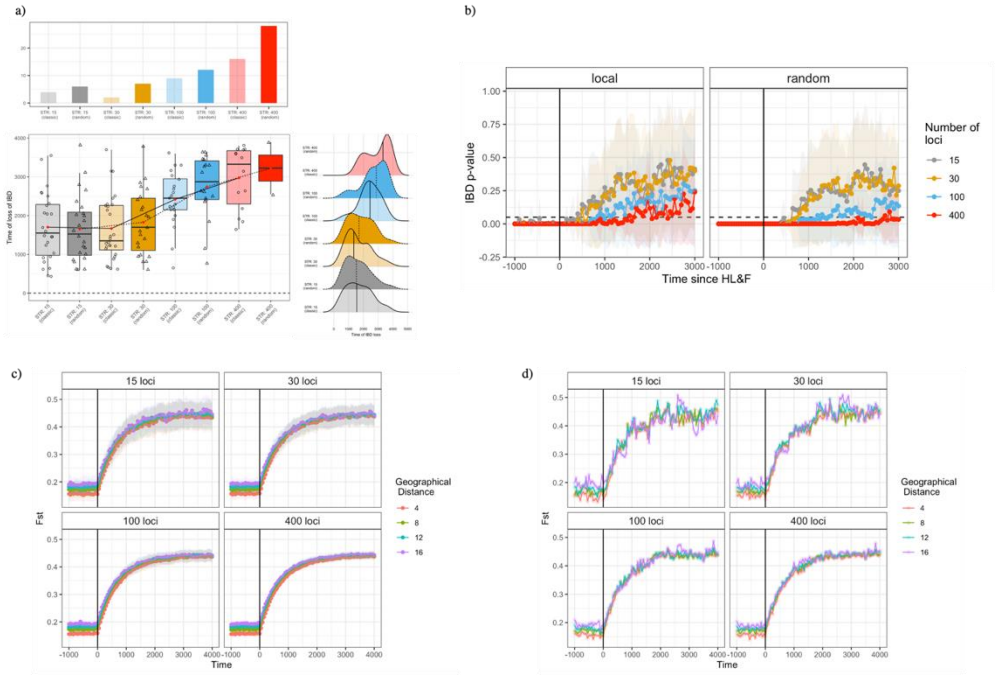


Figure S7. Influence of the number of genotyped markers on the time of IBD loss since HL&F, based on F_{st} isolation-by-distance. a) Time at which IBD is no longer significant as a function of the number of markers (15 – 400) and sampling scheme (*local* vs *random*). Density plots equivalent to box plot results are shown for visualization purposes. Upper histograms represent the number of replicates for which no T_{IBD} could be identified, because IBD was not lost within the time of the simulation. **b)** Mean (dots) and standard deviation (shaded area) of IBD p-value across generations. **c)** Pairwise F_{st} over time as function of geographical distance averaged across simulation replicates (Mean: dots, standard deviation: shade). **d)** Pairwise F_{st} over time as function of geographical distance for a single simulation replicate.

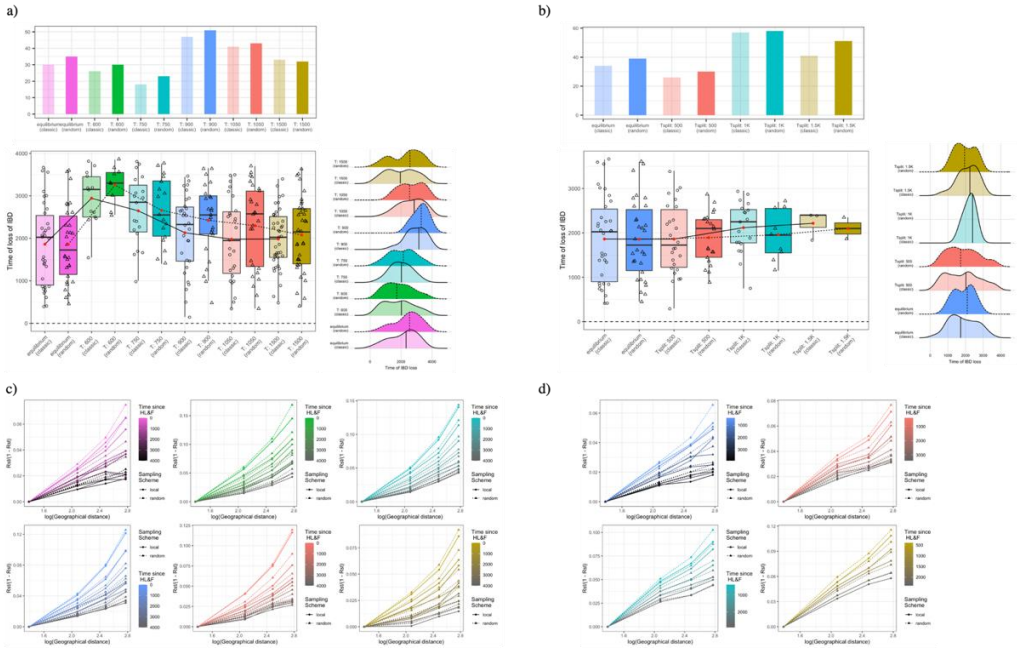


Figure S8. The effect of ancient range expansion prior to HL&F and spatio-temporal variation in HL&F on loss of R_{st}-based isolation-by-distance (IBD). Time at which IBD is no longer significant (T_{IBD}) for each simulation replicate as a function of sampling scheme (*local* vs *random*), **a)** time since the onset of the range expansion (T_{exp} = 600 – 1500, $K = 50$, $m = 0.04$ and μ = 5 x 10^{-4}), **b)** the time since the 1st HL&F event (T_{split} =500 – 1500, $K = 50$, $m = 0.04$ and μ = 5 x 10^{-4}). Density plots equivalent to box plot results are shown for visualization purposes. Upper histograms represent the number of replicates for which no T_{IBD} could be identified, because IBD was not lost within the time of the simulation. Mean pairwise R_{st} as function of log(geographic distance) over time, distinguishing *random* (dashed line) and *local* (solid line) sampling, and **c)** time since the onset of the range expansion or **d)** time since the 1st HL&F event.

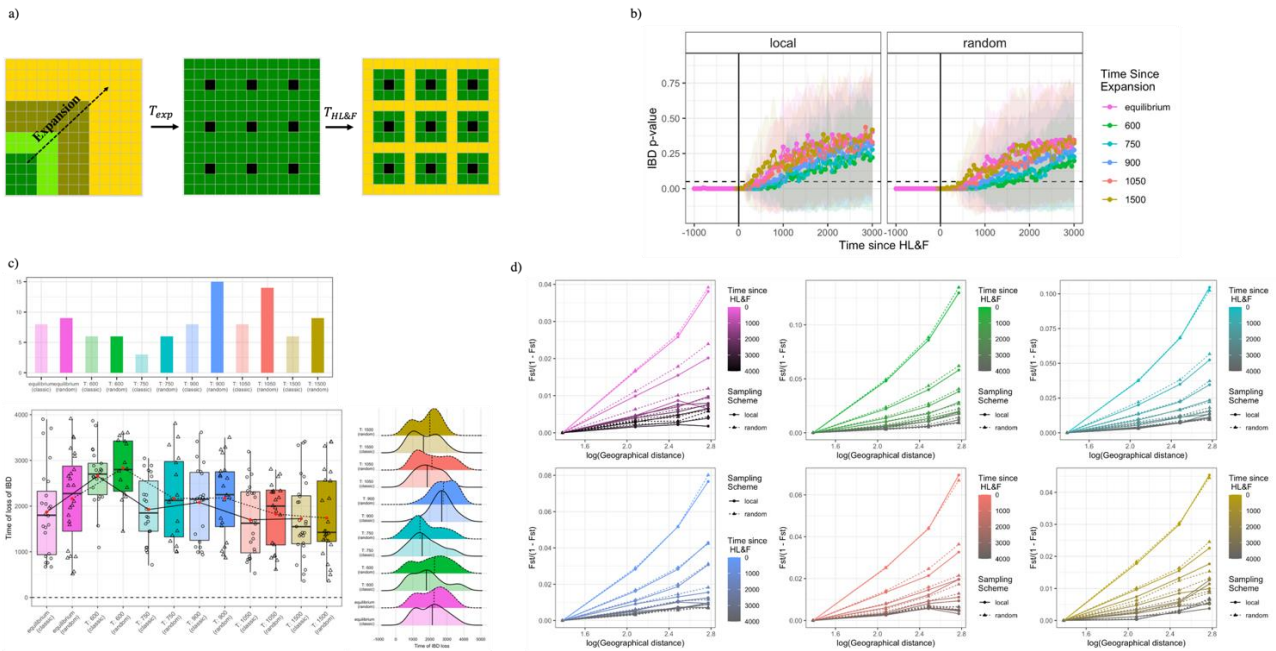


Figure S9. The effect of ancient range expansion prior to HL&F on loss of F_{st}-based isolation-by-distance. a) Simulated 13 x 13 square grid of demes, where T_{exp} is the time since the onset of the range expansion and $T_{HL\&F}$ is the time of instantaneous HL&F. Range expansion was simulated for a scenario with $K = 50$, $m = 0.04$ and $\mu = 5 \times 10^{-4}$. **b**) Mean (dots) and standard deviation (shaded area) of IBD p-value across generations. **c**) Time at which IBD is no longer significant (T_{IBD}) for each simulation replicate as a function of sampling scheme (*local* vs *random*) and time since the onset of the range expansion. Density plots equivalent to box plot results are shown for visualization purposes. Upper histograms represent the number of replicates for which no T_{IBD} could be identified, because IBD was not lost within the time of the simulation. **d**) Mean pairwise F_{st} as function of log(geographic distance) over time, distinguishing *random* (dashed line) and *local* (solid line) sampling, and time since the onset of the range expansion.

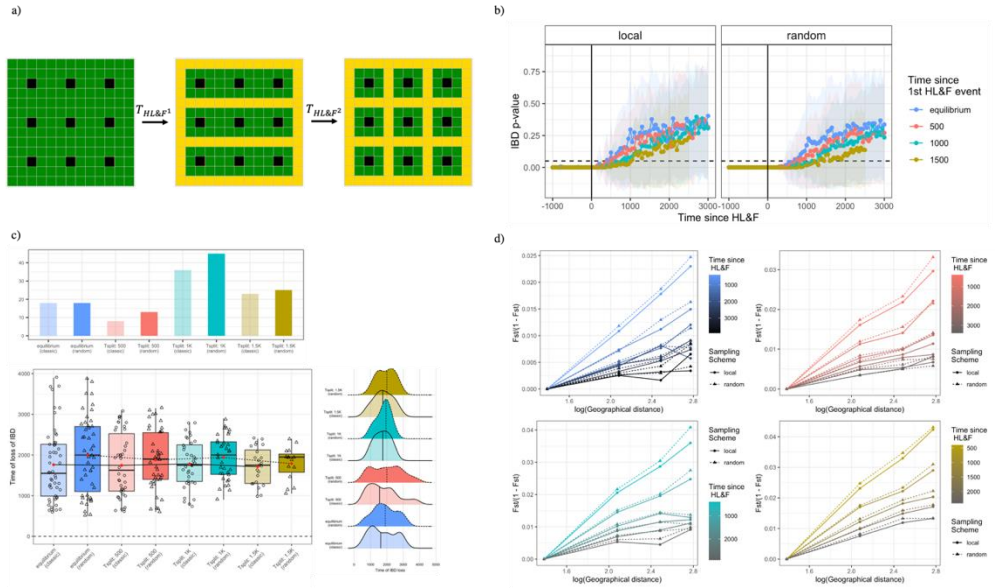


Figure S10. The effect of spatio-temporal variation in HL&F on loss of F_{st}-based isolation-by-distance. a) Simulated 13 x 13 square grid of demes, where $T_{HL\&F^1}$ and $T_{HL\&F^2}$ indicate, respectively, the time of the 1st and 2nd HL&F event. Piecewise HL&F was simulated for a scenario with $K = 50$, $m = 0.04$ and $\mu = 5 \times 10^{-4}$. **b**) Mean (dots) and standard deviation (shaded area) of IBD p-value across generations. **c**) Time at which IBD is no longer significant (T_{IBD}) for each simulation replicate as a function of sampling scheme (*local* vs *random*) and time since the 1st HL&F event. Density plots equivalent to box plot results are shown for visualization purposes. Upper histograms represent the number of replicates for which no T_{IBD} could be identified, because IBD was not lost within the time of the simulation. **d**) Mean pairwise F_{st} as function of log(geographic distance) over time, distinguishing *random* (dashed line) and *local* (solid line) sampling, and time since the 1st HL&F event.

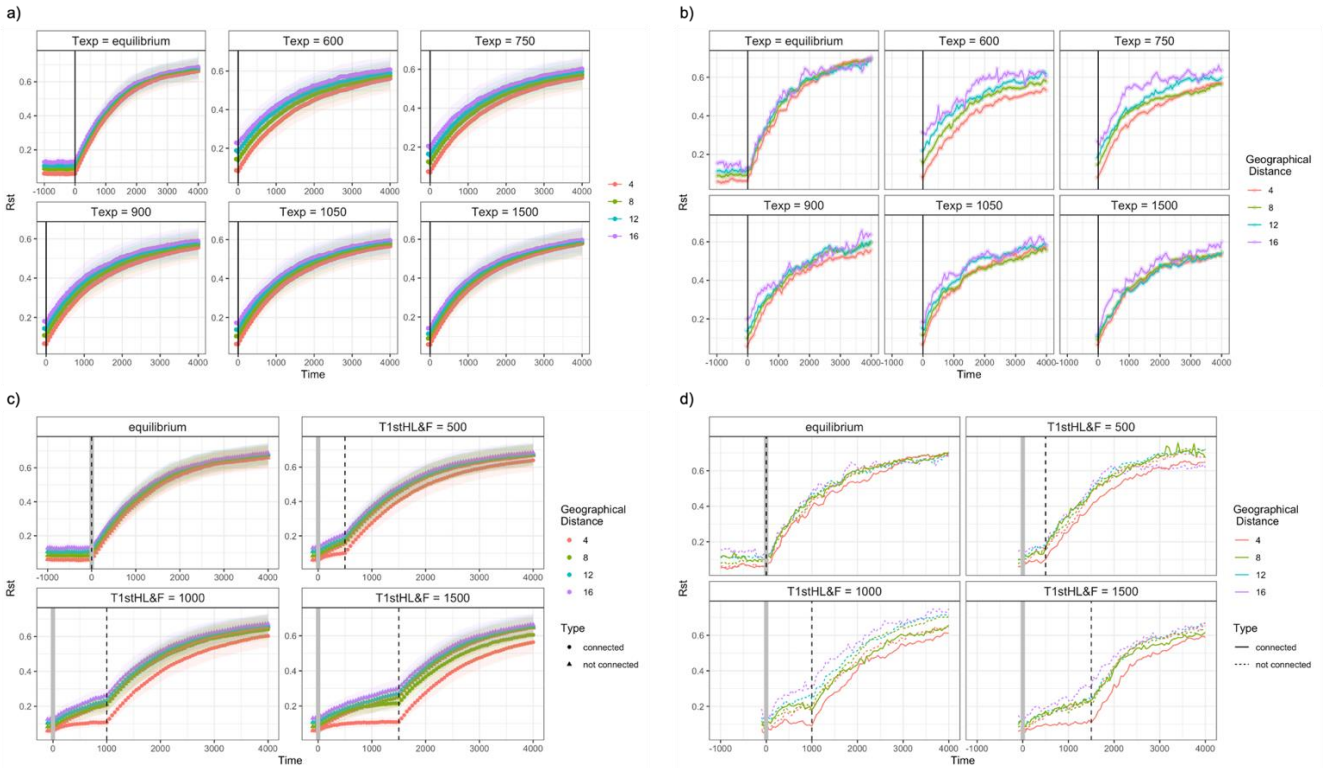


Figure S11. The effect of ancient range expansion prior to HL&F and spatio-temporal variation in HL&F on pairwise R_{st}. First row: effect of the time since the onset of the onset of the range expansion (T_{exp}); Second row: effect of the time since the 1st HL&F event. In both scenarios, carrying capacity, dispersal rate and mutation rate were fixed at $K = 50$, $m = 0.04$ and μ=5 x 10^{-4}. Panels **a)** and **c)** represent the mean (dots) and standard deviation (shade) of pairwise R_{st} across simulation replicates. Panels **b)** and **d)** indicate pairwise R_{st} over time for a single simulation replicate.

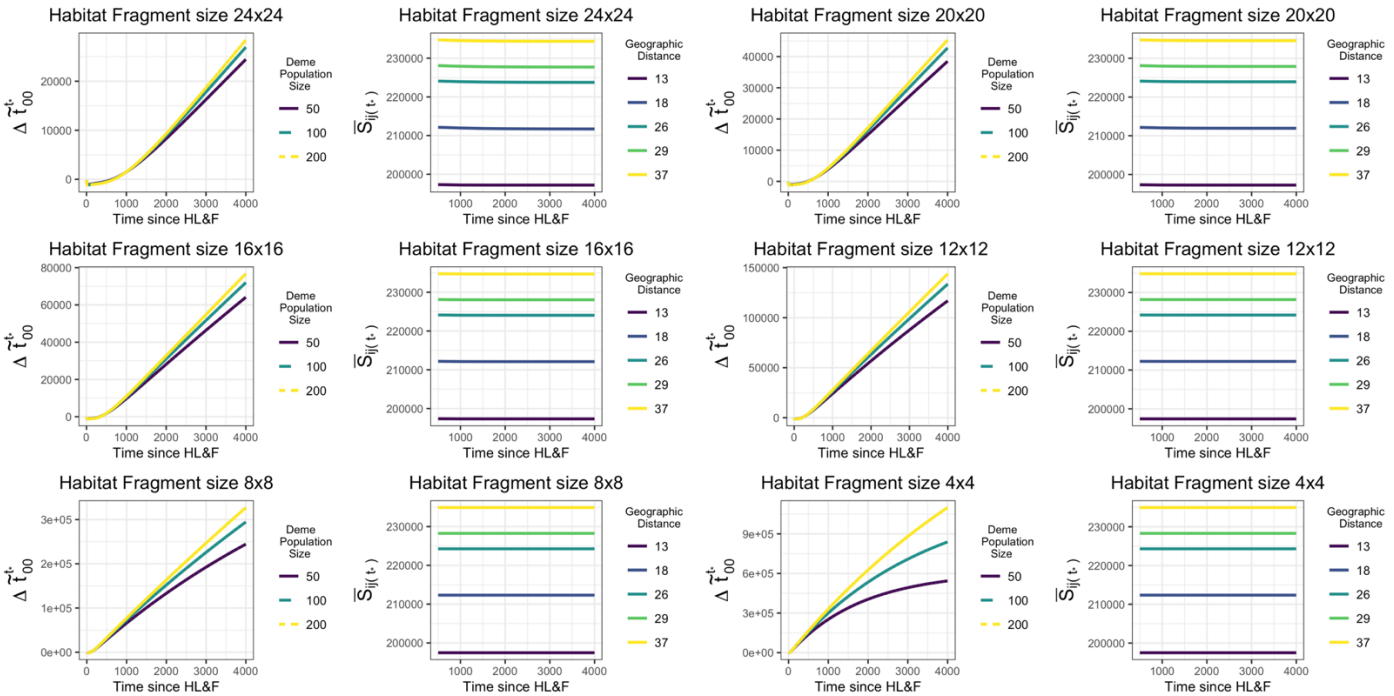


Figure S12. *Spatial* and *drift-isolation* term dynamics after HL&F as function of deme population size and habitat loss. An initial toroidal stepping stone model of size 78x78 is subject to HL&F, leading to the formation of nine habitat fragments of size 24x24, 20x20, 16x16, 12x12, 8x8 or 4x4. Dispersal rate was fixed at $m = 0.02$. The *drift-isolation* term is given by $\Delta\tilde{t}_0^{\,t_*} \equiv \bar{t}_{00} + t_* - \tilde{t}_0^{\,t_*}$ and the *spatial* term by $\bar{S}_{ij(t_*)}$. The results show that $\bar{S}_{ij(t_*)}$ remains nearly constant over time since HL&F and that $\Delta\tilde{t}_0^{\,t_*}$ changes slightly with increasing deme population size, expect for severe habitat loss (e.g., 4x4).

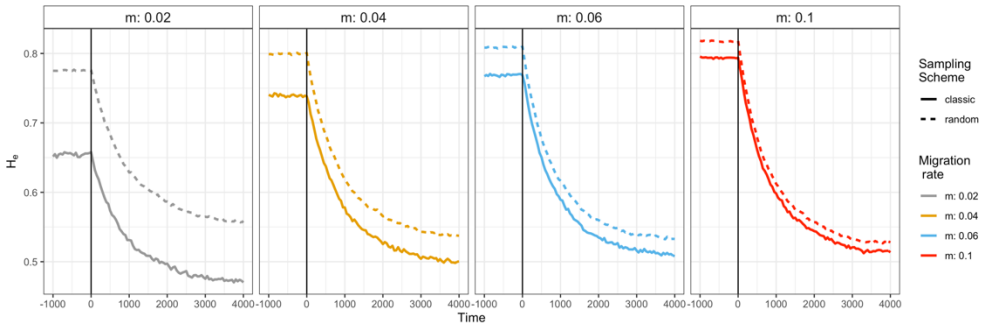


Figure S13. Temporal dynamic of expected heterozygosity calculated using *random* or *local* sampling within habitat fragment. Solid vertical lines indicate the HL&F event under the instantaneous HL&F scenario. Solid and dashed coloured lines refer, respectively, to *local* and *random* sampling. Across simulation, deme population size was fixed at $K = 50$. The results show that, at low dispersal rate, large differences in heterozygosity between sampling schemes can be identified, which tend to disappear with increasing dispersal rate.

Chapter 4

4 Landscape genomics of tufted-tailed rats (*Eliurus* spp.) in northern Madagascar: reconstructing past forest cover changes from forest-dwelling species

Gabriele Maria Sgarlata[1], Emmanuel Rasolondraibe[3], Barbara Le Pors[1], Jordi Salmona[1,2], Tantely Ralantoharijaona[3], Ando Rakotonanahary[3], Sophie Manzi[2], Amaya Iribar-Pelozuelo[2], John Rigobert Zaonarivelo[4], Nicole Volasoa Andriaholinirina[3], Solofonirina Rasoloharijaona[3], Lounès Chikhi[1, 2]

1. Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, 2780-156 Oeiras, Portugal;
2. Laboratoire Evolution & Diversité Biologique (EDB UMR 5174), CNRS, IRD, UPS, Université de Toulouse Midi-Pyrénées, Toulouse, France;
3. Département de Biologie Animale et Ecologie, Faculté des Sciences, Université de Mahajanga, Mahajanga, Madagascar
4. Département des Sciences de la Nature et de l'Environnement, Université d'Antsiranana, 201 Antsiranana, Madagascar

AUTHOR CONTRIBUTIONS: GMS, LC and JS designed the project. GMS collected part of the samples, performed DNA extraction for most samples and analyzed genomic data. BLP and ER led the fieldwork and collected most of the samples. JS collected part of the samples. TR, AR contributed in the collection of samples. SM and AI-P prepared the libraries for genomic sequencing. SR, NVA and JRZ contributed to the organization of the fieldwork. GMS wrote the manuscript and LC revised it.

Highlights

- Identified higher genetic diversity in tuft-tailed rat (genus *Eliurus*) individuals sampled in recently connected forest sites than in isolated forest fragments
- Identified evidence of recent expansion in four out of five investigated *Eliurus* lineages
- Genomic diversity within sites and pairwise differentiation are correlated among co-occurring humid-associated species, while differing from those of the dry-associated *E. carletoni* species
- HL&F and rivers influence genetic diversity similarly in the two *Eliurus* species with similar sampling distribution, the dry-associated *E. carletoni* and the humid-associated *E. myoxinus*
- Climatic changes through the last 25,000 years have likely influenced the distribution and diversity of dry and humid associated species in different ways.

4.1 Abstract

Natural and anthropogenic environmental changes have shaped the distribution and diversity of many taxa across the world. Historical records, such as palaeoecological data, have provided insights on how species have responded to such environmental changes (e.g., habitat loss and fragmentation, HL&F). Still, the palaeoecological approach has drawbacks that limit our understanding of fine-scale species responses (e.g., low sample size). Genetic data, with its own limitations, can complement paleoenvironmental reconstructions, by inferring, for instance, past changes in population connectivity and abundance across species. This is particularly relevant in biodiversity hotspots such as Madagascar, where species are increasingly threatened by ongoing HL&F. Here, we undertake a multi-species approach and investigate spatial population genomic patterns across five endemic rodents (*Eliurus* spp.) occurring in northern Madagascar: *E. carletoni*, *E. myoxinus*, *E. ellermani*, *E. minor* and *Eliurus* sp. nova. The *Eliurus* spp. are forest-dwelling species associated mainly to either dry or humid forests, suggesting that historical changes in connectivity within these species can shed light on species response to forest loss and fragmentation, as well as past changes in dry and humid forest cover across Madagascar. We show that i)) HL&F and rivers influence genetic diversity similarly in the two *Eliurus* species with similar sampling distribution, the dry-associated *E. carletoni* and the humid-associated *E. myoxinus*, ii) genomic patterns are consistent among co-occurring humid-associated species, while differing from those of dry-associated; and iii) that climatic changes through the last 25,000 years have likely influenced the distribution and diversity of dry and humid associated species in different ways.

4.2 Introduction

Natural and anthropogenic habitat loss and fragmentation have shaped the configuration of present-day landscapes, consequently influencing species distribution and diversity (e.g., Hewitt, 2000; Sandel et al., 2011; Weigelt et al., 2016). During the climatic oscillations of the Late Quaternary (between 0.5 – 1 million years ago), habitats have contracted and expanded, allowing species to increase or reduce their distribution over time scales ranging from thousands to hundreds of years (Graham et al., 1996; Williams et al., 1998; Lyons, 2005, 2003). For example, different mammal groups are thought to have moved either south or north in Eurasia, as late Quaternary climate was cooling or warming, respectively (Graham et al., 1996). Over the last century, anthropogenic activities have modified most natural environments, shaping their configuration and vegetation composition, with overall negative consequences for biodiversity and most ecosystems (Lewis and Maslin, 2015; IPBES, 2021). Palaeoecological data have shed light on environmental changes that have occurred over wide time scales, from hundreds to over millions of years ago (Mayewski et al., 2004; Gibbard and Lewin, 2016; Seppä, 2018). Remote-sensing data, on the other hand, have been successfully used to quantify landscape structure and changes over the last 150 years, since the first aerial photograph taken from a hot air balloon by Gaspard-Félix Tournachon in 1858 (Roerink et al., 2003; Xie et al., 2008; Quéméré et al., 2012; Hansen et al., 2013; Knauer et al., 2014; Mitchell et al., 2017; Singh, 2021).

For instance, fossil and sub-fossil records can show if a given species was present at a given location before and after major environmental shifts (Jungers et al., 1995). Still, paleoenvironmental analyses tell us relatively little on fine-scale species response to environmental changes, such as changes in connectivity or abundance, between or within populations (Hu et al., 2009).

Genetic data represent an additional source of information that, with their own limitations, can complement paleoenvironmental reconstruction and contribute to our understanding of how species respond to habitat loss and fragmentation (Hewitt, 2000; Hu et al., 2009; Mitchell and Rawlence, 2021; Epps and Keyghobadi, 2015). This is achieved by interpreting intraspecific genetic diversity using models describing the processes of gene inheritance, that is selection, genetic drift, gene flow, mutation and recombination (Charlesworth and Charlesworth, 2017; Ellegren and Galtier, 2016). In the field of population genetics, models have been developed to reconstruct the history of species, often assuming that DNA sequences are neutral, that is they have negligible impact on the fitness or persistence of a given species (Nielsen and Beaumont, 2009; Loog, 2021). Under the 'neutral' assumption, genetic data contain information on the dispersal and abundance of species, and thus can be used to reconstruct past changes in connectivity and population size (Kuhner et al., 1995; Beerli and Felsenstein, 1999; Nielsen, 2000; Hey and Nielsen, 2004; Beaumont, 2010; Mazet et al., 2016; Arredondo et al., 2021).

Since both genetic and paleoenvironmental approaches have their own advantages and limitations, linking the two sources of information can ultimately help draw a clearer picture of how paleoenvironmental changes have shaped the distribution and evolution of species (Hu et al., 2009). For instance, Teixeira et al., (2021) have integrated paleoenvironmental reconstructions obtained from lake core sediments with genomic-based demographic inferences of a forest-dwelling mouse lemur (*Microcebus arnholdi*) occurring in northern Madagascar. They showed that the climatic and forest cover changes that have occurred between 15,000 and 5,000 years ago have affected population size and connectivity in *M. arnholdi*.

Investigating population genetic patterns on a single species can provide useful insights on its response to environmental changes. Under the "niche conservatism" hypothesis, defined as the tendency of species to retain aspects of their fundamental niche over time (Wiens et al., 2010), it is often

tempting to use demographic inferences on one species as proxy for reconstructing the temporal dynamic of the niche (or habitat) where the considered species occurs. However, this procedure can be misleading, since we often do not actually know if the “niche conservatism” is a reasonable assumption but also because each species may be subject to idiosyncratic processes that influence species genetic diversity in ways that are not linked to past habitat changes. In these circumstances, a practical solution may be to carry out, whenever possible, comparative population genomic studies across several species. In fact, this can help identify the evolutionary processes, environmental events or landscape features that have played a major role in determining genetic diversity across species.

In the present study we investigate spatial population genetic patterns in five endemic rodent species (*Eliurus* spp.) occurring in northern Madagascar: *E. carletoni*, *E. myoxinus*, *E. ellermani*, *E. minor* and the new putative species which we call *Eliurus* sp. nova (Chapter 5). The *Eliurus* spp. genus is composed by forest-dwelling species that occur mainly in native forests and present very diverse ecological traits, such as scansorial, semi-fossorial or saltatorial behaviours (Carleton, 1994; Marquart, 2014; Jansa et al., 2019). In Chapter 5, we have seen that these five species co-exist in northern Madagascar, several of them often in co-occurrence or syntopy. In addition, these species are typically associated to two major vegetation types: *E. carletoni* and *E. myoxinus* mainly to dry deciduous forests; *E. ellermani* and *E. minor* to humid evergreen forests; *Eliurus* sp. nova as far it has only been found in the humid forest of Montagne d’Ambre. However, several studies have shown that such distinctions are not necessarily strict, and some *Eliurus* species may have more habitat flexibility than others (Goodman and Soarimalala, 2002; Jansa et al., 2019; see Chapter 5). This is the case for *E. carletoni* and *E. myoxinus*, which have been documented in both dry and humid habitats (Carleton et al., 2001; Goodman and Soarimalala, 2002; Carleton and Goodman, 2007; Shi et al., 2013; see Chapter 5). As forest-dwelling, ecologically diverse and co-occurring species, the five *Eliurus* spp.

studied in this work represent a good model i) to assess how ecologically diverse species respond to the same or similar landscape features; ii) identify major barrier to gene-flow across northern Madagascar; and iii) provide some insights on the humid and dry forest cover changes that occurred during the late Holocene.

In northern Madagascar, genetic studies on several mammals have provided useful and preliminary insights on putative past environmental changes. The study by Quéméré et al. (2012) used genetic data of *Propithecus tattersalli*, an endemic lemur occurring in a restricted area (~440 km^2) within northern Madagascar, and detected a bottleneck signal with a strong support for an ancient (and likely pre-human arrival) event of habitat degradation, possibly between 4,000 and 10,000 years ago. Later, the work of Salmona et al., (2017) on *P. tattersalli* and the neighbouring species *P. perrieri* found a major signal for a bottleneck and habitat loss and fragmentation that started at least 3000 years ago and possibly even before 10,000 years ago, depending on the generation time assumed, in agreement with Quéméré et al. (2012) findings. The recent genomic study of Teixeira et al., (2021) on *Microcebus arnholdi*, a small body-size lemur associated to the humid forests of northern Madagascar (e.g., Montagne d'Ambre), detected two signal of recent population declines at 1,000 and 5,000 years ago, in agreement with paleoenvironmental reconstruction suggesting an increase in drier conditions between 5,500 and 900 years ago.

Differences in species response to habitat loss and fragmentation were identified in the genetic studies of Rakotoarisoa et al. (2013a, b) on two *Eliurus* species associated to different habitat types: *E. carletoni* (dry deciduous forests) and *E. ellermani* (humid forests). Based on mtDNA sequences (*cytb* and *D-loop*), they found i) evidence for a recent population decline in *E. carletoni*; and ii) no signal of population size change and perhaps a slight continuous increase in *E. ellermani*. These results were interpreted as consequences of long-term habitat changes during the Holocene, which led to an expansion of the distribution of *E. ellermani* and a reduction and

fragmentation of the distribution of *E*. *carletoni*. Similar conclusions, but probably on a different time scale, were drawn by the study of Sgarlata et al., (2019), where the authors compared the distribution of two mouse lemur species from northern Madagascar, known to occur in dry (*Microcebus tavaratra*) and humid (*Microcebus arnholdi*) forests. Their findings showed that northern Madagascar was probably covered by humid vegetation in the recent past, and that recent Holocene environmental changes have increased the distribution of the dry-associated *M. tavaratra*, while reducing the distribution of the humid-associated *M. arnholdi*, in agreement with other types of data (Jungers et al., 1995; Simons et al., 1995; Godfrey et al., 1999).

The aim of this work is to compare population genomic patterns across five forest-dwelling rodents associated to different habitat type (humid *versus* dry) for reconstructing recent changes in landscape configuration and forest cover in northern Madagascar. Overall, our results show that i) HL&F and rivers influence genetic diversity similarly in the two *Eliurus* species with similar sampling distribution, the dry-associated *E. carletoni* and the humid-associated *E. myoxinus*, ii) genomic patterns are consistent among co-occurring humid-associated species, while differing from those of dry-associated species; and iii) that climatic changes through the last 25,000 years have likely influenced the distribution and diversity of dry and humid associated species in a different way.

4.3 Materials and Methods

4.3.1 Study area

We surveyed 35 forest sites between 2010 and 2018 in northern Madagascar, a region considered to be a transition zone between the humid ecoregion of eastern and the dry ecoregion of western Madagascar (Fig. 1). The landscape of northern Madagascar is characterized by forest fragments

of diverse habitat types (dry deciduous, transition, humid evergreen, mountainous) separated by a matrix of savannah and grassland. Although, some forests of northern Madagascar have remained nearly unchanged over the last 50 years (Quéméré et al., 2012), many and probably most have suffered from habitat loss and fragmentation (HL&F) in the last decades (Vieilledent et al., 2018). The main drivers of HL&F in recent times are human activities, including slash-and-burn subsistence farming, illegal logging, charcoal production (Jarosz, 1993; Campbell, 2013; McConnell and Kull, 2014; Zhu, 2017). The study region is crossed by six main rivers which further divide northern Madagascar in five inter-river systems: the Irodo, the Loky, the Manambato, the Manambery, the Fanambana and the Bemarivo rivers. We thus group the sampled sites of the present study in five sub-regions, each characterized by specific climate and vegetation types, and by different anthropogenic pressures on the forested habitat. A detailed description of each subregion is provided in the following subsections.

North Loky-Manambato region

In the North Loky-River region (NL), there are mainly fragmented dry forests, frequently surrounded by riparian corridors along narrow canyons, or more extended riparian forests in proximity of the Irodo river, as in the valleys of Andrafiamena. The forest sites we visited are within three major massifs: Ankarana, Analamerana, Andrafiamena-Andavakoeira and the humid forest massif of Montagne d'Ambre.

The Special Reserve of Ankarana (IUCN category IV) is a limestone massif of (~250 km^2) covered mostly by pinnacle karst ("tsingy") and dry forests. In addition, a large diversity of vegetation types can be observed, such as grazed savannah, scrubby secondary forest, moist primary seasonal forest and specialized xerophytic scrub on the dry limestone plateau (Rossi, 1974; Fowler et al., 1989). Surface water is rare and only present along canyons, where underground rivers get to the surface. The thick soil on the basalt formation of the Anakarana massif, the water locally provided by

underground rivers and the protection from drying winds create optimal conditions for the persistence of moist forests within canyons (Fowler et al., 1989; Nicoll and Legrand, 1989). There are also humid deciduous forests on basaltic soil in the north of the Ankarana Massif, extending down from Montagne d'Ambre (Goodman et al., 2021, pp. 558–572). Hawkins et al., (1990) found that, during the dry season, some animal taxa (e.g., lemurs) seek refuge in these moist areas (limited to 20 – 50 km^2), while using more extensively the dry deciduous forests during the wet season. Secondary grassland occurs at the edges of the massif. The Ankarana Special Reserve is currently managed by Madagascar National Parks (MNP).

The Speciale Reserve of Analamerana (~470 km^2) is characterised by the same vegetational types as Ankarana Special Reserve. It is drier than Ankarana forests and lies entirely on limestone. The dominant vegetation is dry deciduous forest. Moist semi-deciduous vegetation can be found near river streams, with tree species typical from rainforest habitats (Hawkins et al., 1990). Two major blocks of forests can be identified within the Analamerana Special Reserve, one in the northeast that is well preserved, and one in the southwest which has been heavily fragmented by fires (Goodman et al., 2021, pp. 515–527). The Analamerana Special Reserve is currently managed by MNP.

The Andrafiamena-Andavakoeira Protected Harmonious Landscape (IUCN category V; ~730 km^2) is composed of two large forests, Andrafiamena and Andavakoeira differing in geological substrate types: limestone karst and sandstone, respectively (Buřivalová, 2011). As a consequence, the vegetation occurring in the two forests is also different. Buřivalová, (2011) sampled large trees and found that only 4% of the species were shared between the two forests. In particular, Andrafiamena forest has an alkaline soil layer, is more water deprived and supports a more xeric and deciduous vegetation (Fowler et al., 1989) than Andavakoeira forest, which occurs on sandstone, where trees are less water deprived and often related to the tree

species occurring in the eastern rainforest. The Andrafiamena-Andavakoeira Protected area is managed by the Malagasy NGO Fanamby.

Montagne d'Ambre is an isolated volcanic massif with a peak of altitude at 1475 m. It is placed in a dry ecoregion, but due to its exposition to the trade winds, it receives more precipitation than neighbouring areas. Most of the massif is covered by medium-altitude moist evergreen forest, and only close to the summit the moist vegetation tends to transit to an ericoid vegetation (Nicoll and Legrand, 1989; Raxworthy and Nussbaum, 1994). Montagne d'Ambre is protected under the status of National Park (IUCN category II) and is managed by MNP (Goodman et al., 2021, pp. 500–514).

Loky-Manambato region

The Loky-Manambato (LM) region (~2,500 km^2) is a biogeographical transition zone between the western dry and the eastern humid phytogeographical ecoregions (Goodman and Wilmé, 2006), which is delimited by the Loky and Manambato rivers, in the north and south, respectively. This region is crossed by the relatively shallow Manankolana River, bordered by riparian forests along most of its course, and by a national dirt road. It is covered by ~360 km^2 of forest (Goodman et al., 2021, pp. 528–545), fragmented into a dozen major forest patches surrounded by human-altered grasslands, dry scrub and agricultural lands. Most forests are situated at low- to mid-elevations and mostly consist of dry deciduous vegetation. The two most southwestern massifs, Binara and Antsahabe are covered by moist evergreen forests above ~800 m of altitude and by moist semi-deciduous forests between 200 and 800 m of altitude (Gautier et al., 2006; Goodman et al., 2021, pp. 528–545). Despite the presence of human pressure, such as sustained grassland fires, slash-and-burn agriculture and charcoal production, as well as exploitation of wood, gold and sapphires (Goodman et al., 2021, pp. 528–545), deforestation rate in the LM region has been relatively low between 1950 and 2002 (Quéméré et al., 2012; Goodman et

al., 2021, pp. 528–545). The LM region is protected under the status of Protected Harmonious Landscape V and is managed by “Fanamby”.

South Manambato region

In this region, we group forests fragments located, from north to south, i) between the Manambato and the Manambery rivers; ii) between the Manambery and the Fanambana rivers; and iii) one forest south of the Fanambana river. Between the Manambato and the Manambery rivers, there are the lowland dry deciduous forest of Analafiana and the mountain humid forest of Salafaina, which topology and vegetation are similar to the Binara massif of the LM region. Between the Manambery and the Fanambana rivers, there is the evergreen humid forest of Bezavona-Ankirendrina. Ultimately, the southernmost visited forest of Analalava, south of the Fanambana river, is an evergreen humid forest.

COMATSA region

In the COMATSA region, we include sites that are located in the continuous forest of the Corridor of Marojejy-Anjanaharibe Sud-Tsaratanana Nord protected area (COMATSA Nord) and those lying in the small unprotected forest fragments at the north-east of the protected area, Ankinjanala and Anketrakabe. The vegetation composition of the COMATSA region is poorly studied. What is currently known is that it is mostly covered by continuous medium altitude evergreen forests, which in few locations of the north-central portion transits to an ericoid vegetation. The forests of the Andravory Massif are isolated from the continuous forest by a 3-4 km stretch of secondary grassland along the Fanambana river (Goodman et al., 2021, pp. 597–611). The Ankinjanala and Anketrakabe forests are disconnected from the north-east portion of the COMATSA protected area by secondary grassland and pastures on soils with traces of erosion, with only some riparian forests along river streams. The COMATSA protected area is classified as a

Natural Resources Reserve (VI IUCN category) and is currently managed by WWF.

North Marojejy

In the North Marojejy, we include three sites that are located in unprotected forests at the north of the Anakarangona river and Marojejy National Park: Lohananjialava, Beamalonkely, Andranomenabe, Ambatomainty, Andohanibemarivo. They are moist semi-deciduous forest fragments separated by secondary grassland. Deforestation in this area is most likely recent.

4.3.2 Study species

The tuft-tailed rats (*Eliurus* spp.) are nocturnal and easily recognizable by a noticeable tuft of long hairs over the distal portion of the tail. They are associated with forested habitats, but some species show tolerance to mild forest disturbance. In the present work, we focus on five *Eliurus* species distributed in northern and north-eastern Madagascar: *E. carletoni*, *E. ellermani*, *E. minor*, *E. myoxinus* and *Eliurus* sp. nova.

E. carletoni (Goodman et al., 2009) is a relatively large member of the *Eliurus* genus (head-body length: 143 – 150 mm; Goodman et al., 2009) known to occur in the extreme north of Madagascar. It inhabits the relatively mesic forest conditions between the pinnacle karsts (called in Malagasy *tsingy*) (Bésairie, 1964; Puy and Moat, 1996; Veress et al., 2008) found at the SR of Ankarana, SR of Analamerana and a nearby zone of metamorphic rock in the Loky-Manambato protected area (Rakotoarisoa et al., 2010; Rakotoarisoa et al., 2013b). *E. carletoni* can be considered a scansorial forest-dwelling species of dry deciduous habitats and it has been caught in both undisturbed and disturbed forests. Little is known on the ecology of this species. Documented records report an elevation range of occurrence between 37 m and 1,002 m (Jansa et al., 2019; see Chapter 5).

E. ellermani (Carleton, 1994) is another large member of the *Eliurus* genus (head-body length: 126 – 174 mm; Jansa et al., 2019) restricted to the north and north-eastern montane humid forests of Madagascar. In particular, individuals of *E. ellermani*, whose species identity has been genetically confirmed, have been collected from the northernmost Montagne d'Ambre to the central mountains of the Masoala Peninsula. The habitat of *E. ellermani* consists of evergreen humid forest in montane formations, characterised by moist or sclerophyllous plant communities. It occurs in an elevation range between 289 m and 1,558 m (Jansa et al., 2019; see Chapter 5).

E. minor (Major, 1896) is the smallest member of the *Eliurus* genus (head-body length: 101 – 124 mm; Goodman and Carleton, 1996). The litter size of *E. minor* ranges from 2 to 4 (Carleton, 2003). Trap capture data suggest that *E. minor* can be found on lianas, small trees, in bamboo grass, on tree ferns as well as on dead wood. It has been observed to climb along filigree vegetation and in dense filigree undergrowth (Marquart, 2014), up to heights of several meters (Goodman and Carleton, 1996; Carleton and Goodman, 1998). Observations by Katrin Marquart (Marquart, 2014) suggests that in comparison to other *Eliurus* species (*E. tanala* and *E. webbi*), *E. minor* can be 'considered both a good large-tree-climbers as well as a good 'small-branch and bamboo-grass-climbers'. *E. minor* occurs in a wide elevation range between 289 m and 2,016 m in correspondence of moist evergreen forests (near sea level to 1875 m; Garbutt, 1999; Carleton, 2003; see Chapter 5). Individuals whose species identity has been confirmed genetically show that *E. minor* has a wide distribution, extending from the south-east to the north of Madagascar (Everson et al., 2020; see Chapter 5).

E. myoxinus (Milne Edwards, 1885) (head-body length: 117 – 136 mm; Carleton et al., 2001) occurs in the dry forests of the west and in the humid forests of the far north Madagascar (Soarimalala and Goodman, 2003; see Chapter 5). It has been documented from localities in xerophytic scrub

(spiny bush), dry deciduous forest, dry-humid transitional forest (Carleton, 1994; Goodman and Ganzhorn, 1994; Goodman and Rasoloarison, 1997; Goodman et al., 1999). It is strictly forest-dependent and occurs at an elevation range between 40 m and 1,284 m (Soarimalala and Goodman, 2011; see Chapter 5).

Eliurus sp. nova has been described by Sgarlata et al., (see Chapter 5) using genetic data from individuals sampled at Montagne d'Ambre, in northern Madagascar. Current data suggest that this new taxon may occur only in Montagne d'Ambre, however, there are unpublished data suggesting that this might not be the case. In fact, a poster by (Tysdal and Jansa, 2014) show a phylogenetic tree, estimated using three nuclear and two mtDNA markers, where a still undescribed taxon is placed at the same evolutionary position as the *Eliurus* sp. nova lineage found in (see Chapter 5). Interestingly, the undescribed *Eliurus* taxon shown in (Tysdal and Jansa, 2014) corresponds to individuals sampled in northwest Madagascar (Jansa pers. comm.). No ecological data are available concerning this new taxon except that it occurs between 812 m and 1,057 m of altitude (see Chapter 5).

4.3.3 Sampling and Laboratory procedures

We collected small ear biopsies from individuals captured with Sherman traps (H.B. Sherman Traps®) in northern Madagascar (Fig. 1 and Table 1), during the dry seasons of 2011, 2012, 2013, 2014, 2015, 2017 and 2018 according to Rakotondravony and Radespiel, (2009) field procedures. We ensured DNA preservation in field conditions by storing the samples in Queen's lysis buffer (Seutin et al., 1991) and at −20 °C once in the laboratory. We extracted total genomic DNA from 354 ear biopsies using a DNeasy® blood and tissue kit (Qiagen®). To increase the amount of extracted genomic DNA, tissues were incubated overnight at 56 °C in 300 µL of ATL digestion buffer (Qiagen®) including proteinase *K* (C_f = 1 mg/ mL), and 20 µL of

dithiothreitol (1 M). The final DNA solution was then eluted in 60 µL of AE buffer. We quantified nucleic acids and checked for DNA purity using the Nanodrop spectrophotometer (ND-1000, Thermo Fisher) and quantified double stranded DNA (dsDNA) using a Qubit 2.0 Fluorometer (Invitrogen). For maximizing the amount of genomic data recoverable from high-throughput DNA sequencing, we selected samples with a dsDNA concentration superior to 4 ng/µL and a 260/280 absorbance ratio higher than 1.75.

4.3.4 RAD-sequencing

Genomic data were sequenced using the Restriction site-Associated DNA sequencing (RAD-seq) approach, which consist in sequencing regions neighbouring restriction sites, to obtain homologous sequences across many individuals, spread across the genome, with a high coverage and at a reasonable cost (Baird et al., 2008; Andrews et al., 2016). Library preparation and RAD sequencing were conducted at the GeT-PlaGe platform (Toulouse, France). The libraries were prepared at the GeTPlaGe platform (Toulouse, France) and sequenced using 150bp paired-end reads on an Illumina HiSeq3000. A detailed description on library preparation and sequencing can be found in Sgarlata et al., (see Chapter 5).

Raw reads were first checked for quality using FastQC (Andrews, 2010) and MultiQC (Ewels et al., 2016), then demultiplexed, adapter-removed and quality-filtered in STACKS v2.2 (Rochette et al., 2019) using default parameters. PCR duplicates were removed using the *clone_filter* program in STACKS v2.2 under default settings. The last filtering step was carried out using the *kmer_filter* program in STACKS v2.2 with default settings.

4.3.5 RAD-seq data processing and *de novo* assembly

Filtered short-read sequences were assembled *de-novo* in STACKS v2.2 using *ustacks*, *cstacks*, *sstacks*, *tsv2bam*, *gstacks* and *populations* programs. The assembly of nuclear loci from RAD-seq data relies on several parameters. The choice of the best combination of parameters is difficult, since several factors can influence the amount and quality of genomic information recoverable from the dataset, such as the biology and natural history of the studied population(s) or the deployed RAD-seq technique (Paris et al., 2017). Therefore, we followed the protocol proposed by (Paris et al., 2017) to optimize the three main parameters of the *de novo* assembly (*m*, *M*, *n*). The analysis for the optimization of *de novo* assembly parameters was performed on a subset of the whole dataset, as suggested by Rochette and Catchen, (2017). In particular, Rochette and Catchen, (2017) found that adding all samples does not increase significantly the number of widely shared *catalog* loci in the dataset and can add noise in the estimation of *m*, *M* and *n*. We selected 20 individuals for *E. carletoni*, 10 for *E. ellermani*, 10 for *E. minor*, 20 for *E. myoxinus* and 8 for *Eliurus sp.* nova (Tab. S1 and Fig. S1). We do not discuss the details of the *de novo* pipeline and parameter selection procedure inhere, since it has been thoroughly described in Sgarlata et al., (see Chapter 5).

4.3.6 RAD-seq datasets building

The *populations* program of STACKS 2.2 was used to generate the final RAD-seq dataset for each of the five species. Additional filtering parameters were specified in the *populations* program, allowing to further increase confidence on the recovered genomic information by reducing biases due to sequencing errors, paralogy or repetitive regions. A nucleotide site at a given locus was processed only if it was satisfying the following requirements: i) a minimum minor allele frequency (*min_maf*) equal to 5%; ii)

a maximum observed heterozygosity (*max_obs_het*) equal to 70%; iii) at least 80% of individuals within populations needs to be genotyped at that given locus in order to be processed; and iv) at least 80% of the populations, within species, need to be genotyped at that given locus in order to be processed, which corresponded to 14 sites for *E. myoxinus*, 17 sites for *E. carletoni*, 8 sites for *E. ellermani*, 6 sites for *E. minor* and 2 sites for *Eliurus* sp. nova). In order to reduce linkage disequilibrium (LD) among genotyped SNPs, only one random SNP per locus was kept in the final datasets. In addition, we further mitigated the effect of LD by thinning SNPs that had R^2 value greater than 0.1 with any other SNP within a 50-SNP sliding window (advanced by 10 SNPs each time). This analysis was performed with *plink-v1.90* (Purcell et al., 2007) using the *--indep-pairwise* command. At last, we removed loci that were missing in more than 10% of individuals, and excluded individuals with more than 40% missing genotypes.

4.3.7 Clustering analyses (ADMIXTURE, PCA

Clustering methods were used to investigate population genetic structure within each of the five *Eliurus* species. We used the maximum likelihood model-based approach implemented in ADMIXTURE v1.3 (Alexander et al., 2009). This program uses the same likelihood model as STRUCTURE software (Pritchard et al., 2000), but instead of sampling the posterior probability of a model given the data, it maximises the likelihood of the model. In the model, it is assumed that individuals are sampled from an admixed population with contribution from K *a priori* clusters. Then, individuals are assigned to K-clusters by maximizing the likelihood of the model. For each species, we ran the program for several K values and repeated the analysis for 30 independent runs. ADMIXTURE analysis was carried out using the default optimization method and the default termination criterion, which stops the analysis when the log-likelihood increases by less than 10^{-4} between iterations. K values were compared using the cross-validation method by setting $--cv = 20$ and by comparing likelihoods

across runs. We used CLUMPAK (Kopelman et al., 2015) to permute the 30 independent runs so that clusters align across runs. The final assignment of individuals to the best K-value was obtained by taking the mode containing most runs (*Major.Cluster*).

In addition, we used a model-free approach, the Principal Components Analysis (PCA), through the function *glPca* implemented in the R package ADEGENET 2.0.1 (Jombart and Ahmed, 2011). PCA constructs low-dimensional projections of the data that explain the gross variation in marker genotypes, and thus it can be useful to assess major pattern of genomic differentiation within species.

4.3.8 Genetic Diversity and Differentiation

We assessed the population- and individual-based observed and expected heterozygosity (H_{obs} and H_{exp}) for each *Eliurus* species. We used the Variant Call Format (VCF) file generated in STACKS v2.2 and estimated heterozygosity in R using the function *gl.report.heterozygosity* from the *dartR* R package (Gruber et al., 2018). Individual H_{obs} is measured as the proportion of heterozygous loci for each individual, also taking into account missing values. Population H_{obs} is obtained by averaging Individual H_{obs} over the individuals in a given population. Population H_{exp} is the expected proportion of heterozygotes under Hardy-Weinberg equilibrium for each locus, which is then averaged across loci. The calculations of expected heterozygosity use the unbiassed estimates of Nei, (1987). Computing heterozygosity from a VCF file, which include esclusively nucleotide sites that are variants within the dataset, overestimate the true level of genetic diversity. Thus, the purpose of this analysis is not to provide absolute measures of genetic diversity, but rather compare genetic diversity among forest fragments within species. The same reasoning applies concerning comparisons of genetic diversity among species.

We used pairwise F_{st} to measure pairwise genetic differentiation between forest fragments within species. We used the *gl.fst.pop* function implemented in the *dartR* R package (Gruber et al., 2018), which uses the method of Wright, (1949), updated by Weir and Cockerham, (1984). In addition, we estimated pairwise genetic distances between individuals (d_{xy}) using the individual genetic distance measure of Smouse and Peakall, (1999). We investigated patterns of isolation by distance (IBD) to assess the impact of geographically-limited dispersal on *Eliurus* spp. genomic structure (Wright, 1943). We computed IBD by correlating either pairwise $d_{xy}/(1-d_{xy})$ or pairwise $F_{st}/(1-F_{st})$ with the logarithm of Euclidean geographic distance, and assessed the significance of such correlation using Spearman correlation test.

4.3.9 Effective migration surfaces

Within each species, we estimated spatial variation in effective dispersal rate with the Estimating Effective Migration Surfaces (EEMS) software (Petkova et al., 2016). EEMS identify changes in effective dispersal rate in a landscape due, for instance, to the presence of a physical barrier or to environmental condition that are unsuitable for the dispersal and survival of a species. EEMS does not require prior knowledge on the landscape features that could influence gene-flow, but as showed by Tang et al., (2020), it is more suitable for investigating large-scale rather than small-scale changes in dispersal. For a correct interpretation of the results, it is important to point out that EEMS, although providing useful insights on dispersal, makes impossible i) to separately infer population size and dispersal rate; and ii) to disentangle historical and current dispersal, as the inferred effective dispersal rates are an historical average across time.

EEMS is based on a two-dimensional stepping-stone model and estimate effective migration rates between adjacent subpopulations (demes).

Following Petkova et al., (2018), we used the genetic dissimilarity matrix generated by the EEMS program *bed2diffs_v2* and raw geographic coordinates for each sampling sites as input files for EEMS. We ran EEMS with different grids, differing in number of demes, to investigates how and if results would change with grid size. We tested several numbers of demes, namely 200, 400 and 600 demes. We checked whether results would change across deme size and if not, choose the larger number of demes in order to retain more fine scale information of effective dispersal rate. Since EEMS appears to perform well at large spatial scale, we used EEMS for each species to estimate changes in effective dispersal across the entire geographic range of sampled individuals. We performed the analysis eight times for each species (eight chains), using MCMC length of 8×10^6 iterations with a burn-in of 10^6 iterations and set 999 as number of iterations to thin between two writing steps. To ensure MCMC convergence and reliable parameters estimation, we checked for convergence in the posterior trace plot generated in *rEEMSplot* package and made sure that the parameters values of the models would be accepted about 20% − 30% of the time, or less than 10% (Petkova et al., 2018). In conclusion, we combined results from different chains using the *rEEMSplot* package and plotted surfaces of effective migration rates.

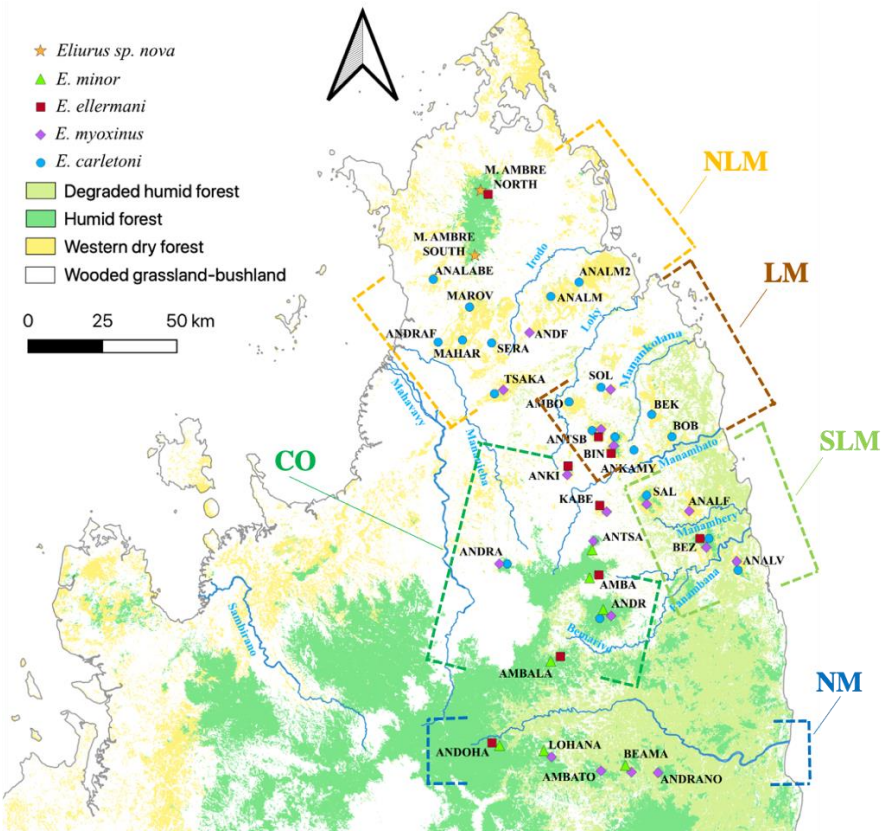


Figure 1. Geographic distribution of sampled *Eliurus* individuals across northern Madagascar. *E. carletoni* (N = 142), *E. myoxinus* (N = 116), *E. ellermani* (N = 26), *E. minor* (N = 38) and *Eliurus* sp. nova (N = 13). NLM: North Loky-Manambato; LM: Loky-Manambato; SLM: Southern Loky-Manambato; CO: COMATSA region; NM: North Marojejy.

4.4 Results

4.4.1 RAD-data processing: assembly of nuclear loci *de novo*

The optimization analysis of *de novo* assembly parameters shows that the highest amount of polymorphism for $r = 80\%$ was at: *m* = 2, *M* = 5, *n* = 6 for *E. carletoni*; *m* = 2, *M* = 6, *n* = 6 for *E. myoxinus; m* = 2, *M* = 2, *n* = 4 for *E. ellermani*; *m* = 2, *M* = 2, *n* = 4 for *E. minor*; *m* = 2, *M* = 2, *n* = 4 for *Eliurus sp.* nova (Fig. S2). The cleaned *catalog* was composed of 232,788 loci for *E. carletoni*, 227,531 loci for *E. myoxinus*, 220,305 loci for *E. ellermani*, 249,551 loci for *E. minor* and 203,426 loci for *Eliurus sp.* nova. After the three filtering steps described in *Material and Methods* (*process_radtags*, *clon_filter*, *kmer_filter*) and data processing (*ustacks*, *cstacks*, *sstacks*, *tsv2bam*, *gstacks*) of each RAD-seq data set, we obtained ~488 million (M) short-reads for *E. carletoni* (N_{ind} = 142), ~516 M short-reads for *E. myoxinus* (N_{ind} = 116), ~110 M short-reads for *E. ellermani* (N_{ind} = 26), ~122 M short-reads for *E. minor* (N_{ind} = 38) and ~60 M short-reads for *Eliurus sp*. nova (N_{ind} = 13). The number of reads per individual (N_{reads}) ranged between ~950 K and ~9.1 M for *E. carletoni* (average N_{reads} = ~3.4 M), ~1.5 M and ~11 M for *E. myoxinus* (average N_{reads} = ~4.4 M), ~2 M and ~7.5 M for *E. ellermani* (average N_{reads} = ~4.2 M), ~1 M and ~6 M for *E. minor* (average N_{reads} = ~3.2 M) and, ~2.3 M and ~8.5 M for *Eliurus sp*. nova (average N_{reads} = ~4.6 M) (Tab. S2 and Fig. S3). The weighted mean individual coverage ($\overline{W}$) was between ~5 and ~35X for *E. carletoni* (average $\overline{W}$ = ~14X), ~6 and ~40X for *E. myoxinus* (average $\overline{W}$ = ~17X), ~9 and ~26X for *E. ellermani* (average $\overline{W}$ = ~17X), ~6 and ~22X for *E. minor* (averge $\overline{W}$ = ~14X) and, ~8 and ~29X for *Eliurus sp*. nova (average $\overline{W}$ = ~16X) (Tab. S2). The final filtered datasets comprised 9,924 SNPs for *E. carletoni*, 21,556 SNPs for *E. myoxinus*, 2,668 SNPs for *E. ellermani*, 5,605 SNPs for *E. minor* and 61,617

SNPs for *Eliurus sp.* nova. A visual representation of the SNPs data for each species, show little missing data across individuals (Fig. S4).

Table 1. Localities and summary statistics of nuclear genomic data across *Eliurus* species sampled in northern Madagascar. Region: Abbreviation of pre-defined regions; Forest: complete forest name; Site: forest abbreviation; Forest type: major vegetation type(s) within site; N: number of samples used for genomic analyses; Coverage: mean coverage; WCoverage: weighted mean coverage calculated by giving more weight to loci that are more represented across the dataset; Nloci: number of genotyped loci; NFWReads: number of used forward reads; SNPs number: average number of SNPs within site; Hobs – mean and -sd: average and standard deviation of within-site individual observed heterozygosity averaged; a) *E. carletoni*; b) *E. myoxinus*; c) *E. ellermani*; d) *E. minor*; e) *Eliurus* sp. nova.

a)

Region	Forest	Site	Forest type	Latitude	Longitude	Altitude (m)	N	Coverage	WCoverage	Nloci	NFWReads	SNPs number	Hobs - mean	Hobs - sd
LM	Ambohitsitondroina	AMBO	D	-13.13039	49.4511	132-165	7	10.9305	11.336	124827	1371577	8502	0.202	0.005
NL	Analabe	ANALABE	D/H	-12.76553	49.02506	119-138	2	16.2805	17.0855	126316	2085721.5	8775	0.185	0.009
NL	Analamerana	ANALM	D	-12.81209	49.38934	215-491	22	12.588	13.174	124409.5	1575327	3071	0.210	0.006
NL	Ampasimaty	ANALM_2	D	-12.76715	49.47568	77-144	9	12.321	12.972	123517	1479403	6086	0.210	0.007
SM	Analalava	ANALV	H	-13.61952	50.00191	98	1	5.065	5.114	96117	486790	6740	0.167	-
CO	Andoharano	ANDR	H	-13.77516	49.52557	1002	1	13.689	14.137	123591	1691852	9341	0.197	0.021
CO	Andraharikely	ANDRA	H	-13.6246	49.24428	856	1	16.555	17.334	131829	2182490	9572	0.204	-
NL	Andrafiabe	ANDRAF	D	-12.95576	49.04309	37-138	12	12.276	12.857	119299	1371452.5	4545	0.210	0.007
LM	Ankaramy	ANKAMY	?	-13.27271	49.65465	439, 439	2	14.411	15.3375	132716.5	2048095	8741	0.175	0.019
LM	Antsahabe	ANTSB	D/T/H	-13.20946	49.56042	571-737	6	10.5415	10.921	125683.5	1319830	7386	0.205	0.007
LM	Bekaraoka	BEK	D	-13.16378	49.70763	121-270	16	12.0415	12.5175	126873	1512211.5	4900	0.210	0.007
SM	Bezavona - Ankirendrina	BEZ	H	-13.55742	49.89869	409	1	13.771	14.337	127616	1757407	9340	0.198	-
LM	Binara	BIN	D/T/H	-13.25624	49.61579	287-507	3	11.947	12.479	122649	1456062	8794	0.195	0.005
LM	Bobankora	BOB	D	-13.24654	49.76853	174-391	11	13.419	14.079	126443	1696730	6386	0.206	0.005
NL	Mahamasina/Ankarana	MAHAR	D	-12.94873	49.11851	108-127	11	12.43	12.865	123484	1562426	4401	0.209	0.008
NL	Marovato/Ankarana	MAROV	D	-12.84803	49.13856	150-335	14	13.692	14.323	124866	1697643	4192	0.211	0.006
SM	Salafaina	SAL	H	-13.44336	49.72103	336-380	6	14.8975	15.626	130793	1971236.5	7318	0.197	0.007
NL	Antserasera/Masiamposa	SERA	D	-12.95584	49.20876	233-337	8	15.4285	16.093	128986.5	1988415	6352	0.213	0.008
LM	Solaniampilana	SOL	D	-13.0925	49.58031	110-210	8	14.2365	14.8935	139144.5	1989476	8464	0.210	0.007
NL	Andavakoera	TSAKA	D/H	-13.1176	49.21146	271	1	10.839	11.33	119764	1298170	9346	0.214	-
	Total	-	-	-	-	37-1002	142	13.0035	13.6265	125274.8	1633589.5	7352	0.206	0.007

b)

Forest	Site	Forest type	Latitude	Longitude	Altitude (m)	N	Coverage	WCoverage	Nloci	NFWReads	SNPs numbe	Hobs - mean	Hobs - sd
Ambatomainty	AMBATO	H	-14.25017	49.61373	857-917	2	17.976	18.86	141311	2553914.5	20671	0.210802574	0.003044483
Analafiana	ANALF	D	-13.4542	49.82948	40-64	3	26.119	27.79	147455	3851412	20576	0.273279549	0.004705203
Analalava	ANALV	H	-13.61609	49.991	180-194	2	11.9145	12.227	127646.5	1523044	19542	0.24749258	0.002134853
Andrafiamena	ANDF	D	-12.9226	49.32506	421-521	3	16.945	17.659	134006	2270792	20300	0.260180624	0.006686515
Andoharano	ANDR	H	-13.74913	49.59581	937	1	18.215	19.368	145232	2645416	21361	0.251065025	-
Andraharikely	ANDRA	H	-13.62457	49.24422	783-990	8	11.7165	12.098	131233.5	1555386	14984	0.279311111	0.006826093
Andranomenabe	ANDRANO	H	-14.24828	49.74902	620-840	2	25.892	27.027	136040.5	3517481.5	20188	0.222879929	0.001225913
Ankinjanala	ANKI	H	-13.35139	49.44964	801-1084	12	13.7645	14.234	131484	1799665	14192	0.294267501	0.004314077
Antsahavary	ANTSA	H	-13.5416	49.54489	976-1043	7	15.128	15.764	131507	1940523	18428	0.267372486	0.003234179
Antsahabe	ANTSB	H	-13.21014	49.56031	571-737	7	17.205	17.8925	132874	2288981	18574	0.298777	0.004011957
Beamalonkely	BEAMA	H	-14.24036	49.65551	737-1032	3	18.479	19.186	135569	2505135	16257	0.205121896	0.006618565
Bezavona - Ankirendrina	BEZ	H	-13.55385	49.89297	223-579	12	12.521	12.949	131306	1660709.5	14402	0.270063208	0.016128507
Binara	BIN	H	-13.27597	49.61046	289-913	18	17.7155	18.4935	136223	2435099.5	12893	0.297934214	0.009110544
Anketrakabe	KABE	H	-13.46114	49.5738	705-826	11	12.35	12.704	133628	1645868	16261	0.278426576	0.007022852
Lohananjialava	LOHANA	H	-14.18806	49.39257	953	1	19.003	19.759	136225	2588641	20954	0.194234991	-
Salafaina	SAL	D	-13.44309	49.71178	336-669	11	13.617	14.12	132691	1822420	15022	0.29142224	0.004492191
Solaniampilana	SOL	D	-13.09044	49.57916	110	1	13.857	14.276	131809	1826501	20738	0.28667181	-
Andavakoera	TSAKA	D/H	-13.11661	49.22099	151-474	12	14.0305	14.4895	133020	1860078.5	11429	0.278474631	0.006538214
Total	-		-	-	40-1084	116	16.0365	16.7115	133324	2105657.5	18501	0.271671378	0.004705203

c)

Region	Forest	Site	Forest type	Latitude	Longitude	Altitude (m)	N	Coverage	WCoverage	Nloci	NFWReads	SNPs numbe	Hobs - mean	Hobs - sd
CO	Ambaliha	AMBA	H	-13.65099	49.55507	988-1018	5	9.011	9.43	93753	844773	3103	0.18825599	0.00437173
CO	Ambalasoa	AMBALA	H	-13.91712	49.40907	918-1148	6	12.077	12.9485	119141	1549723.5	3898	0.19858199	0.01740371
NM	Andohanibemarivo	ANDOHA	H	-14.1517	49.24149	938-1344	7	13.88	14.577	113740	1625191	3435	0.18313507	0.01521156
CO	Andoharano	ANDR	H	-13.75711	49.56868	930-1298	9	14.419	15.579	125540	1999289	3886	0.20461006	0.01639861
CO	Antsahavary	ANTSA	H	-13.55831	49.53822	962-1049	7	11.595	12.282	107856	1250540	3515	0.18875349	0.01823541
NM	Beamalonkely	BEAMA	H	-14.22791	49.64715	1032	1	11.609	12.091	114319	1327160	4911	0.14274079	-
NM	Lohananjialava	LOHANA	H	-14.18806	49.39257	942-1127	3	13.606	14.375	115838	1576102	5468	0.18343087	0.0099027
	Total	-	-	-	-	918-1344	38	12.077	12.9485	114319	1549723.5	3886	0.18825599	0.01580508

d)

Region	Forest	Site	Forest type	Latitude	Longitude	Altitude (m)	N	Coverage	WCoverage	Nloci	NFWReads	SNPs number	Hobs - mean	Hobs - sd
CO	Ambaliha	AMBA	H	-13.65282	49.55326	1018	1	14.604	15.31	113191	1653024	2458	0.328722539	-
CO	Ambalasoa	AMBALA	H	-13.89701	49.44081	918	1	13.019	13.766	134933	1756733	2586	0.367362722	-
NM	Andohanibemarivo	ANDOHA	H	-14.1517	49.24149	926	1	11.815	12.224	92867	1097216	2130	0.258215962	-
CO	Ankinjanala	ANKI	H	-13.34506	49.44872	933-1084	2	12.6295	13.322	132482.5	1719220.5	2531	0.351837218	0.010337002
LM	Antsahabe	ANTSB	H	-13.21235	49.55932	571-737	9	16.519	17.295	122916	2050124	1746	0.38289464	0.014508251
SM	Bezavona - Ankirendrina	BEZ	H	-13.55604	49.87817	409-605	2	15.417	15.9585	116971.5	1806571.5	2515	0.327833002	0.015463568
LM	Binara	BIN	H	-13.28065	49.60196	767-913	6	17.0225	17.9835	123855	2262536.5	2145	0.363293851	0.018471171
CO	Anketrakabe	KABE	H	-13.45813	49.56894	812-818	2	19.3405	20.6385	138659.5	2715244	2599	0.350711812	0.024214122
NM	Montagne d'Ambre	M_AMBRE_NORTH	H	-12.52451	49.18241	938-1118	2	18.399	19.6855	149407	2745101.5	2657	0.359804291	0.007983893
	Total	-	-	-	-	409-1118	26	15.417	15.9585	123855	1806571.5	2515	0.351837218	0.014983909

e)

Forest	Site	Forest type	Latitude	Longitude	Altitude (m)	N	Coverage	WCoverage	Nloci	NFWReads	NPs numbe	Hobs	Hexp
Montagne d'Ambre	M_AMBRE_NORTH	H	-12.52451	49.18241	923-1050	4	10.1775	10.4155	138241.5	1299438	61617	0.24854748	0.01212686
Montagne d'Ambre	M_AMBRE_SOUTH	H	-12.69158	49.15266	812-1057	9	15.405	15.924	151723	2051905	43540	0.26727825	0.01048124
Total	-	-	-	-	812-1057	13	12.79125	13.16975	144982.25	1675671.5	52578.5	0.25791286	0.01130405

4.4.2 Genomic diversity within *Eliurus* species

Across the five *Eliurus* species, population genomic diversity was computed by averaging the estimated individual observed heterozygosity (H_{obs}) for each sampling site (Table 1, Fig. S10).

In *E. carletoni*, H_{obs} ranged between 0.167 (ANALV; N = 1) and 0.214 (TSAKA; N = 1). The lowest values were detected for the ANALV, ANKAMY, ANALABE and ANDR forest fragments, whereas the highest were found at the north of the Loky river (NL region) and in few forest fragments of the LM region. We computed correlation between site H_{obs} and geographical distance from a set of candidate sites, to identify putative origins of expansion (Fig. 2). The strongest correlation was found for ANALM (Spearman ρ = -0.69; p-value = 0.001), corresponding to a steady decrease in H_{obs} moving away from the site. Similar results were obtained by performing the same analysis among individuals rather than sites (Fig. S11).

In *E. myoxinus*, H_{obs} varied between 0.194 (LOHANA; N = 1) and 0.299 (ANSTB; N = 7). Among the lowest values of H_{obs}, we found all sites in the North Marojejy region (NM), at the southern-edge of our *E. myoxinus* sampling distribution, followed by ANALV and ANDR (as in *E. carletoni*). The highest levels of genetic diversity were observed in the two neighbouring (and recently connected) forest fragments of the LM region, ANTSB and BIN. We identified a significant decrease in H_{obs} with increasing geographical distance from ANTSB (Spearman ρ = -0.86; p-value = 0.0) (Fig. 2 and Fig. S11).

In *E. ellermani*, the minimum H_{obs} was 0.258 (ANDOHA; N = 1) and the highest was 0.383 (ANTSB; N = 9), the latter very similar to the neighbour BIN forest. Consistently and similarly to *E. myoxinus*, we found a linear decrease in H_{obs} with increasing geographical distance from ANTSB (Spearman ρ = -1; p-value = 4e-4) (Fig. 2 and Fig. S11). The same analysis performed at individual level showed statistical significance (Spearman ρ = -0.71; p-value = 0.0; Fig. S11).

In *E. minor*, H_{obs} ranged between 0.143 (BEAMA; N = 1) and 0.205 (ANDR; N = 9). As for *E. myoxinus*, the lowest genetic diversity was detected in the three sites in the NM region (BEAMA, ANDOHA and LOHANA) whereas the highest values were in ANDR and AMBALA sites (CO region). A non-significant linear decrease in site H_{obs} was identified with increasing distance from ANDR site (Spearman ρ = -0.75; p-value = 0.07) (Fig. 2). However, the same analysis performed at individual level showed statistical significance (Spearman ρ = -0.5; p-value = 0.0; Fig. S11).

For *Eliurus sp.* nova, the number of sampled sites is limited to two, but we found that the northern site of the Montagne d'Ambre had lower H_{obs} than the southern site, respectively 0.249 and 0.267.

The comparison of population genomic diversity (H_{obs}) across *Eliurus* species was performed for those sampling sites in which more than two *Eliurus* species were co-occurring (Fig. S12). We found that among the five possible pairwise comparisons, which include three among humid-habitat associated species (*E. minor - E. ellermani; E. myoxinus - E. ellermani; E. myoxinus - E. minor*) and two among humid- and dry-habitat associated species (*E. carletoni - E. ellermani; E. carletoni - E. myoxinus*), only *E. myoxinus - E. ellermani* comparison showed significant Spearman correlation value (Spearman ρ = 1.0; p-value = 0.016).

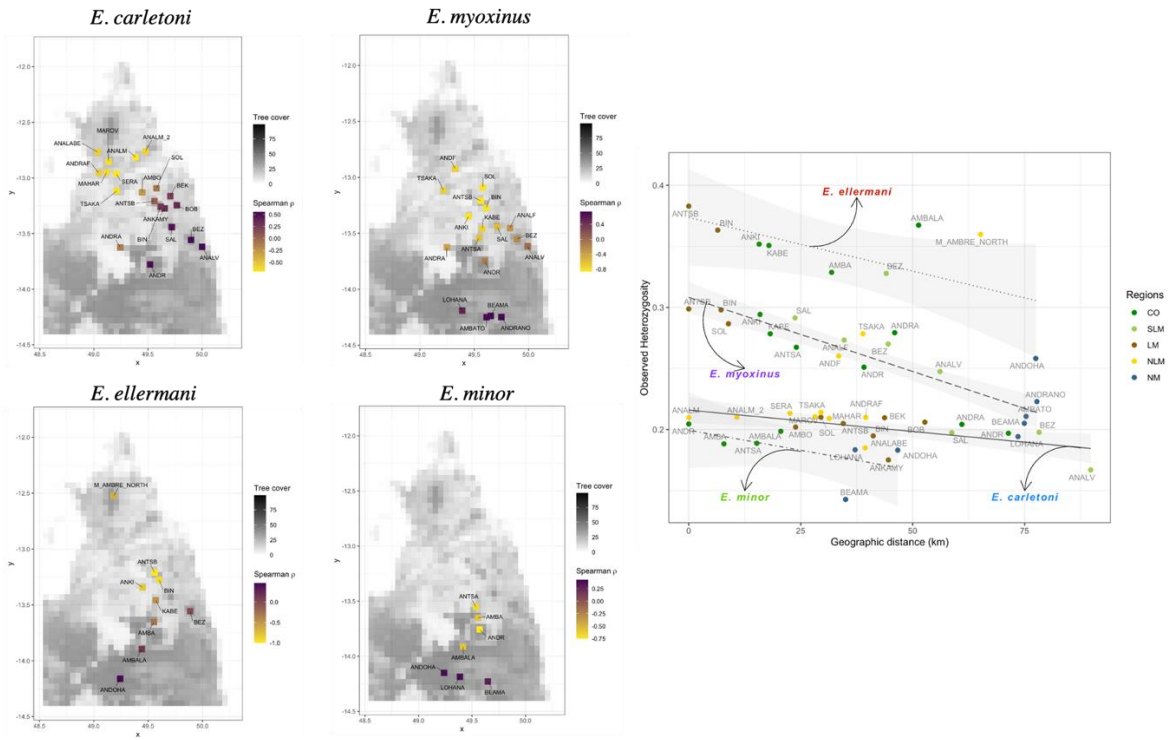


Figure 2. Decline in observed heterozygosity and putative expansion origin in *E. carletoni*, *E. myoxinus*, *E. ellermani* and *E. minor*. Map: squared cells indicate sampling sites, with colour code proportional to the amount of correlation (Spearman ρ) between observed heterozygosity and geographical distance between the focal cell and the remaining sampling sites. Plot: Linear regression between geographic distance and observed heterozygosity from the site with highest correlation, namely ANALM for *E. carletoni*, ANTSB for *E. myoxinus* and *E. ellermani* and ANDR for *E. minor*.

4.4.3 Population Genomic structure within *Eliurus* species

Population genomic structure was investigated within each *Eliurus* species using Principal Component Analysis (PCA) (Fig. 3), ADMIXTURE analysis (Fig. 4), isolation-by-distance analysis (Fig. 5) and by estimating effective migration surfaces (Fig. 6).

4.4.3.1 Principal Component Analysis

In *E. carletoni*, the first principal component (PC1 = 2.6%) divides sampling sites in the north (NL region) and south (LM, SLM and CO regions) of the Loky river. The second principal component (PC2 = 1.7%) shows i) a major separation of the LM region sampling sites at the west (BIN, ANTSB, AMBO, SOL) and east (BEK and BOB) of the Manankolana river; ii) that sampling sites south of the Manambato river cluster with BEK and BOB forest fragments; and iii) that ANKAMY sampling site, which is placed near the head source of the Manankolana river, is equally close to the sites on the east and on the west of the Manankolana river.

In *E. myoxinus*, PC1 (3.2%) shows three clusters, one composed of individuals from the neighbouring sampling sites ANTSA and KABE, a second including individuals from the southern edge of *E. myoxinus* sampling distribution (ANDRANO, AMBATO, LOHANA and BEAMA) and a third cluster including all other sites. PC2 (3%) separates the sites at the southern edge from all other sites. The PCA analysis, thus, identify three main genetic clusters: i) ANTSA and KABE; ii) ANDRANO, AMBATO, LOHANA and BEAMA; and iii) a third cluster, which include all other sampling sites, where PCA coordinates mirrors the geography of the sampled individuals. Within the third cluster, two major gap can be identified i) between individuals sampled at the north of Loky river (TSAKA and ANDF) and within LM (BIN, ANTSB, SOL) and CO region (ANDRA, ANKI), and ii) between individuals north (NLM, LM, CO regions) and south of the Manambato river (SLM region).

In *E. ellermani*, PC1 (6.1%) separates three individuals from ANTSB from the rest of the sampled individuals, whereas PC2 (5.5%) shows a separation between individuals from the neighbouring ANTSB and BIN forests, and the rest of the *E. ellermani* sampling distribution.

In *E. minor*, PC1 (5.7%) separate sites of the nearly-continuous forest of the CO region from the individuals sampled in the isolated ANDR forest fragment. PC2 (5.2%) distinguishes roughly sites from the north and south of CO region, with the exception of ANDR site.

In *Eliurus sp.* nova, PC1 (12.2%) splits individuals from the north and the south of the Montagne d'Ambre, while PC2 (9.2%) distinguishes individuals from the southern site, whose distribution is wider than those from northern site, suggesting that genomic data mirror the geography of *Eliurus sp.* nova individuals.

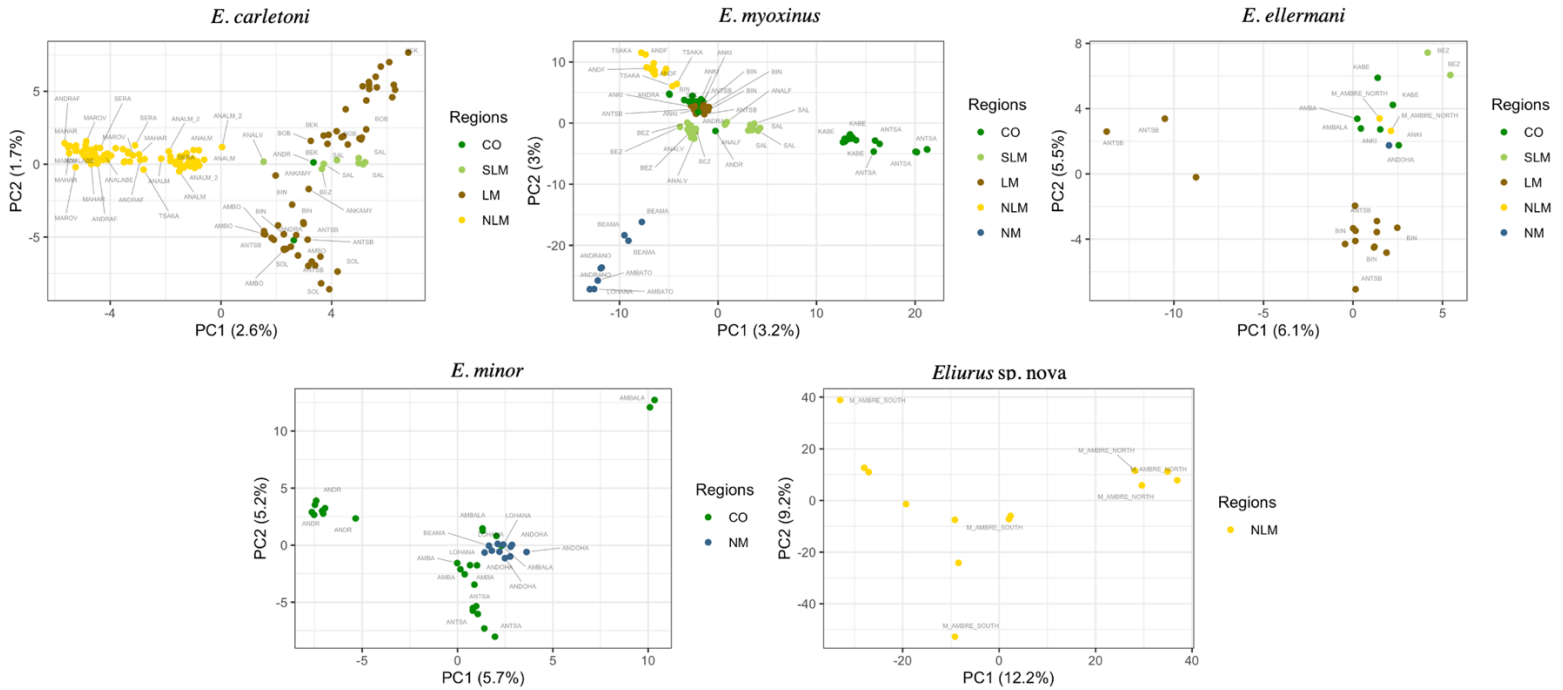


Figure 3. Principal Component Analysis (PCA) of within species Eliurus sp. genomic diversity. PC1 and PC2 were plotted for each Eliurus species. Colours refer to the regions from which individuals were sampled. NLM: North Loky-Manambato; LM: Loky-Manambato; SLM: Southern Loky-Manambato; CO: COMATSA; NM: North Marojejy.

4.4.3.2 ADMIXTURE analysis

In *E. carletoni*, at K = 2 a gradual change in proportions of individual genetic ancestry is observed from the northern (MAHAR) to the southern sampling sites (ANALV), in agreement with their geographical location (Spearman ρ = 0.94, p-value = 0.0; Fig. S13). Both Fig. 4 and Fig. S13 show a major change in genetic ancestry between sites south and north of the Loky river. At K = 3, sites from the Loky-Manambato region (ANTSB, BIN, AMBO and SOL) form a third cluster (Cluster 3). Sites at the south of the Manambato river and at the north of the Loky river show proportions of co-ancestry with Cluster 3, whereas BEK and BOB, which are both at the east of the Manankolana river, form together another unique genetic group (Cluster 2). At K = 4, a fourth genetic cluster includes sites located in the Analamerana massif. At K = 5, BEK forest fragment forms a new genetic cluster, separated from the neighbour BOB population, which instead show strong similarity with populations south of the Manambato River (ANDR, ANDRA, ANALV, BEZ and SAL). We also performed ADMIXTURE analyses within each region, whenever the number and location of the sampled individuals were sufficiently large (LM and NL region; Fig. S14). For LM region, at K = 2, a major genetic separation is observed for individuals in the east and west of the Manankolana river. At K=3, BOB forms a unique cluster, separated from BEK forest fragment. At K = 4 and K = 5, SOL and AMBO sites form distinct genetic clusters, Cluster 4 and Cluster 5 respectively, leaving the two other neighbour forest fragments (BIN and ANTSB) grouping into Cluster 2. For NL region, we found biologically reasonable results only at K=2, which identify a major genetic gap between sites in the Analamerana (ANALM and ANALM_2) and Ankarana-Anadavakoera massifs.

In *E. myoxinus*, at K = 2, we identified major genetic differences between KABE+ANTSA and the remaining sampling sites, among which TSAKA and ANDF were assigned with nearly 100% of genetic ancestry to

Cluster 1 (Fig. 4). At K = 3, the southern NM populations are identified as a unique genetic cluster (Cluster 3). For higher K values, *i*) TSAKA and ANDF form Cluster 4; *ii*) BIN, ANTSB and SOL form Cluster 1; iii) BEZ individuals show 100% genetic ancestry with Cluster 5; and iv) ANALF and SAL show similar pattern of genetic ancestry. Changes in % genetic ancestry correlate with geographical distance from the northern ANDF site (Spearman ρ = 0.70, p-value = 0.001; Fig. S13). As in *E. carletoni*, we also performed ADMIXTURE analysis within two regions (CO and SM) (Fig. S14). For CO region, ADMIXTURE results further confirm the high proportion of genetic co-ancestry between KABE and ANTSA sites. Within SM region, at K = 2, ANALF and SAL sites are assigned to Cluster 1, BEZ individuals to Cluster 2, whereas ANALV show mixed ancestry between the two clusters. For higher K values, each forest fragment is assigned to a unique genetic cluster.

In *E. ellermani*, at K = 2, three individuals from ANTSB forest fragment cluster apart from the other individuals, as observed in the PCA analysis (Fig. 3). At K = 3, the two individuals genotyped in BEZ forest form a distinct genetic cluster (Cluster 3). In addition, at K = 4, similar patterns of genetic co-ancestry are inferred for two groups of sampling sites: i) ANTSB and BIN forests from LM region; and ii) the other remaining sites from NM, CO and Montagne d'Ambre sites. We found a significant correlation between % ancestry and geographical distance from ANKI, further suggesting that geography shapes *E. ellermani* genomic diversity (Fig. S13).

In *E. minor*, ADMIXTURE analysis for K = 2 separates individuals occurring in the ANDR forest fragment from the other locations in CO and NM regions. Note that individuals from AMBA, ANTSA and AMBALA sites show higher genetic co-ancestry with sites geographically far apart (~ 50 km; BEHAMA, LOHANA and ANDOHA) than with the geographically close ANDR population (~ 15 - 25 km). At K = 3, two additional clusters group together AMBA and ANTSA sites on one side, and BEHAMA, LOHANA and ANDOHA,

on the other, in agreement with their relative geographic proximity. At K = 3 and K = 4, AMBALA individuals show higher genetic co-ancestry with sites in the NM region rather than CO region. Also in *E. minor*, % genetic ancestry correlated with geographical distance from ANTSA (Spearman ρ = 0.96, p-value = 8e-4) (Fig. S13). ADMIXTURE analysis within *E. minor* regions (CO and NM) confirms the results obtained from the global analysis (Fig. S14 and Fig. 4).

In *Eliurus sp.* nova, individual ancestry estimation distinguishes individuals sampled in the north and south of Montagne d'Ambre (Fig. 4).

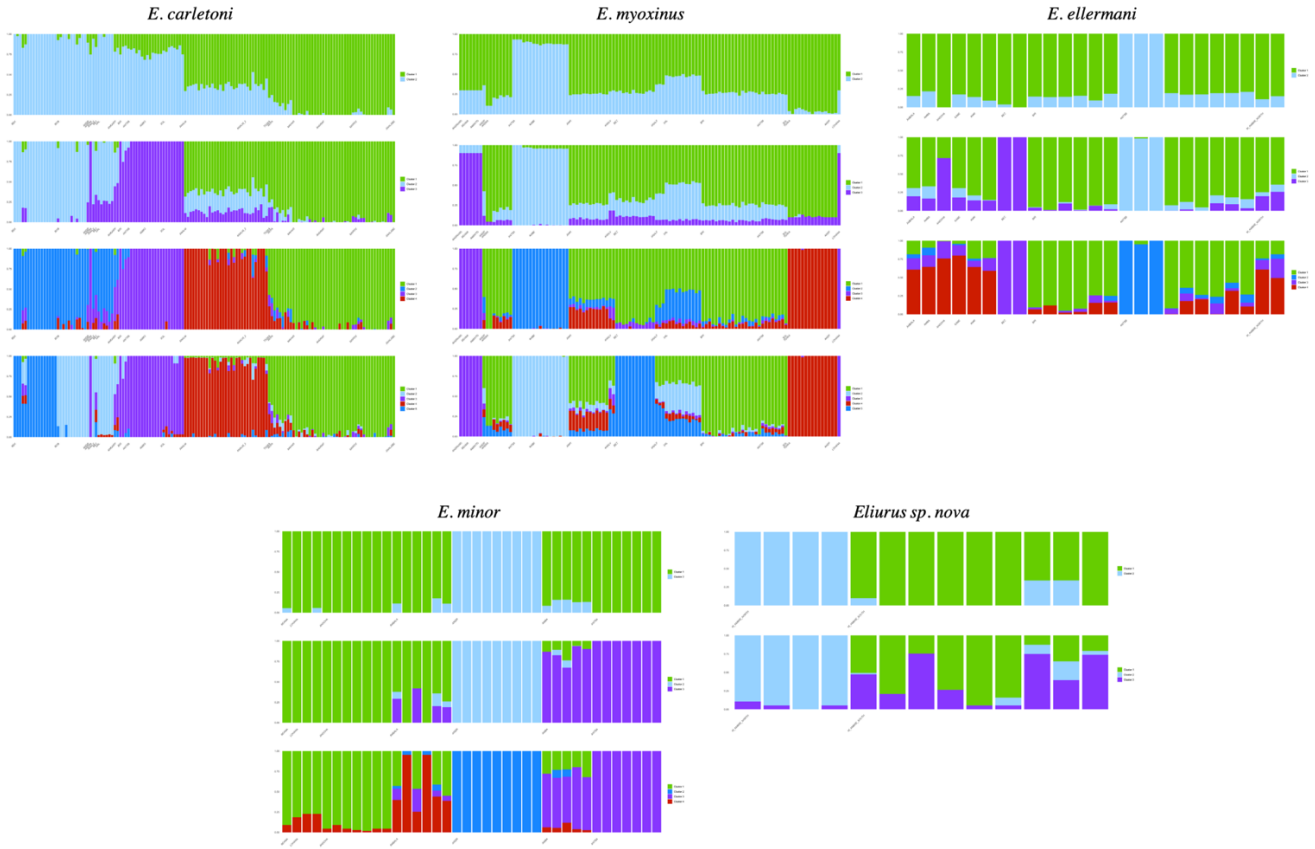


Figure 4. ADMIXTURE analysis within *Eliurus* species. For each species, several values of a prior number of clusters (K) were tested: K = 2 -5 for *E. carletoni*; K = 2 -5 for *E. myoxinus*; K = 2 – 4 for *E. ellermani*; K = 2 – 4 for *E. minor*; and K = 2, 3 for *Eliurus* sp. nova.

4.4.3.3 Isolation-by-distance (IBD)

Within each *Eliurus* species, we investigated isolation-by-distance pattern (IBD) by regressing pairwise genetic differentiation with the logarithm of geographic distance, using as sampling unit either individuals (Fig. 5) or populations (Fig. S15). In addition, we also looked at IBD patterns within each region, aiming at comparing landscape and spatial genetic structure among regions. It was not always possible to compare population-based IBD and genetic differentiation among regions, due to the limited number of surveyed sites within some of the regions. Overall, our results show that all five *Eliurus* species exhibit spatial genetic structure across the whole species sampling distribution (Fig. 5): *E. carletoni*, slope = 0.0096 (p-value <2e-16); *E. myoxinus*, slope = 0.022 (p-value <2e-16); *E. minor*, slope = 0.013 (p-value <2e-16); *E. ellermani*, slope = 0.011, (p-value <2e-16); *Eliurus* sp. nova, slope = 0.020 (p-value = 4.7e-10).

In *E. carletoni*, we detected significant individual-based IBD within LM (slope = 0.013, p-value<2e-16), NLM (slope = 0.006, p-value<2e-16) and SLM (slope = 0.011, p-value=0.01) regions. We found significant differences in IBD slope between LM and NLM regions (T-test = 14.282, p-value<0.0001) and between SLM and NLM regions (T-test = 2.898, p-value= 0.011). No significant difference was detected between LM and SLM regions (T-test = -1.166, p-value= 0.474). Controlling for geographic distance, we also compared within-region individual genetic distances among LM, NLM and SLM regions and found significant differences among all pairwise comparisons ($\Delta_{LM-SLM} = 0.049$ [0.039, 0.059], p-value = 0; $\Delta_{NLM-SLM} = 0.033$ [0.023, 0.043], p-value = 0; $\Delta_{LM-NLM} = 0.016$ [0.014, 0.018], p-value = 0). Non-significant differences in population-based IBD slope were found between LM (slope = 0.023, p-value = 2.63e-05) and NLM (slope = 0.012, p-value = 5.93e-06) regions (T-test = 2.006, p-value = 0.055). Average pairwise F_{st} was higher in LM than NLM sampling sites ($\Delta\bar{F}_{st}$ = 0.045; p-value<0.0001). Average genetic differentiation between LM and NLM sites was lower than

pairwise comparisons within LM region, but higher than comparisons within NLM region (Fig. S14). Any other inter-region pairwise population comparison shows values of genetic differentiation that increase with geographic distance in a similar way as within LM region.

In *E. myoxinus*, we found significant IBD within LM (slope = 0.006, p-value= 1.72e-14), CO (slope = 0.023, p-value < 2e-16), NLM (slope = 0.017, p-value = 1.33e-13), SLM (slope = 0.023, p-value < 2e-16), NM (slope = 0.006, p-value = 0.036) regions. We detected lower individual-based IBD slope among sampling sites occurring at the NM region (AMBATO, BEAMA, ANDRANAO) in comparison to SLM (T-test = 0.017; p-value= 0.0003) and CO (T-test = 4.351; p-value=0.0001) regions (Fig. 5). Non-significant differences were identified between IBD slope in the NM region and LM (ANTSB, BIN and SOL; T-test = - 0.070, p-value = 1.0) or NLM (ANDF and TSAKA; T-test = 2.47, p-value =0.098) regions.

In *E. ellermani*, we compared IBD slopes computed within CO (ANKI, KABE, AMBA and AMBALA) and within LM (ANTSB and BIN) regions and found no significant differences between the two regions ($\Delta slope$ = 0.010, p-value= 0.352).

In *E. minor*, we found lower IBD slope within NM region (slope = 0.007, p-value = 8.6e-05) than CO region (slope = 0.016, p-value < 2e-16), with a significant difference of $\Delta slope = 0.009$ (T-test= 2.97, p-value = 0.009). Both regions present similar average individual genetic distances (Δ_{NM-CO} = -0.012 [-0.025, 0.0001]; p-value = 0.052). Similarly, pairwise F_{st} within NM and CO regions show small non-significant differences ($\Delta\bar{F}_{st}$ = 0.012, p-value= 0.601), although such statistical comparison should be interpreted with caution given the small sample size of F_{st} values in NM region (only 3 sites).

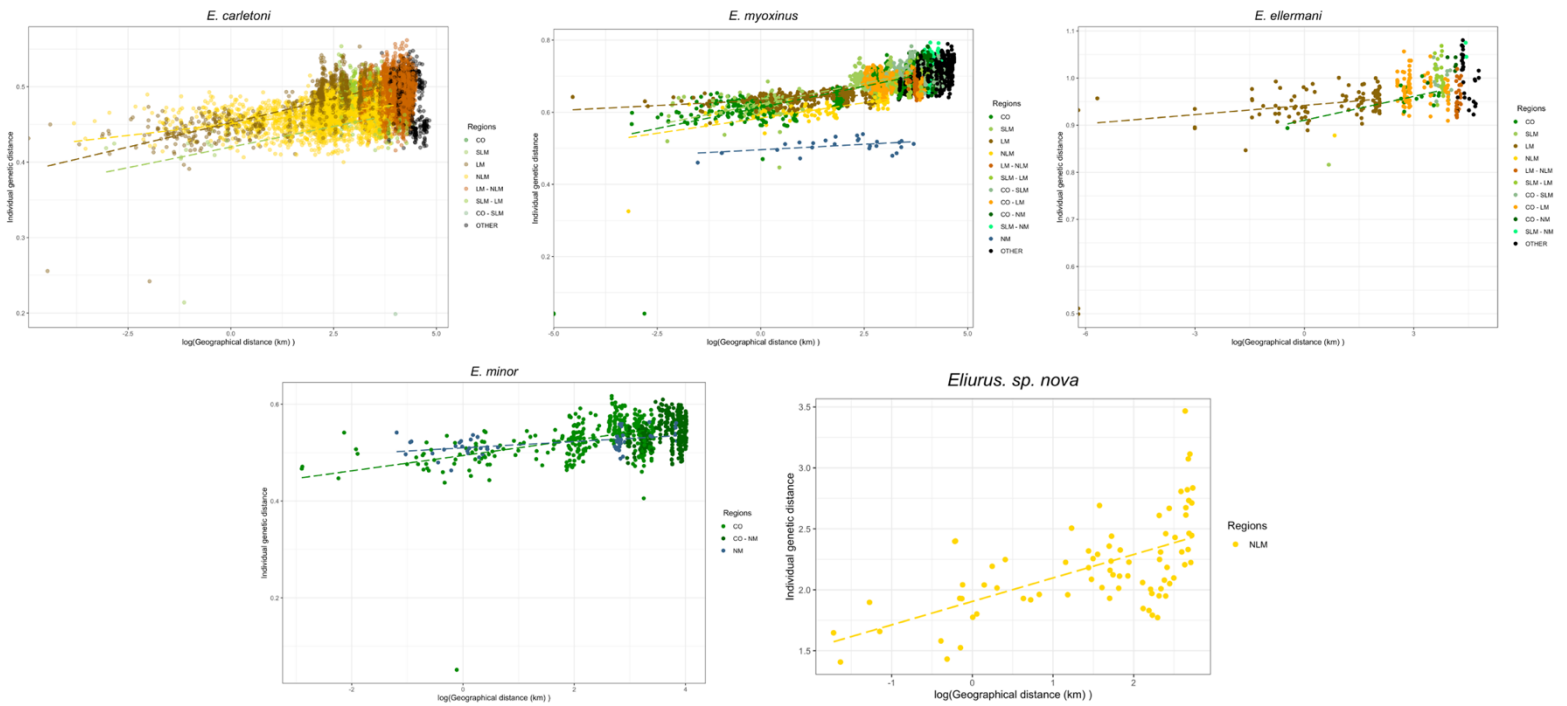


Figure 5. Isolation-by-distance across *Eliurus* species sampled in northern Madagascar. Isolation-by-distance patterns were computed by regressing pairwise genetic distance d_{xy} between individual x and y with the logarithm of geographical distance. Colours identify within or between-regions pairwise comparisons. CO: COMATSA region; SLM: Southern Loky-Manambato; LM: Loky-Manambato; NLM: North Loky-Manambato.

4.4.3.4 Genomic differentiation within Eliurus species

As for *Eliurus* spp. genomic diversity, we compared population pairwise genetic differentiation between *Eliurus* species pairs. The analysis was constrained to populations pairs where at least two *Eliurus* species were co-occurring, and where there were at least two individuals genotyped (Fig. 6). The analysis included two pairwise comparisons between humid-associated *Eliurus* species (*E. myoxinus* - *E. ellermani*; *E. myoxinus* – *E. minor*) and two between humid- and dry-associated *Eliurus* species (*E. carletoni* – *E. ellermani*; *E. carletoni* – *E. myoxinus*). Among all pairwise comparisons, *E. myoxinus* - *E. ellermani* showed significant correlation (Spearman ρ = 0.84, p-value < 0.01), whereas no significant relationships were found for all the other species pairs (*E. myoxinus* – *E. minor*: Spearman ρ = 0.14, p-value = 0.78; *E. carletoni* – *E. ellermani*: Spearman ρ = 0.9; p-value = 0.08; *E. carletoni* – *E. myoxinus*: Spearman ρ = 0.18, p-value = 0.26). Given the limited number of sampling sites within regions, it was possible to quantify correlation in population pairwise F_{st} between *Eliurus* species only for *E. myoxinus*-*E.ellermani* comparison which, in agreement with individual-based genetic distances, show significant correlation between the two species (Spearman ρ = 1, p-value=0.016) (Fig. S15).

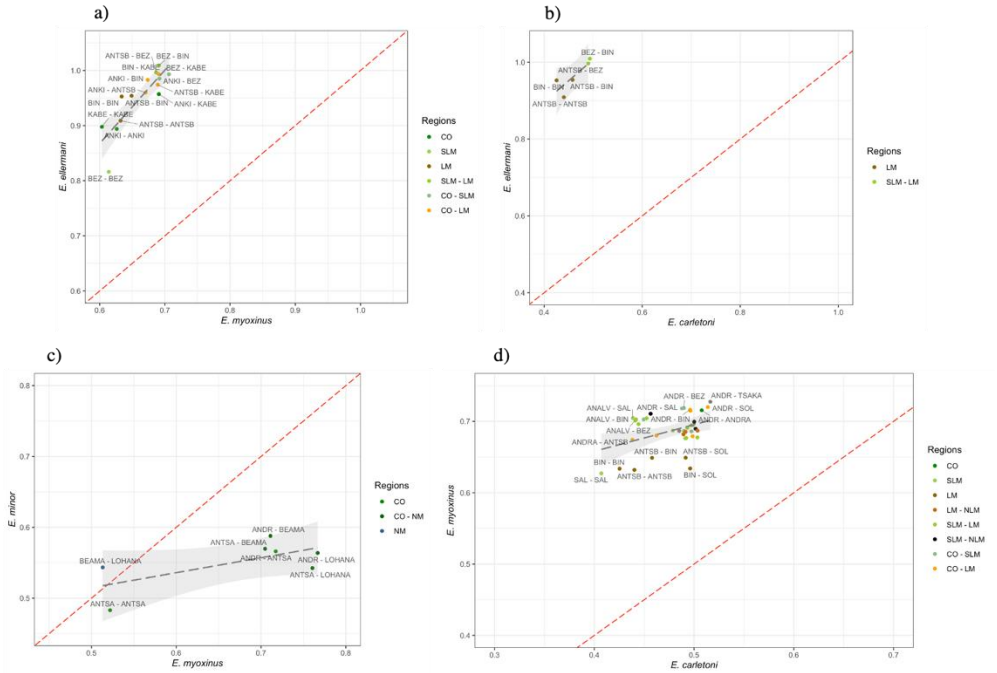


Figure 6. Comparison of population pairwise genetic differentiation between co-occurring *Eliurus* species pairs. a) *E. myoxinus* - *E. ellermani*: Spearman ρ = 0.84, p-value < 0.01; b) *E. carletoni* – *E. ellermani*: Spearman ρ = 0.9, p-value = 0.08; c) E. myoxinus – *E. minor*: Spearman ρ = 0.14, p-value = 0.78; d) *E. carletoni* – *E. myoxinus*: Spearman ρ = 0.18, p-value = 0.26. Colours in dots refer to within-region or between-regions comparisons. CO: COMATSA region; SLM: Southern Loky-Manambato; LM: Loky-Manambato; NLM: North Loky-Manambato.

4.4.3.5 Estimating Effective Migration Surfaces (EEMS)

The above genetic analyses have shown that *Eliurus* species are genetically structured in space, consistent with geographically-limited dispersal in these species. We performed EEMS analysis to investigate patterns that deviate from average IBD across the whole distribution of each species (Fig. 7).

In *E. carletoni*, the surveyed areas with higher effective migration are within the NLM region (ANDRAF, MAHAR, MAROV, SERAM, ANALM and ANALM2), where sites are connected by either average (white) or high (blue) migration rate. Within NLM region, TSAKA (in the south) and ANALABE (in

the north) are the only sites with migration rate lower than average (brown). Within LM region, migration rate ranges between average ($\log_{10}(m) = 0$) and slightly low ($\log_{10}(m) = -1$) values. Nearly average migration rate was estimated among AMBO, ANTSB, BIN, ANKARAMY and BOB sites, as also between BEK and BOB. Low migration rate was inferred between SOL and BEK sites. In the SLM region, average connectivity was identified between SAL and BEZ, whereas we found low migration between these two sites and the southern ANALV forest fragment. The CO region (ANDRA and ANDR) was characterized by low migration rate among sites. Overall, among regions we found evidence of low connectivity.

In *E. myoxinus*, high connectivity was identified within LM (in the north east) and NM (in the south) regions. In the LM region, we found highest migration between ANTSB and BIN, and average migration between these two sites and SOL. In the NM region, the three sites (LOHANA, AMBATO and ANDRANO) showed connectivity values ranging between $\log_{10}(m) = 0$ and $\log_{10}(m) = 2$. Low migration rate was estimated among forest fragments in the SLM region (SAL, ANALF, BEZ and ANALV). In the CO region, we found lower connectivity than average across most sites, however ANDR and ANDRA showed values closer to the average across the whole distribution.

In *E.ellermani*, surprisingly, high connectivity was detected between ANDOHA (in the NM region) and M_AMBRE_NORTH (in the NLM region). Average to high connectivity were inferred also between the two most geographically distant samples, AMBALA, AMBA in the CO regio and M_AMBRE_NORTH in the NLM region. Lower connectivity ($0 \geq \log_{10}(m) > -1$) was estimated among the other sites of CO region (AMBA, KABE and ANKI) and within LM region. Finally, the lowest migration rate ($\log_{10}(m) \leq -1$) was found between BEZ (in the SLM region) and all other sampling sites across *E.ellermani* sampling distribution.

In *E. minor*, average to high migration rate was estimated within NM region ($0 \leq \log_{10}(m) < 1$), whereas lower connectivity than average was

inferred within CO region ($0 > \log_{10}(m) > -1$). In particular, the lowest connectivity was found between ANDR forest fragment and the other sites within CO region.

For all *Eliurus* species, MCMC analyses were checked for convergence (Fig. S17). Moreover, we repeated the same analysis changing grid density within each species, and found that the number of demes assumed in the model did not have major effect on EEMS results (Fig. S18).

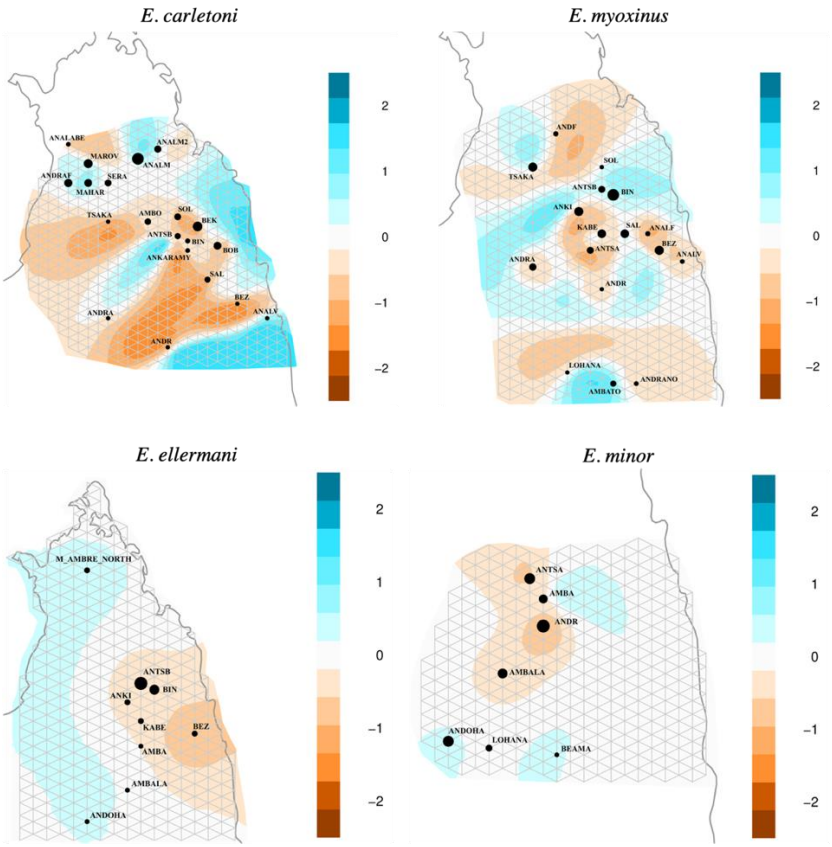


Figure 7. Estimated effective migration surfaces (EEMS) across *Eliurus* species. Coloured contour plots are interpolation of averaging draws from posterior distribution of the migration rates. Colours represent relative rates of migration, ranging from low (orange) to high (blue). $log(m)$ correspond to the effective migration rate on the log_{10} scale, relative to the overall migration rate across the species distribution.

4.5 Discussion

In the present study, we analyzed genomic RAD-seq data from 335 *Eliurus* individuals sampled across the forest fragments of northern Madagascar (Fig. 1). The dataset included genomic data for four recognized and an additional putative new *Eliurus* species (Tab. 1): *E. carletoni* (N = 142), *E. myoxinus* (N = 116), *E. ellermani* (N = 26), *E. minor* (N = 38) and *Eliurus sp.* nova (N = 13). The present work represents the first population genomic study on Malagasy endemic rodents (*Nesomynae*), since previous population genetic studies used a small number of (linked) mtDNA loci (Rakotoarisoa et al., 2010; Rakotoarisoa et al., 2013a,b). It also represents one of the few population genomic studies in Madagascar and beyond that undertake a multi-species approach to link landscape structure and genomic diversity across species associated to different habitat types. Our results show that i) HL&F and rivers influence genetic diversity similarly in the two *Eliurus* species with similar sampling distribution, the dry-associated *E. carletoni* and the humid-associated *E. myoxinus*, ii) genomic patterns are consistent among co-occurring humid-associated species, while differing from those of dry-associated species; and ii) that climatic changes through the last 25,000 years have likely influenced the distribution and diversity of dry and humid associated species in a different way. We conclude by proposing a scenario of past HL&F in northern Madagascar, distinguishing changes in dry and humid forest cover, which could be tested in future analyses.

4.5.1 Population genomic diversity of *Eliurus carletoni*

The nuclear genomic diversity (H_{obs}) of the dry habitat-associated *E. carletoni* was lower in the isolated forests of ANALV, ANKARAMY, ANALABE and ANDR (Fig. S10). On the contrary, the highest levels of genetic diversity were detected in sampling sites located within the fragmented forest massifs of the NLM region, in the northern part of *E. carletoni* distribution: Ankarana, Analamerana, Andrafiamena and Andavakoera. The high genetic diversity

found in these apparently fragmented forest massifs could be explained by the presence of forest corridors that connect most forests of the NLM region. These corridors are located in correspondence of the canyons where ground water and protection from the dry eastern winds create favorable conditions for the persistence of dry deciduous and mesic vegetations, even during the dry season (Hawkins et al., 1990).

Beyond NLM region, we found a spatial cline in *E. carletoni* genetic diversity, which decreases linearly from the sites in the north (e.g., NLM region) to the sites in the north-east (e.g., ANALV) (Fig. 2 and Fig. S11). Previous theoretical studies have shown that range expansion can generates a decrease in heterozygosity with increasing distance from the origin of expansion, caused by the series of colonization events that characterize the expansion process (e.g., Austerlitz et al., 1997; Edmonds et al., 2004; Wegmann et al., 2006; Slatkin and Excoffier, 2012). The persistence of a cline in genetic diversity due to range expansion depends on the velocity of the increase in species range area and on the age of the expansion. In fact, Garnier and Lewis, (2016), have demonstrated that under a model of range expansion (or range shift) due to climate change, a fast expansion and range shift erode genetic diversity more than slow range expansion and range shift. Accordingly, in Chapter 3, we have shown that the shape of IBD pattern generated during range expansion depends on the time since the onset of the expansion. Such behavior in IBD pattern occurs also in populations that undergo HL&F after range expansion, as it might be the case for *E. carletoni*. Therefore, the geographic cline in genetic diversity observed in *E. carletoni* could be interpreted as the result of a relatively recent range expansion.

PCA analysis shows that *E. carletoni* genomic diversity mirrors the geography of the sampled individuals and a possible role of rivers and/or non-forested habitat as barriers to gene-flow (Fig. 3). In particular, we found that the first principal component (PC1), representing the direction of largest variation in the data, separated sampling sites at the north and south of the Loky river, while PC2 separated sampling sites at the west and east of the

Manankolana river. We recovered similar patterns in ADMIXTURE, yet achieving further details on fine-scale genomic structure. At K = 2 we observed a major change in percentage of genetic ancestry (% ancestry) between samples south and north of the Loky river (e.g., SOL *vs* ANALM; Fig. 4). Such change in % ancestry is gradual and shows that, unlike sites at the west of the NLM region, ANALM and ANALM2 sites (at the east) are somewhat admixed with sites from the LM region. Considering that the gradient of % ancestry (for K=2) appears to correlate with geographic distance from ANALM/ANALM2 (putative origins of expansion) (Fig. 8), this simple observation may suggest that past *E. carletoni* dispersal between LM and NLM regions was likely occurring through the north-east rather than the north-west. ADMIXTURE results for larger K values showed i) that BEK and BOB sites (at the east of the Manankolana river) share larger % ancestry with sites south of the Manambato river (SLM region) than with the other sites of the LM region (at the west of the Manankolana river), and ii) that SLM region show admixture with sites at the east and west of the Manankolana river. These patterns could be interpreted as the result of two waves of dispersal towards the south, one from BEK-BOB and the other from the LM sites west of the Manankolana river, passing through BIN forest fragment. Importantly, these results provide also some insights on the role of rivers and habitat loss and fragmentation as barriers to gene-flow for *E. carletoni*. In fact, although major changes in genetic ancestry seems to overlap with the presence of rivers (Loky, Manankolana and Manambato), we cannot exclude that this interpretation is actually confounded by the concomitant presence in these areas of non-forested habitat. In fact, if rivers would represent strict barrier to gene-flow we would expect to observe strong genetic differentiation between sites on the opposite sides of the Manambato river (BEK-BOB *versus* SLM sites), as observed between BEK-BOB and sites at the west of the Manankolana river. Instead, ADMIXTURE results show significant level of admixture between BEK-BOB and SLM sites, even at larger K values ($2 < K < 5$).

The spatial gradient in % ancestry and the PCA-geography relationship also suggest that *E. carletoni* is characterised by geographically-limited dispersal. Accordingly, we detected significant IBD patterns across the whole *E. carletoni* sampling distribution, as well as within regions (Fig. 5 and Fig. S15). We found higher IBD slope in severely fragmented regions (LM and SLM) compared to the relatively continuous NLM region. As shown in Chapter 3, an increase in IBD slope in a fragmented landscape compared to a continuous habitat can occur when HL&F is not instantaneous, but rather follow gradual changes in landscape configuration. That is, when habitat connectivity is maintained for longer between neighbouring forests, while disrupted between forests far apart. Thus, we propose that higher IBD slope in LM region is the result of non-instantaneous HL&F compared to the NLM region, where forest corridors may maintain connectivity among sites, still nowadays. This interpretation is further supported by ADMIXTURE analysis within LM region, which showed higher genetic similarity between the neighbouring BEK and BOB sites, as well as ANTSB and BIN forests. Indeed, aerial photographs from 1949 suggests that ANTSB and BIN forests were likely connected until few decades ago (Quéméré et al., 2012), further supporting our interpretation of IBD patterns.

In conclusion, we estimated effective migration surfaces across the whole *E. carletoni* sampling distribution using a spatially-explicit model implemented in EEMS (Fig. 7). The results confirm the patterns recovered from ADMIXTURE, PCA and IBD analyses. In fact, we found that the areas with higher connectivity are placed within NLM region, with the exception of the isolated TSAKA and ANALBE forest fragments. Average to slightly low migration rates were estimated between BIN, ANTSB and AMBO, or between BEK and BOB, in agreement with regional ADMIXTURE results.

Nuclear genomic data have provided important insights on the ecology and recent history of *E. carletoni*. To summarize, we found that *E. carletoni* may have undergone range expansion in the recent past, which might have been promoted by a possible expansion of dry deciduous

vegetation. Spatial patterns across and within regions suggest that *E. carletoni* is characterized by geographically-limited dispersal. We found that rivers may not represent a strict barrier to gene-flow in this dry-habitat associated species, but instead that HL&F may impair dispersal between forest sites.

Moreover, our study shows how the large amount of information recovered from nuclear genomic data can provide insights that with lower number of genetic markers may otherwise remain hidden. In fact, our results differ in part from those of the two other genetic studies carried out on *E. carletoni*, which were based on two mitochondrial DNA loci (Rakotorisoa et al., 2013a, b). In particular, Rakotorisoa et al., (2013a) sequenced the mtDNA *cytb* and *D-loop* of individuals sampled in twelve sites within the LM region (N = 192), in two sites within NLM region (roughly, ANDRAF and ANALM; N = 10) and one site at the north of the SLM region (N = 13). Similar to our results, Rakotorisoa et al., (2013a) found the highest level of mtDNA haplotype diversity ($\overline{H}$) and proportion of unique haplotypes ($\bar{P}_u$), averaged across sites within region, in NLM ($\overline{H}$ = 1.0, $\bar{P}_u$ = 80%) compared to LM ($\overline{H}$ = 0.78, $\bar{P}_u$ = 78% and 67%) and SLM ($\overline{H}$ = 0.86, $\bar{P}_u$ = 63%) regions. They found a weak effect of Manankolana and Manambato rivers as barriers to gene-flow and higher dispersal rate between neighboring forest fragments (i.e., BEK-BOB, BOB-ANKAMY). Similar to our ADMIXTURE results, Rakotorisoa et al., (2013a) showed that a large proportion of mtDNA variation (~30%) is associated with the delineation of existing forest blocks, pointing out to the effect of HL&F on *E. carletoni* genetic differentiation. Surprisingly, they found relatively low dispersal rate between ANTSB and BIN forests, whereas our results suggest the opposite. As mentioned above, remote-sensing data show that ANTSB and BIN were likely connected in the recent past (Quéméré et al., 2012), in agreement with the results obtained from nuclear genomic data, and in disagreement with mtDNA data. Rakotorisoa et al., (2013a) also assessed the impact of geography on genetic differentiation, finding no evidence of geographically limited dispersal in *E. carletoni*, either at local or

broad regional scales. These results again differ from what nuclear genomic data suggest. Moreover, Rakotorisoa et al., (2013a, b) performed various neutrality tests (Tajima's D, Fu's F_s and R_2) and Bayesian Skyline Plot analysis to reconstruct the demographic history of *E. carletoni*. Overall, their results pointed out to a demographic bottleneck that may have occurred between 7,500 and 18,750 years ago. Conversely, spatial genomic pattern across *E. carletoni* suggest that this species may have undergone recent range expansion. The discrepancy between our and Rakotorisoa et al., (2013a, b) results may be explained by i) the strong assumptions of the demographic analysis, as well as ii) the local dynamic of the expansion process. In particular, several studies have shown that population structure leaves genetic signatures that, under the assumption of no population structure (i.e., panmixia), can be wrongly interpreted as a demographic bottleneck (Städler et al., 2009; Chikhi et al., 2010; Heller et al., 2013; Mazet et al., 2016). On the other hand, Austerlitz et al., (1997) have shown that range expansion occurs through a series of founder effect, which lead to a steady decrease in genetic diversity moving away from the origin of expansion. Thus, the signatures of local demographic bottleneck inferred with mtDNA may, together with population structure, also be caused by the local effect of the expansion process on genetic diversity. However, different pattern may be expected if genetic data were collected from the origin of expansion, which likely would show significant signatures of expansion due to a rapid demographic growth (Ray et al., 2003).

4.5.2 Population genomic diversity of *Eliurus myoxinus*

In *E. myoxinus*, the sites with lower genomic diversity were found at the southern edge of *E. myoxinus* sampling distribution, in the NM region (LOHANA, BEAMA, AMBATO and ANDRANO). Whereas, the highest H_{obs} was detected in the ANTSB and BIN neighbouring forest fragments, at the northern edge of *E. myoxinus* sampling distribution, in the LM region. We

found a steady decrease in H_{obs} with increasing geographical distance from ANTSB and BIN. As for *E. carletoni*, the linear decrease in H_{obs} is considered evidence of a recent range expansion, in this case from ANTSB-BIN, towards the southern edge of *E. myoxinus* sampling distribution. The case of *E. myoxinus* is quite peculiar, since unlike other *Eliurus* species, it has been shown to occur in both dry and humid habitats. Thus, we cannot associate *E. myoxinus* to a unique habitat its wide distribution. However, in Chapter 5 we have shown that, unlike most of its species' distribution (south and south-west Madagascar), the populations of *E. myoxinus* from northern Madagascar occur mainly in humid habitats. Thus, we will consider that the signature of a possible recent range expansion recovered in *E. myoxinus*, may be interpreted as evidence for an expansion of humid vegetation across northern Madagascar, sometime in the recent past.

PCA and ADMIXTURE analyses shows that *E. myoxinus* genomic diversity mirrors the geography of the sampled individuals, as well as recent history of habitat connectivity among neighboring forests and and a possible role of rivers as barriers to gene-flow. Since ADMIXTURE results are in agreement with PCA analysis, we will mainly discuss PCA results. In fact, it is shown that the largest direction of variation in the dataset (PC1 = 3.2%) distinguishes mainly the neighboring, but fragmented, KABE-ANTSA sites from the rest of the samples. PC2 (3%) instead separates individuals sampled in the recently connected southern edge (AMBATO, LOHANA, ANDRANO, BEAMA) from the other individuals. PC1-PC2 identifies a third cluster, composed of individuals sampled from NLM, LM and SLM regions, that follows a northwest-southeast orientation. Two observations regarding the large cluster are worth to mention. First, major gaps in the PCA plot coincide with the presence of the Loky and Manambato rivers, suggesting a possible role of rivers in limiting gene-flow. Second, sites from LM and CO regions are somewhat closely related, suggesting possible routes of connectivity between these regions through the north (LM – CO). In addition, the observation that

KABE-ANTSA are closely related to sites from SLM region, instead than the other sites of the CO region, suggest that perhaps additional habitat corridors were primarily connecting the north-east of CO region with SLM region. Previous studies have shown that differences in sample sizes across sites can distort PCA plot and thus lead to wrong conclusions (McVean, 2009). We performed additional PCA analysis by sampling equal number of individuals per sites, and the results did not change (results not shown), such that we can exclude the impact of sample size on the recovered genomic pattern.

The effect of geography in shaping *E. myoxinus* genomic diversity can also be observed from the correlation of individual geographic and genetic distances (that is, isolation-by-distance), suggesting that also *E. myoxinus* is characterized by geographically-limited dispersal. We found lower IBD slopes in the southern edge sites (NM region) and in the LM region, compared to NLM and SLM region. This is consistent with a scenario of recent connectivity within NM region and between the two forest fragments of the LM region (ANTSB-BIN). In fact, as mentioned above, ANTSB-BIN as well as AMBATO, LOHANA, ANDRANO, BEAMA (NM region) were likely connected in the recent past, no more than 70 years ago (Fig. S19) (Green and Sussman, 1990; Sussman et al., 1994; Harper et al., 2007; Quéméré et al., 2012; Vieilledent et al., 2018). Conversely, the higher IBD slope measured in NLM and SLM regions is consistent with a scenario of reduced gene-flow or relatively ancient HL&F. Estimates of effective migration surfaces across *E. myoxinus* distribution were overall supporting PCA, IBD and ADMIXTURE results. In fact, we found the highest connectivity within LM and NM regions, and lower connectivity in SLM and NLM region. However, unlike ADMIXTURE and PCA results, EEMS identified lower connectivity than average between the two neighboring ANTSA and KABE forests, suggesting that perhaps those two forests have been indeed connected for long time, but that more recently forest fragmentation has limited dispersal between them.

Besides the present work, Shi et al., (2013) is the only other genetic study focusing on *E. myoxinus*. In particular, Shi et al., (2013) sequenced one mtDNA locus (*cytb*) and two nuclear loci (*β-fib* and *GHR*) for 130 *E. myoxinus* individuals sampled in eight sites from the south-west, eight sites from the west and two sites from the north of Madagascar. Compared to our sampling sites, one of the two northern sites (site 1) included in Shi et al., (2013) corresponds to BIN (LM region), whereas the other (site 2) is located at ~15 km from AMBATO (NM region). They performed two neutrality tests across all sampling sites (Tajima's D and Fu's Fs), obtaining non-significant results for most site estimates. Interestingly, they found a negative Tajima's D and Fu's Fs for 'site 1' (=BIN) and, positive Tajima's D and Fu's Fs for site 2 (close to AMBATO), in both cases non-significant. Under panmixia and neutrality, a negative Tajima's D and Fu's Fs is interpreted as a signature of a recent population expansion (in site 1), whereas positive Tajima's D and Fu's Fs suggest population contraction (in site 2). The results of Shi et al., (2013) on the two northern sites indeed echo the cline in observed heterozygosity that we identified from ANTSB-BIN to the sites of the NM region, and that we interpreted as a possible signature of a recent expansion from ANTSB-BIN. The neutrality tests of 'site 1' support our hypothesis, whereas 'site 2' shows inconsistencies with our results, similar to those observed in relation to *E. carletoni* neutrality tests (Rakotoarisoa et al., 2013a). Briefly, we could consider as a possible explanation that individuals sampled at the 'center' or at the 'periphery' of a range expansion will not necessarily show similar local coalescent patterns, as shown in Ray et al., (2003) regarding mismatch distribution analysis. Further support of our hypothesis is given by an additional spatially-explicit analysis that Shi et al., (2013) performed to identify putative origins of expansion. They used a continuous Brownian diffusion model implemented in BEAST to infer spatial pattern of colonization and divergence across *E. myoxinus* sampling distribution. If we look only at the two northern sites, given that those are the sites included in our study, Shi et al., (2013) inferred an expansion starting at 'site 1' (=BIN) that reached

relatively recently ‘site 2’ (close to AMBATO), which again is in agreement with our results. Shi et al., (2013) have also estimated migration rate between locations on opposite sides of two large rivers: Tsiribihina river in the west and Mangoky river in the south-west. They found evidence of no or very low migration across Tsiribihina river, but evidence of asymmetrical migration across the Mangoky river, suggesting that Tsiribihina and Mangoky may represent a strict or weak barrier to dispersal, respectively. Overall, these results point out to an ambiguous role of rivers as barrier to gene-flow in *E. myoxinus*. Our PCA and ADMIXTURE results seem to support the hypothesis that rivers may represent barriers to gene-flow, however our conclusions are rather correlative and will required proper hypothesis testing.

4.5.3 Population genomic diversity of *Eliurus ellermani*

Similar to *E. myoxinus*, we recovered in *E. ellermani* a steady decline in H_{obs} from ANTSB-BIN (LM region) to ANDOHA (NM region), corresponding to the southern edge of the sampling distribution. We may interpret this pattern as evidence of a recent range expansion towards the south. Considering that *E. ellermani* is associated to humid habitat across its whole species range (Jansa et al., 2019; see Chapter 5), the inferred range expansion represents additional evidence supporting the hypothesis that humid forests were probably covering part of northern Madagascar.

PCA analysis showed a major separation between three individuals sampled in ANTSB forest and the remaining samples. The second direction of largest variation divides individuals sampled in ANTSB-BIN forest fragments (cluster 1) from all other samples (cluster 2), perhaps due to a much more recent connectivity between these two forest fragments compared to the other surveyed forests. Surprisingly, we observed that ‘cluster 2’ groups individuals sampled across the wide distribution of *E. ellermani*, with no clear correlation with their geographical location. For instance, we observed that

individuals from the far north M_AMBRE_NORTH were clustering together with individuals from AMBA, AMBALA and ANDOHA, at the southern edge of its distribution. We could exclude that this pattern is the result of unequal sample sizes across sites, given that PCA results did not change when we included only two individuals per site (results not shown). As we will discuss further below, this pattern may be interpreted as the result of routes of migration that were connecting Montagne d'Ambre with the major forest massif of CO region. Isolation by distance patterns suggest that dispersal is limited by geographical distance also in *E. ellermani*. More precisely, we found significantly higher IBD slope within the CO region compared to LM region. This is consistent with the fact that ANTSB and BIN were connected in the recent past, thus maintaining relatively low genetic differentiation. Instead, CO region includes sampling sites placed in both continuous and fragmented forests which were already fragmented in 1949 (Fig. S19), thus suggesting that ancient HL&F is the probable cause for the higher genetic distances measured within CO region. The migration surface estimated for *E. ellermani* exhibits an apparent surprising pattern. In fact, we found the highest migration rate between two sites that are placed at the opposite margins of *E. ellermani* distribution (M_AMBRE_NORTH and ANDOHA), and average connectivity between M_AMBRE_NORTH and the other sites of the CO region. Indeed, previous genetic studies on Arnhold's mouse lemur (*Microcebus arnholdi*) pointed to the same pattern (Weisrock et al., 2010; Sgarlata et al., 2019). In particular, Sgarlata et al., (2019) (see Chapter 5) sequenced 42 mouse lemurs at two mtDNA loci (*cytb* and *cox2*) and reconstructed a phylogenetic tree including all mouse lemur species for which mtDNA data were available. Their results showed that the only *M. arnholdi* individual sampled at Ambanja (far west compared to CO region) was genetically more similar to conspecifics from the far north Montagne d'Ambre, at ~150 km of distance, than with conspecifics sampled at shorter geographical distances, ~50 km (e.g., ANTSA, AMBA and ANKI). Indeed, previous studies across various taxa have provided multiple evidence of a

biogeographic link between Tsaratanana (far east of CO region) and Montagne d'Ambre, for instance, in tenrecs (*Microgale fotsyfotsy*, *Nesogale talazaci*; Everson et al., 2020), rodents (*Eliurus majori*; Everson et al., 2020), frogs (*Mantidactylus ambreensis*; Rasolonjatovo et al., 2020; *Boophis blommersae*, *B. septentrionalis, Gephyromantis ambohitra*; Andreone et al., 2009), snakes (*Pseudoxyrhopus ambreensis*; Andreone et al., 2009) and chameleons (*Brookesia minima*; Andreone et al., 2009). Moreover, in Chapter 5 we presented a new *Eliurus* taxa that we identified from individuals sampled in Montagne d'Ambre, and that may probably be genetically close to other individuals sampled in north-western Madagascar (Tysdal and Jansa, 2014), further supporting the hypothesis of a biogeographic link between CO region, north-west and far north Madagascar.

In conclusion, we compare our results on *E. ellermani* with those from the only population genetic study carried out on this species. We refer to Rakotoarisoa et al., (2013a), where the authors sequenced two mtDNA loci (*cytb* and *control region*) in 61 *E. tanala* (now *E. ellermani*; Jansa et al., 2019) individuals sampled in BIN (N = 33), ANTSB (N = 22) and Ambohibe (N = 6), a small forest fragment at ~5 km from BIN. In agreement with our results and with evidence of recent connectivity between ANTSB and BIN, they found that many haplotypes were shared between ANTSB and BIN populations, but none with Ambohibe. Accordingly, lower values of genetic differentiation were estimated between ANTSB and BIN (Φ_{ST} = 0.12), compared to Ambohibe (BIN-Ambohibe: 0.25; ANTSB-Ambohibe: 0.30). Rakotoarisoa et al., (2013a) carried out neutrality tests (Tajima's D, Fu's F_s and Ramos-Onsins and Rozas' R_2), mismatch distribution and Bayesian Skyline plot analyses to investigate the demographic history of *E. ellermani*. Neutrality tests' results, although most of them non-significant, and mismatch distribution analysis were pointing out to a historical population expansion, whereas results from BSP reconstructed a rather stable demographic history. Besides BSP analysis, historical demography results suggested that *E. ellermani* underwent

population expansion in the recent past, in agreement with the steady decline in H_{obs} that we detected from ANSTB-BIN towards the NM region.

4.5.4 Population genomic diversity of *Eliurus minor*

The lowest levels of nuclear H_{obs} were found in the three sites of the NM region (0.143), while the highest H_{obs} was measured in the ANDR forest fragment and AMBALA site. The similar H_{obs} estimated in ANDR and AMBALA sites may be related to their recent connectivity, as suggested from the forest cover from 1950s (Fig. S19). Correlation of H_{obs} with geographical distance identify ANDR as a possible origin of recent expansion, and thus as a possible refugia for *E. minor* during drier climatic periods. Analyses of population structure separate individuals from the continuous forests of the CO region from the isolated ANDR forest fragment. PCA and ADMIXTURE results show that genomic differentiation mirrors the geography of the sampled individuals. This is consistent with the significant isolation-by-distance signature detected across the study area. IBD slope was lower in NM region than CO region, in agreement with evidence of recent connectivity in NM region (Fig. S19). Accordingly, the migration rate surfaces estimated in EEMS show high migration rate across sites in the NM region, whereas lower migration rate was estimated between ANDR forest fragment and the other site of the CO region. In summary, population genomics patterns in *E. minor* confirm the likely recent connectivity of sites in NM region and less recent connectivity of ANDR forest fragment with the other site of CO region, thus suggesting that in *E. minor* dispersal is impaired by HL&F.

4.5.5 Population genomic diversity of *Eliurus sp. nova*

Eliurus sp. nova individuals were found in two sites within the montane humid forest, Montagne d'Ambre in the far north. Observed heterozygosity differed between the two sites, showing lower H_{obs} in the southern than in the northern site. PCA and ADMIXTURE analyses suggest that genomic pattern are consistent with geographically-limited dispersal, since individuals from the south of its distribution can be distinguished from those of the north. This is

further confirmed by the estimated isolation-by-distance pattern. In an unpublished study, Tysdal and Jansa, (2014) found that *Eliurus* individuals sampled in north-western Madagascar belonged to a lineage that diverged from the two closely-related species, *E. myoxinus* and *E. minor*. The position of this undescribed lineage in the *Eliurus* phylogeny coincides with the *Eliurus* sp. nova lineage identified in Chapter 5. Therefore, future genomic studies on *Eliurus* sp. nova may need to clarify whether the north-western individuals sampled by Tysdal and Jansa, (2014) would belong to the *Eliurus* sp. nova lineage. In case they do, population genomic analyses including both populations may aid, for instance, quantifying the divergence time between the two populations, and thus provide some insights on when Montagne d'Ambre lost connectivity with the other humid forests of northern Madagascar, as suggested by EEMS results in *E. ellermani*.

4.5.6 Insights on forest cover changes: dry versus humid vegetation

In the present study we have investigated nuclear population genomic pattern across five *Eliurus* species, three with a wide distribution across northern Madagascar (*E. carletoni*, *E. myoxinus* and *E. ellermani*), and two limited to fewer sites, mainly within well connected forests (*E. minor* and *Eliurus* sp. nova). The three species with larger distribution could be designated as being suitable 'model' organisms to reconstruct forest cover changes in northern Madagascar. This is because they i) occur in both fragmented and connected forest sites; and ii) present differences in habitat preferences, namely dry (*E. carletoni*) *versus* humid (*E. myoxinus* and *E. ellermani*) vegetation. This may become clearer if we take into account that *Eliurus* sp. are forest-dwelling species (Carleton, 2003), so that dispersal and population size are mainly influenced by forest connectivity and cover. Thus, we expect that pattern of neutral genomic relatedness between fragmented and continuous forests may reliably elucidate on past changes in habitat

connectivity. Moreover, the fact that these three species differ in habitat preferences allows to distinguish past changes in dry and humid forest.

Among the three widespread species, *E. carletoni* and *E. ellermani* are those for which occurrence data suggest a clear preference towards dry and humid vegetations, respectively. Different is the case of *E. myoxinus*, which has been generally associated to dry habitats, although earlier surveys found *E. myoxinus* individuals in two humid sites of northern (BIN and NM region) and southern Madagascar (canyons of Isalo and Analavelona Massif) (Carleton et al., 2001; Goodman and Soarimalala, 2002; Soarimalala and Goodman, 2003). In Chapter 5, we have shown for the first time that the occurrence of *E. myoxinus* in humid habitats should not be considered occasional in northern Madagascar, but rather common. On this regard, the population genomic analyses of the present study have provided further evidence on the preference of *E. myoxinus* populations in northern Madagascar for humid habitat. In particular, we have identified a linear decrease in H_{obs} with increasing distance from ANTSB-BIN forests which coincides with the one found in the humid-associated *E. ellermani*. These results suggest that both species may have undergone recent range expansion following the expansion of humid vegetation. Accordingly, we found strong correlation in population genomic diversity (ρ = 1) and pairwise population genetic distances between *E. myoxinus* and *E. ellermani* (ρ for pairwise F_{st} = 1; ρ for population H_{obs} = 0.84), while none with the dry-associated *E. carletoni*. In few words, occurrence data and genomic pattern suggest that, in northern Madagascar, *E. myoxinus* populations show preference for humid vegetations, and thus information on historical migration routes across those populations can be considered a reasonable proxy for reconstructing the history of humid forest cover changes in northern Madagascar.

We should start by considering that genomic diversity patterns from the two humid associated species (*E. myoxinus* and *E. ellermani*) suggest that, at some point in the recent past, humid forests were connecting ANTSB-

BIN sites with other humid forests in the north-east. Similarly, nuclear genomic data from the humid-associated Anrhold's mouse lemur (*Microcbeus arnholdi*) sampled in BIN and in several forest fragments of the SLM region showed a steady decline in H_{obs} from SLM region to BIN forest, suggesting an expansion from the northeast to the north (see Annex). The implicit assumption underlying the expansion scenario is that humid vegetations persisted in ANTSB and BIN massifs during the drier period preceding humid forest expansion. Indeed, this is in agreement with previous studies suggesting that ANTSB and BIN forests likely acted as montane refugia for various taxa during the drier periods of the Late Quaternary (Raxworthy and Nussbaum, 1994; Goodman and Wilmé, 2006; Rakotoarisoa et al., 2013a; Salmona, 2015; Sgarlata et al., 2019; Salmona et al., 2021).

The other observation to be made is that *E. carletoni* shows evidence of range expansion from the nearly continuous dry forests of the NLM region (Analamerana, Ankarana and Andrafiamena massifs), towards the south of its distribution. Evidence of a recent range expansion were also detected from mtDNA in populations (within LM region) of the dry-associated mouse lemur (*M. tavaratra*), which distribution, similar to *E. carletoni*, extend from the far north to north-eastern Madagascar (Sgarlata et al., 2018, 2019). Furthermore, nuclear genomic analyses on *M. tavaratra* and on the dry-associated sportive lemur *Lepilemur ankaranensis* show the same decline in H_{obs} from NLM region down to the north-east Madagascar, once again pointing out to a recent range expansion of dry ecosystems from the NLM region (see Annex).

As we have discussed, genomic and genetic data across various taxa suggest that both dry and humid vegetations underwent range expansion at some point in the past. It is reasonable to assume that expansion of these two forest types may not have happened simultaneously, but without proper estimates of expansion time is difficult to know which one occurred first. However, we can make use of independent sources of information (e.g., palaeoecological data) to propose a parsimonious scenario explaining biogeographic, genetic and genomic patterns across all these species.

It is now well established that a shift from humid to dry forest cover has occurred in northern Madagascar. For instance, several palaeoecological studies carried out in the present-day dry NLM region (Ankarana, Andavakoera and Montagne des Français) found subfossils of extant lemur species (*Propithecus cf. diadema*, *Indri indri*, and *Prolemur simus*) which currently occur only in the humid forests of eastern Madagascar (Jungers et al., 1995; Simons et al., 1995; Godfrey et al., 1999). The radiocarbon dating analysis of *P. simus* and *I. indri* specimens estimated that those individuals were alive about 4,560 ± 70 Cal. YBP and 4,400 ± 60 Cal. YBP, respectively, suggesting that northern Madagascar was more humid than today until at least ~4,500 years ago (Simons et al., 1995; Jungers et al., 1995). Consistently, recent paleoenvironmental reconstructions from a sedimentary record collected in Montagne d'Ambre revealed that dry-humid climatic oscillations occurred during the last 25,000 years in northern Madagascar (Teixeira et al., 2021). Notably, Teixeira et al., (2021) identified five climatic time intervals associated with different vegetation assemblages, that is: i) a first cold and dry period between 25,000 and 15,200 years ago; ii) a second and third period characterized by higher humidity and temperature between 15,200 and 5,500 years ago, iii) a fourth period associated to reduced precipitation between 5,500 and 900 years ago; and iv) a fifth period between 900 years ago and present-day marked by an important increase in charcoal particles and amount of *Poaceae* pollen, which are indicators of higher fire frequency and occurrence of open-grassland ecosystems, respectively.

Subfossil, palynological, species distribution, genomic and genetic data, together, help us to draw a clearer picture of what happened across northern Madagascar during the last 25,000 years.

According to Teixeira et al., (2021), climatic conditions were drier and cooler between 25,000 and 15,000 years ago than in modern times, such that dry vegetations may have expanded during this period. Similar dry conditions have been recorded during the same period in tropical eastern Africa (Ivory and Russell, 2016; Los et al., 2019), which were accompanied by a savanna

expansion in African lowlands (e.g., Flenley, 1998). Consequently, in northern Madagascar, humid vegetations were likely limited to mid-altitude montane massifs (e.g., Montagne d'Ambre, Binara and Antsahabe), since altitude would maintain favourable conditions for the persistence of humid microclimate even during dry periods, as it has been demonstrated for other tropical regions (e.g., Anderson et al., 2018). Therefore, between 25,000 and 15,000 years ago, dry-associated species would have had a wide distribution, in contrast to the scattered (and fragmented) distribution of humid-associated species.

Between ~15,200 and ~5,000 years ago, climate was more humid, not only in northern Madagascar (Teixeira et al., 2021), but also in mainland Africa, so that this period has been called 'African Humid Period' (AHP). This period was characterised by major ecosystem changes in mainland Africa, such as the greening of the Sahara (Kutzbach and Street-Perrott, 1985; Gasse, 2000). Therefore, it seems plausible to assume that in northern Madagascar, as in mainland Africa, humid vegetations were covering areas that today are composed by dry vegetation. The transition from a dry to a humid climate and forest cover promoted the expansion of humid associated species, such as *E. ellermani*, *E. myoxinus* and *M. arnholdi* and the contraction of dry-associated species. The latter is supported by genetic studies that have dated a population decline between 10,000 and 4,000 years ago in two medium-size lemur species, *Propithecus tattersalli* and *Propitehcus perrieri*, occurring in the dry deciduous forests of LM and NLM regions, respectively (Quéméré et al., 2012; Salmona et al., 2017). Under this scenario, it makes sense to consider that major corridors of humid vegetations were being established between ANSTB, BIN and the other humid forests south of ANTSB-BIN massifs, since these are geographically closer and have been keeping humid habitat even in the dry modern times.

Towards the end of the AHP, about ~5,000 years ago, palaeoecological data recorded a reduction in precipitations (Teixeira et al., 2021), which may have led to the expansion of dry vegetation. Consequently,

humid vegetation may have undergone HL&F by shifting upward its lower altitudinal limit, similar to what it has been reported for eastern Africa (Ivory and Russell, 2018). We propose that this last major shift, from humid to dry vegetation, promoted the recent expansion of dry-associated species, such as *E. carletoni*, *M. tavaratra* and *L. ankaranensis*, from the north to north-eastern Madagascar. Under this phylogeographic model, however, it remains unclear why the Analamerana, Akarana, Andavakoera massifs of NLM region would have represented the origin of expansion for dry-associated species. In other words, were the massifs of NLM region somehow maintaining dry local conditions, while parts of northern Madagascar were covered by humid vegetation? The presence of the lemur Perrier's sifaka (*Propithecus perrieri*) in the forest massifs of NLM region could give some insights on this regard. In fact, *P. perrieri* is a medium-size lemur species with a narrow distribution (~210 km^2) limited to the dry deciduous and semi-evergreen forests of Analamerana, Antsoy and Andriafiamena massifs (Lehman and Mayor, 2004; Banks, 2013; Salmona et al., 2013). Although it is currently associated to dry deciduous habitat, surprisingly, it is genetically closer to the humid-associated sifakas species of eastern Madagascar than to the dry-associated sifakas species of western Madagascar (Mayor et al., 2004; Rumpler et al., 2011). This indicates that perhaps the ancestor of *P. perrieri* came from the humid forests of eastern Madagascar and genetically adapted to survive in the dry deciduous forests of NLM region. *P. perrieri* adaptation to dry vegetation and its persistence even in modern times would suggest that perhaps 'islands' of dry forests were occurring in NLM region, even during periods of major vegetation transition from dry to humid (e.g., AHP, 15,000 – 5,000 years ago). Accordingly, the presence of such 'islands' of dry forests may explain why dry-associated species show decline in H_{obs} from NLM region, consistent with a putative origin of expansion from this area.

Future genetic analyses aiming at dating the onset of range expansion and past HL&F in *E. carletoni*, *M. tavaratra*, *L. ankaranensis*, *E. ellermani*, *E.*

myoxinus and *M. arnholdi* are warranted to test the phylogeographic model here proposed.

4.6 References

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19, 1655–1664.

Anderson, J., Keppel, G., Thomson, S.-M., Randell, A., Raituva, J., Koroi, I., Anisi, R., Charlson, T., Boehmer, H.J., Kleindorfer, S., 2018. Changes in climate and vegetation with altitude on Mount Batilamu, Viti Levu, Fiji. Journal of Tropical Ecology 34, 316–325.

Andreone, F., Glaw, F., Mattioli, F., Jesu, R., Schimmenti, G., Randrianirina, J.E., Vences, M., 2009. The peculiar herpetofauna of some Tsaratanana rainforests and its affinities with Manongarivo and other massifs and forests of northern Madagascar. Italian Journal of Zoology 76, 92–110.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics 17, 81–92.

Andrews, S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data.

Arredondo, A., Mourato, B., Nguyen, K., Boitard, S., Rodríguez, W., Noûs, C., Mazet, O., Chikhi, L., 2021. Inferring number of populations and changes in connectivity under the n-island model. Heredity 126, 896–912.

Austerlitz, F., Jung-Muller, B., Godelle, B., Gouyon, P.-H., 1997. Evolution of Coalescence Times, Genetic Diversity and Structure during Colonization. Theoretical Population Biology 51, 148–164.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLOS ONE 3, e3376.

Banks, M.A., 2013. Determinants of abundance and the distribution of primates in northern Madagascar. (PhD thesis). Stony Brook University.

Beaumont, M.A., 2010. Approximate Bayesian Computation in Evolution and Ecology. Annual Review of Ecology, Evolution, and Systematics 41, 379–406.

Beerli, P., Felsenstein, J., 1999. Maximum-Likelihood Estimation of Migration Rates and Effective Population Numbers in Two Populations Using a Coalescent Approach. Genetics 152, 763–773.

Bésairie, H., 1964. Carte géologique de Madagascar, au 1:1,000,000e, trois feuilles en couleur.

Buřivalová, Z., 2011. Remote sensing of vegetation in conservation: A case study from the dry and transitional forests of Andrafiamena, northern Madagascar. Université de Genève and Conservatoire et Jardin botaniques de la Ville de Genève, Genève.

Campbell, G., 2013. Forest Depletion in Imperial Madagascar, c.1790–1861. Contest for Land in Madagascar 63–95.

Carleton, M., Goodman, S., 1998. New taxa of nesomyine rodents (Muroidea: Muridae) from Madagascar's northern Highlands, with taxonomic comments on previously described forms. Fieldiana Zoology 1998, 163–200.

Carleton, M.D., 2003. Eliurus, tufted-tailed rats., in: The Natural History of Madagascar. University of Chicago Press, Chicago, pp. 1373–1380.

Carleton, M.D., 1994. Systematic studies of Madagascar's endemic rodents (Muroidea: Nesomyinae) : revision of the genus Eliurus. American Museum novitates ; no. 3087.

Carleton, M.D., Goodman, S.M., 2007. A new species of the Eliurus majori complex (Rodentia, Muroidea, Nesomyidae) from south-central Madagascar, with remarks on emergent species groupings in the genus Eliurus ; American Museum novitates, no. 3547. Eliurus majori complex.

Carleton, M.D., Goodman, S.M., Rakotondravony, D., 2001. A new species of tufted-tailed rat, genus Eliurus (Muridae : Nesomyinae), from western Madagascar, with notes on the distribution of E-myoxinus.

Charlesworth, B., Charlesworth, D., 2017. Population genetics from 1966 to 2016. Heredity 118, 2–9.

Chikhi, L., Sousa, V.C., Luisi, P., Goossens, B., Beaumont, M.A., 2010. The Confounding Effects of Population Structure, Genetic Diversity and the Sampling Scheme on the Detection and Quantification of Population Size Changes. Genetics 186, 983–995.

Edmonds, C.A., Lillie, A.S., Cavalli-Sforza, L.L., 2004. Mutations arising in the wave front of an expanding population. PNAS 101, 975–979.

Ellegren, H., Galtier, N., 2016. Determinants of genetic diversity. Nat Rev Genet 17, 422–433.

Epps, C.W., Keyghobadi, N., 2015. Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. Molecular Ecology 24, 6021–6040.

Everson, K.M., Goodman, S.M., Olson, L.E., 2020. Speciation and gene flow in two sympatric small mammals from Madagascar, Microgale fotsifotsy and M. soricoides (Mammalia: Tenrecidae). Molecular Ecology 29, 1717–1729.

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048.

Flenley, J.R., 1998. Tropical Forests Under the Climates of the Last 30,000 Years. Climatic Change 39, 177–197.

Fowler, S.V., Chapman, P., Checkley, D., Hurd, S., McHale, M., Ramangason, G.-S., Randriamasy, J.-E., Stewart, P., Walters, R., Wilson, J.M., 1989. Survey and management proposals for a tropical deciduous forest reserve at Ankarana in Northern Madagascar. Biological Conservation 47, 297–313.

Garbutt, N., 1999. Mammals of Madagascar. Yale University Press.

Gasse, F., 2000. Hydrological changes in the African tropics since the Last Glacial Maximum. Quaternary Science Reviews 19, 189–211.

Gautier, L., Ranirison, P., Nusbaumer, L., Wohlhauser, S., 2006. Aperçu des massifs forestiers de la région Loky-Manambato, in: Inventaires de La Faune

et de La Flore Du Nord de Madagascar Dans La Région Loky-Manambato, Analamerana et Andavakoera. Rech. Développement. Sér Sci Biol.

Gibbard, P., Lewin, J., 2016. Partitioning the Quaternary.

Godfrey, L.R., Jungers, W.L., Simons, E.L., Chatrath, P.S., Rakotosamimanana, B., 1999. Past and Present Distributions of Lemurs in Madagascar, in: Rakotosamimanana, B., Rasamimanana, H., Ganzhorn, J.U., Goodman, S.M. (Eds.), New Directions in Lemur Studies. Springer US, Boston, MA, pp. 19–53.

Goodman, S.M., Carleton, M.D., 1996. The rodents of the Réserve Naturelle Intégrale d'Andringitra, Madagascar. Fieldiana: Zoology (new series) 257–283.

Goodman, S.M., Carleton, M.D., Pidgeon, M., 1999. The rodents of the Réserve Naturelle Intégrale d'Andohahela, Madagascar. Fieldiana: Zoology (new series) 217–249.

Goodman, S.M., Ganzhorn, J.U., 1994. Les petits mammiferes, in: Inventaire Biologique Foret de Zombitse,Recherches Pour Le Developpement, Serie Sciences Biologiques. Centre d'Information et de Documentation Scientifique et Technique, Antananarivo, Madagascar, pp. 58–63.

Goodman, S.M., Raheriarisena, M., Jansa, S.A., 2009. A new species of Eliurus Milne Edwards, 1885 (Rodentia: Nesomyinae) from the Réserve Spéciale d'Ankarana, northern Madagascar | Request PDF. Bonner Zoologische Beiträge 56, 133–149.

Goodman, S.M., Raherilalao, M.J., Wohlhauser, S. (Eds.), 2021. The Terrestrial Protected Areas of Madagascar: Their History, Description, and

Biota, Volume 3: Western and southwestern Madagascar - Synthesis. Association Vahatra in Antananarivo.

Goodman, S.M., Rasoloarison, R., 1997. Les petits mammiferes, in: Inventaire Biologique Foret de Vohibasia et d'Isoky-Vohimena. Recherches Pour Le Developpement, Serie Sciences Biologiques. Centre d'Information et de Documentation Scientifique et Technique, Antananarivo, Madagascar, pp. 144–155.

Goodman, S.M., Soarimalala, V., 2002. Les petits mammifères de la Réserve Spéciale de Manongarivo, Madagascar, in: Inventaire Floristique et Faunistique de La Réserve Spéciale de Manongarivo (NW Madagascar). pp. 383–401.

Goodman, S.M., Wilmé, L., 2006. Inventaires de la faune et de la flore du nord de Madagascar dans la région Loky-Manambato, Analamerana et Andavakoera.

Graham, R.W., Lundelius, E.L., Graham, M.A., Schroeder, E.K., Toomey, R.S., Anderson, E., Barnosky, A.D., Burns, J.A., Churcher, C.S., Grayson, D.K., Guthrie, R.D., Harington, C.R., Jefferson, G.T., Martin, L.D., McDonald, H.G., Morlan, R.E., Semken, H.A., Webb, S.D., Werdelin, L., Wilson, M.C., 1996. Spatial Response of Mammals to Late Quaternary Environmental Fluctuations. Science 272, 1601–1606.

Green, G.M., Sussman, R.W., 1990. Deforestation History of the Eastern Rain Forests of Madagascar from Satellite Images. Science 248, 212–215.

Gruber, B., Unmack, P.J., Berry, O.F., Georges, A., 2018. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Mol Ecol Resour 18, 691–699.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342, 850–853.

Harper, G.J., Steininger, M.K., Tucker, C.J., Juhn, D., Hawkins, F., 2007. Fifty years of deforestation and forest fragmentation in Madagascar. Environmental Conservation 34, 325–333.

Hawkins, A., Chapman, P., Ganzhorn, J., Bloxam, Q., Barlow, S., Tonge, S., 1990. Vertebrate conservation in Ankarana special reserve, Northern Madagascar.

Heller, R., Chikhi, L., Siegismund, H.R., 2013. The Confounding Effect of Population Structure on Bayesian Skyline Plot Inferences of Demographic History. PLOS ONE 8, e62992.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907–913.

Hey, J., Nielsen, R., 2004. Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, With Applications to the Divergence of Drosophila pseudoobscura and D. persimilis. Genetics 167, 747–760.

Hu, F.S., Hampe, A., Petit, R.J., 2009. Paleoecology meets genetics: deciphering past vegetational dynamics. Frontiers in Ecology and the Environment 7, 371–379.

IPBES, 2021. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Zenodo.

Ivory, S.J., Russell, J., 2016. Climate, herbivory, and fire controls on tropical African forest for the last 60ka. Quaternary Science Reviews 148, 101–114.

Jansa, S.A., Carleton, M.D., Soarimalala, V., Rakotomalala, Z., Goodman, S.M., 2019. A review of the Eliurus tanala complex (Rodentia, Muroidea, Nesomyidae), with description of a new species from dry forests of western Madagascar. (Bulletin of the American Museum of Natural History, no. 430).

Jarosz, L., 1993. Defining and explaining tropical deforestation: shifting cultivation and population growth in colonial Madagascar (1896-1940). Econ Geogr 69, 366–379.

Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070–3071.

Jungers, W.L., Godfrey, L.R., Simons, E.L., Chatrath, P.S., 1995. Subfossil Indri indri from the Ankarana Massif of northern Madagascar. Am J Phys Anthropol 97, 357–366.

Knauer, K., Gessner, U., Dech, S., Kuenzer, C., 2014. Remote sensing of vegetation dynamics in West Africa. International Journal of Remote Sensing 35, 6357–6396.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15, 1179–1191.

Kuhner, M.K., Yamato, J., Felsenstein, J., 1995. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. Genetics 140, 1421–1430.

Kutzbach, J.E., Street-Perrott, F.A., 1985. Milankovitch forcing of fluctuations in the level of tropical lakes from 18 to 0 kyr BP. Nature 317, 130–134.

Lehman, S.M., Mayor, M., 2004. Dietary patterns in Perrier's sifakas (Propithecus diadema perrieri): a preliminary study. Am J Primatol 62, 115–122.

Lewis, S.L., Maslin, M.A., 2015. Defining the Anthropocene. Nature 519, 171–180.

Loog, L., 2021. Sometimes hidden but always there: the assumptions underlying genetic inference of demographic histories. Philosophical Transactions of the Royal Society B: Biological Sciences 376, 20190719.

Los, S.O., Street-Perrott, F.A., Loader, N.J., Froyd, C.A., Cuní-Sanchez, A., Marchant, R.A., 2019. Sensitivity of a tropical montane cloud forest to climate change, present, past and future: Mt. Marsabit, N. Kenya. Quaternary Science Reviews 218, 34–48.

Lyons, S.K., 2005. A Quantitative Model for Assessing Community Dynamics of Pleistocene Mammals. The American Naturalist 165, E168–E185.

Lyons, S.K., 2003. A Quantitative Assessment of the Range Shifts of Pleistocene Mammals. Journal of Mammalogy 84, 385–402.

Major, C.I.F., 1896. Descriptions of four additional new mammals from Madagascar. Annals and Magazine of Natural History 461–463.

Marquart, K., 2014. Habitat use and morphological adaptations of endemic rodents (Muroidea: Nesomyinae) of East Madagascar. Universität Hohenheim, Munich.

Mayewski, P.A., Rohling, E.E., Curt Stager, J., Karlén, W., Maasch, K.A., David Meeker, L., Meyerson, E.A., Gasse, F., van Kreveld, S., Holmgren, K., Lee-Thorp, J., Rosqvist, G., Rack, F., Staubwasser, M., Schneider, R.R., Steig, E.J., 2004. Holocene climate variability. Quaternary Research 62, 243–255.

Mayor, M.I., Sommer, J.A., Houck, M.L., Zaonarivelo, J.R., Wright, P.C., Ingram, C., Engel, S.R., Louis, E.E., 2004. Specific Status of Propithecus spp. International Journal of Primatology 25, 875–900.

Mazet, O., Rodríguez, W., Grusea, S., Boitard, S., Chikhi, L., 2016. On the importance of being structured: instantaneous coalescence rates and human evolution--lessons for ancestral population size inference? Heredity (Edinb) 116, 362–371.

McConnell, W.J., Kull, C.A., 2014. Deforestation in Madagascar: debates over the island's forest cover and challenges of measuring forest change, in: Conservation and Environmental Management in Madagascar. Routledge.

McVean, G., 2009. A Genealogical Interpretation of Principal Components Analysis. PLOS Genetics 5, e1000686.

Milne Edwards, A., 1885. Description d'une nouvelle espèce de rongeur provenant de Madagascar. Annales des Sciences naturelles, Zoologie et Paléontologie.

Mitchell, A.L., Rosenqvist, A., Mora, B., 2017. Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+. Carbon Balance and Management 12, 9.

Mitchell, K.J., Rawlence, N.J., 2021. Examining Natural History through the Lens of Palaeogenomics. Trends in Ecology & Evolution 36, 258–267.

Nei, M., 1987. Molecular Evolutionary Genetics, Molecular Evolutionary Genetics. Columbia University Press.

Nicoll, M.E., Legrand, 1989. Madagascar : revue de la conservation et des aires protégées.

Nielsen, R., 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics 154, 931–942.

Nielsen, R., Beaumont, M.A., 2009. Statistical inferences in phylogeography. Mol Ecol 18, 1034–1047.

Paris, J.R., Stevens, J.R., Catchen, J.M., 2017. Lost in parameter space: a road map for stacks. Methods in Ecology and Evolution 8, 1360–1373.

Petkova, D., Novembre, J., Stephens, M., 2018. EEMS.

Petkova, D., Novembre, J., Stephens, M., 2016. Visualizing spatial population structure with estimated effective migration surfaces. Nature Genetics 48, 94–100.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet 81, 559–575.

Puy, D.D., Moat, J., 1996. A refined classification of the primary vegetation of Madagascar based on the underlying geology: Using GIS to map its distribution and to assess its conservation status.

Quéméré, E., Amelot, X., Pierson, J., Crouau-Roy, B., Chikhi, L., 2012. Genetic data suggest a natural prehuman origin of open habitats in northern Madagascar and question the deforestation narrative in this region. PNAS 109, 13028–13033.

Rakotoarisoa, J.E., Raheriarisena, M., Goodman, S.M., 2013a. Late Quaternary climatic vegetational shifts in an ecological transition zone of northern Madagascar: insights from genetic analyses of two endemic rodent species. Journal of Evolutionary Biology 26, 1019–1034.

Rakotoarisoa, J. E., Raheriarisena, M., Goodman, S.M., 2013b. A phylogeographic study of the endemic rodent Eliurus carletoni (Rodentia: Nesomyinae) in an ecological transition zone of Northern Madagascar. J. Hered. 104, 23–35.

Rakotoarisoa, J.E., Raheriarisena, M., Goodman, S.M., 2010. Phylogeny and species boundaries of the endemic species complex, Eliurus antsingy and E. carletoni (Rodentia: Muroidea: Nesomyidae), in Madagascar using mitochondrial and nuclear DNA sequence data. Molecular Phylogenetics and Evolution 57, 11–22.

Rakotondravony, R., Radespiel, U., 2009. Varying patterns of coexistence of two mouse lemur species (Microcebus ravelobensis and M. murinus) in a heterogeneous landscape. American Journal of Primatology 71, 928–938.

Rasolonjatovo, S.M., Scherz, M.D., Hutter, C.R., Glaw, F., Rakotoarison, A., Razafindraibe, J.H., Goodman, S.M., Raselimanana, A.P., Vences, M., 2020. Sympatric lineages in the Mantidactylus ambreensis complex of Malagasy frogs originated allopatrically rather than by in-situ speciation. Molecular Phylogenetics and Evolution 144, 106700.

Raxworthy, C.J., Nussbaum, R.A., 1994. A rainforest survey of amphibians, reptiles and small mammals at Montagne d'Ambre, Madagascar. Biological Conservation 69, 65–73.

Ray, N., Currat, M., Excoffier, L., 2003. Intra-deme molecular diversity in spatially expanding populations. Mol Biol Evol 20, 76–86.

Rochette, N.C., Catchen, J.M., 2017. Deriving genotypes from RAD-seq short-read data using Stacks. Nat Protoc 12, 2640–2659.

Rochette, N.C., Rivera-Colón, A.G., Catchen, J.M., 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Molecular Ecology 28, 4737–4754.

Roerink, G.J., Menenti, M., Soepboer, W., Su, Z., 2003. Assessment of climate impact on vegetation dynamics by using remote sensing. Physics and Chemistry of the Earth, Parts A/B/C, Applications of Quantitative Remote Sensing to Hydrology 28, 103–109.

Rossi, G., 1974. Morphologie et évolution d'un karst en milieu tropical : l'Ankarana (Madagascar)., in: Mém. et Doc. Du C.N.R.S. Vol. 15. Phénomènes Karstiques II. Paris, pp. 279–298.

Rumpler, Y., Hauwy, M., Fausser, J.-L., Roos, C., Zaramody, A., Andriaholinirina, N., Zinner, D., 2011. Comparing chromosomal and mitochondrial phylogenies of the Indriidae (Primates, Lemuriformes). Chromosome Res 19, 209–224.

Salmona, J., 2015. Comparative conservation genetics of several threatened lemur species living in fragmented environments. A glimpse throught natural history of northern Madagascar lemurs. Instituto Gulbenkian de Ciencia, Lisbon.

Salmona, J., Dresen, A., Ranaivoson, A.E., Manzi, S., Pors, B.L., Hong-Wa, C., Razanatsoa, J., Andriaholinirina, N.V., Rasoloharijaona, S., Vavitsara, M.-E., Besnard, G., 2021. How ancient forest fragmentation and riparian connectivity generate high levels of genetic diversity in a micro-endemic Malagasy tree.

Salmona, J., Heller, R., Quéméré, E., Chikhi, L., 2017. Climate change and human colonization triggered habitat loss and fragmentation in Madagascar. Mol. Ecol. 26, 5203–5222.

Salmona, J., Jan, F., Rasolondraibe, E., Zaranaina, R., Said Ousseni, D., Mohamed-Thani, I., et al., 2013. Survey of the critically endangered Perrier's

sifaka (Propithecus perrieri) across most if its distribution range. Lemur News 17, 9–12.

Sandel, B., Arge, L., Dalsgaard, B., Davies, R.G., Gaston, K.J., Sutherland, W.J., Svenning, J.-C., 2011. The Influence of Late Quaternary Climate-Change Velocity on Species Endemism. Science 334, 660–664.

Seppä, H., 2018. Palaeoecology, in: ELS. John Wiley & Sons, Ltd, pp. 1–8.

Seutin, G., White, B.N., Boag, P.T., 1991. Preservation of avian blood and tissue samples for DNA analyses. Can. J. Zool. 69, 82–90.

Sgarlata, G.M., Salmona, J., Aleixo-Pais, I., Rakotonanahary, A., Sousa, A.P., Kun-Rodrigues, C., Ralantoharijaona, T., Jan, F., Zaranaina, R., Rasolondraibe, E., Zaonarivelo, J.R., Andriaholinirina, N.V., Chikhi, L., 2018. Genetic Differentiation and Demographic History of the Northern Rufous Mouse Lemur (Microcebus tavaratra) Across a Fragmented Landscape in Northern Madagascar. Int J Primatol 39, 65–89.

Sgarlata, G.M., Salmona, J., Pors, B.L., Rasolondraibe, E., Jan, F., Ralantoharijaona, T., Rakotonanahary, A., Randriamaroson, J., Marques, A.J., Aleixo-Pais, I., Zoeten, T. de, Ousseni, D.S.A., Knoop, S.B., Teixeira, H., Gabillaud, V., Miller, A., Ibouroi, M.T., Rasoloharijaona, S., Zaonarivelo, J.R., Andriaholinirina, N.V., Chikhi, L., 2019. Genetic and morphological diversity of mouse lemurs (Microcebus spp.) in northern Madagascar: The discovery of a putative new species? American Journal of Primatology 81, e23070.

Shi, J.J., Chan, L.M., Rakotomalala, Z., Heilman, A.M., Goodman, S.M., Yoder, A.D., 2013. Latitude drives diversification in Madagascar's endemic

dry forest rodent Eliurus myoxinus (subfamily Nesomyinae). Biological Journal of the Linnean Society 110, 500–517.

Simons, E.L., Burney, D.A., Chatrath, P.S., Godfrey, L.R., Jungers, W.L., Rakotosamimanana, B., 1995. AMS 14C Dates for Extinct Lemurs from Caves in the Ankarana Massif, Northern Madagascar. Quaternary Research 43, 249–254.

Singh, R., 2021. Re-envisioning Remote Sensing Applications: Perspectives from Developing Countries. CRC Press.

Slatkin, M., Excoffier, L., 2012. Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. Genetics 191, 171–181.

Smouse, P.E., Peakall, R., 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity 82, 561–573.

Soarimalala, V., Goodman, S.M., 2011. Les Petits Mammifères de Madagascar. Association Vahatra, Antananarivo, Madagascar.

Soarimalala, V., Goodman, S.M., 2003. Diversité biologique des micromammifères non volants (Liptotyphla et Rodentia) dans le complexe Marojejy-Anjanaharibe-Sud. Recherches pour le Développement, Série Sciences Biologiques 231–278.

Städler, T., Haubold, B., Merino, C., Stephan, W., Pfaffelhuber, P., 2009. The Impact of Sampling Schemes on the Site Frequency Spectrum in Nonequilibrium Subdivided Populations. Genetics 182, 205–216.

Sussman, R.W., Green, G.M., Sussman, L.K., 1994. Satellite Imagery, Human Ecology, Anthropology, and Deforestation in Madagascar. Human Ecology 22, 333–354.

Tang, Q., Fung, T., Rheindt, F.E., 2020. ResDisMapper: An r package for fine-scale mapping of resistance to dispersal. Molecular Ecology Resources 20.

Teixeira, H., Montade, V., Salmona, J., Metzger, J., Bremond, L., Kasper, T., Daut, G., Rouland, S., Ranarilalatiana, S., Rakotondravony, R., Chikhi, L., Behling, H., Radespiel, U., 2021. Past environmental changes affected lemur population dynamics prior to human impact in Madagascar. Commun Biol 4, 1–10.

Tysdal, J.A., Jansa, S.A., 2014. Phylogeny and biogeography of an endemic Madagascar rodent.

Veress, M., Lóczy, D., Zentai, Z., Tóth, G., Schläffer, R., 2008. The origin of the Bemaraha tsingy (Madagascar). International Journal of Speleology 37.

Vieilledent, G., Grinand, C., Rakotomalala, F.A., Ranaivosoa, R., Rakotoarijaona, J.-R., Allnutt, T.F., Achard, F., 2018. Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. Biological Conservation 222, 189–197.

Wegmann, D., Currat, M., Excoffier, L., 2006. Molecular Diversity After a Range Expansion in Heterogeneous Environments. Genetics 174, 2009–2020.

Weigelt, P., Steinbauer, M.J., Cabral, J.S., Kreft, H., 2016. Late Quaternary climate change shapes island biodiversity. Nature 532, 99–102.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution 38, 1358–1370.

Weisrock, D.W., Rasoloarison, R.M., Fiorentino, I., Ralison, J.M., Goodman, S.M., Kappeler, P.M., Yoder, A.D., 2010. Delimiting Species without Nuclear Monophyly in Madagascar's Mouse Lemurs. PLOS ONE 5, e9883.

Wiens, J.J., Ackerly, D.D., Allen, A.P., Anacker, B.L., Buckley, L.B., Cornell, H.V., Damschen, E.I., Davies, T.J., Grytnes, J.-A., Harrison, S.P., Hawkins, B.A., Holt, R.D., McCain, C.M., Stephens, P.R., 2010. Niche conservatism as an emerging principle in ecology and conservation biology. Ecology Letters 13, 1310–1324.

Williams, J.W., Summers, R.L., Webb III, T., 1998. Applying plant functional types to construct biome maps from eastern North American pollen data: comparisons with model results. Quaternary Science Reviews 17, 607–627.

Wright, S., 1949. The Genetical Structure of Populations. Annals of Eugenics 15, 323–354.

Wright, S., 1943. Isolation by Distance. Genetics 28, 114–138.

Xie, Y., Sha, Z., Yu, M., 2008. Remote sensing imagery in vegetation mapping: a review. Journal of Plant Ecology 1, 9–23.

Zhu, A., 2017. Rosewood occidentalism and orientalism in Madagascar. Geoforum 86, 1–12.

4.7 Supplementary Information

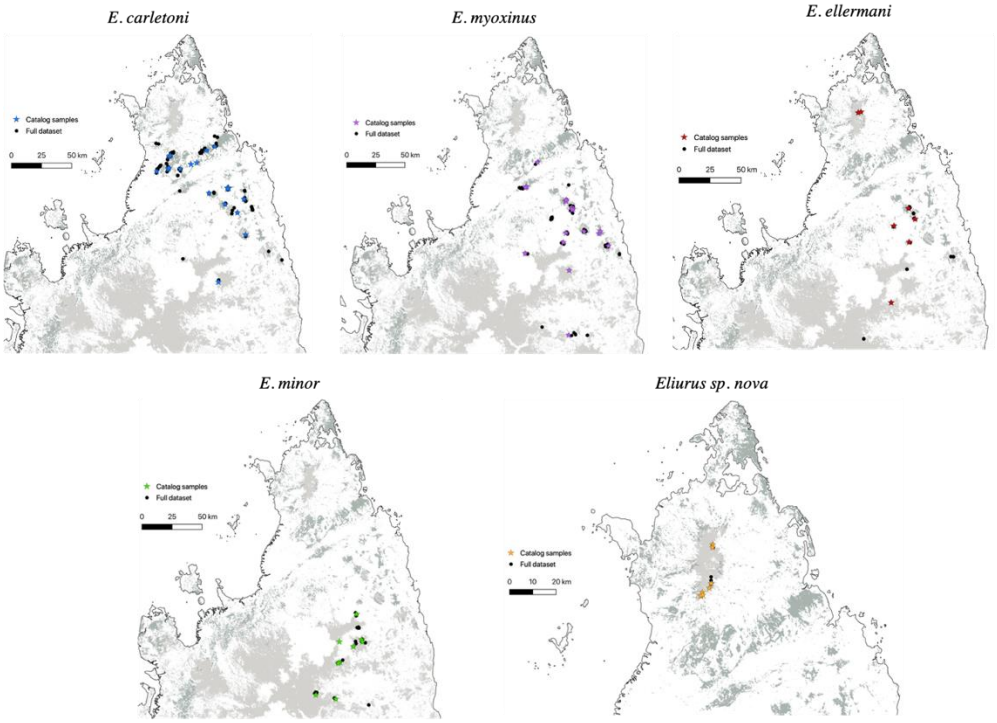


Figure S1. Individuals selected for building the *catalog* in each *Eliurus* species. Black dots indicate the complete dataset for each species: *E. carletoni* (N = 142), *E. myoxinus* (N = 116), *E. ellermani* (N = 26), *E. minor* (N = 38), *Eliurus sp.* nova (N = 13). Coloured stars mark the individuals that have been selected for *catalog* assembly (Tab. S1). Individuals have been selected such to increase the geographic representation of each species, and use individuals with higher genomic coverage.

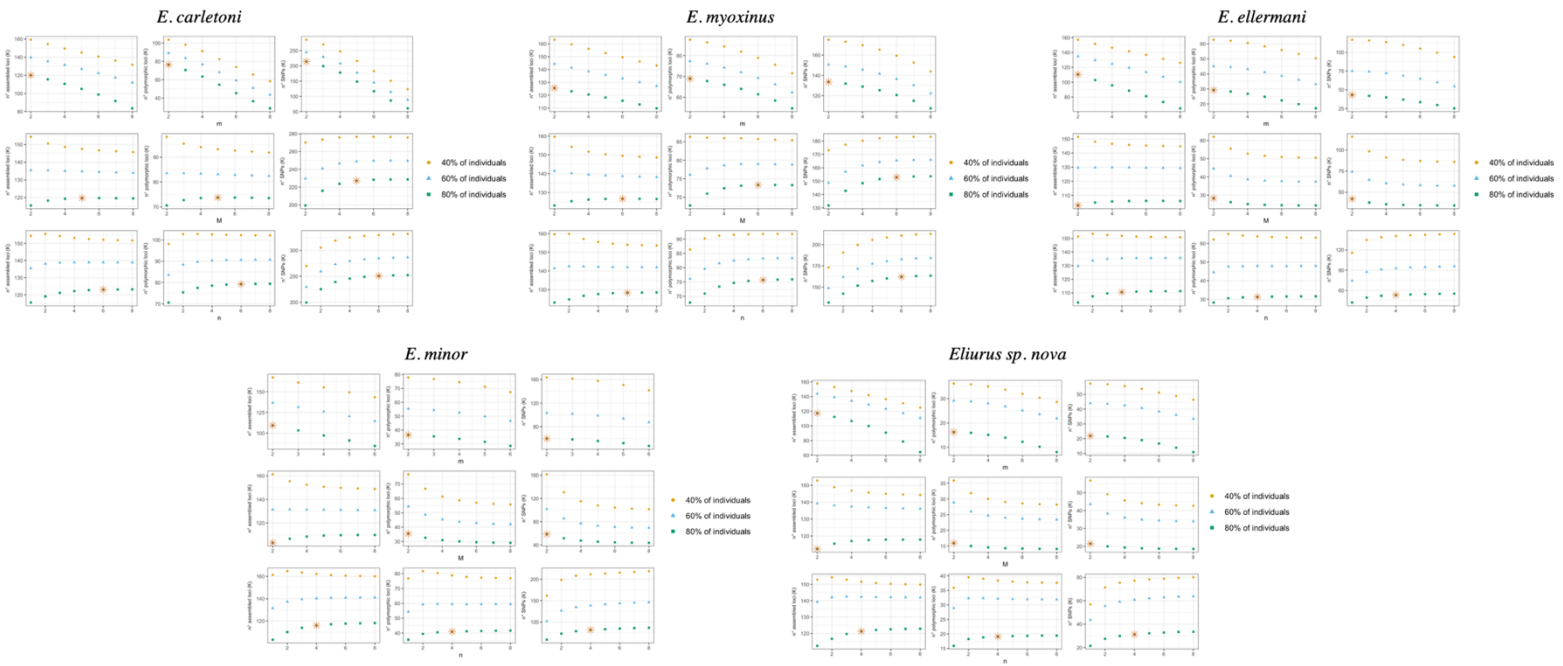


Figure S2. Optimization *de novo* assembly of RAD-seq loci using STACKS for *E. carletoni*, *E. myoxinus*, *E. minor*, *E. ellermani* and *Eliurus sp.* nova. Iterating values of i) the minimum number of raw reads required to form a stack (*m*; First row), ii) the distance allowed between two stacks (*M*; Second row), and iii) the number of mismatches allowed between stacks during construction of the catalog (*n*; Third row). First column: number of assembled loci; Second column: number of polymorphic loci; Third column: number of SNPs. The stars indicate the best parameter combination, that is *m* = 2; *M* = 5 and *n* = 6 for *E. carletoni*, *m* = 2; *M* = 6 and *n* = 6 for *E. myoxinus*, *m* = 2; *M* = 2 and *n* = 4 for *E. minor*, *m* = 2; *M* = 2 and *n* = 4 for *E. ellermani*, *m* = 2; *M* = 2 and *n* = 4 for *Eliurus sp.* nova.

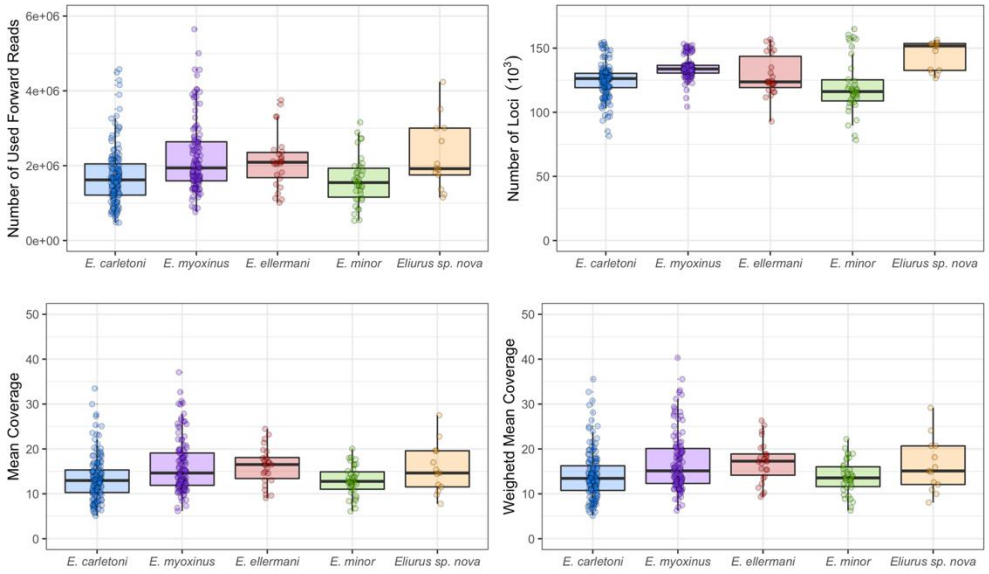


Figure S3. Summary statistics of the assembled RAD-seq loci across *Eliurus* species. Top-left panel: the number of forward reads used for genotyping each individual; Top-right panel: the number of loci called in each individual; Low-left panel: the individual mean coverage; Low-right panel: the individual weighted mean coverage, obtained by giving more weight to loci that are more represented across individuals (see also Table 1).

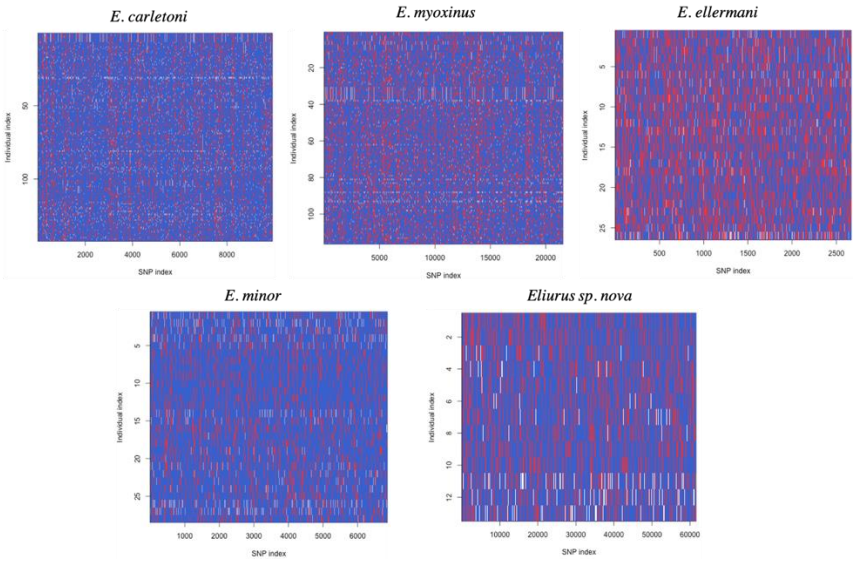


Figure S4. Distribution of SNPs genotypes for the final dataset of each *Eliurus* species. Rows indicates individuals, columns identify single SNPs. Blue and red represent allele type 1 and 2, whereas white colour represents missing data. Overall, the results suggest little missing data across datasets.

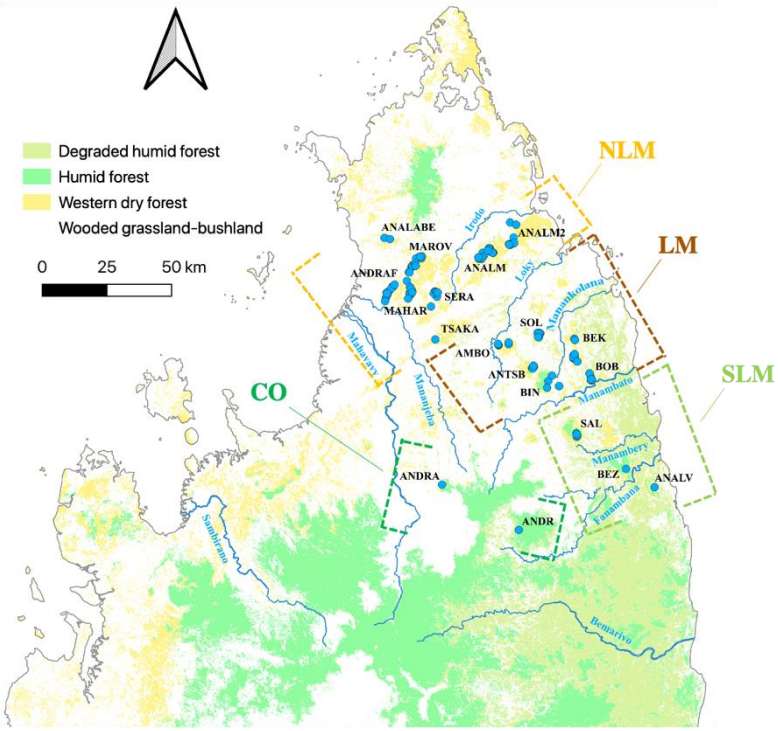


Figure S5. Sampling distribution *Eliurus carletoni*.

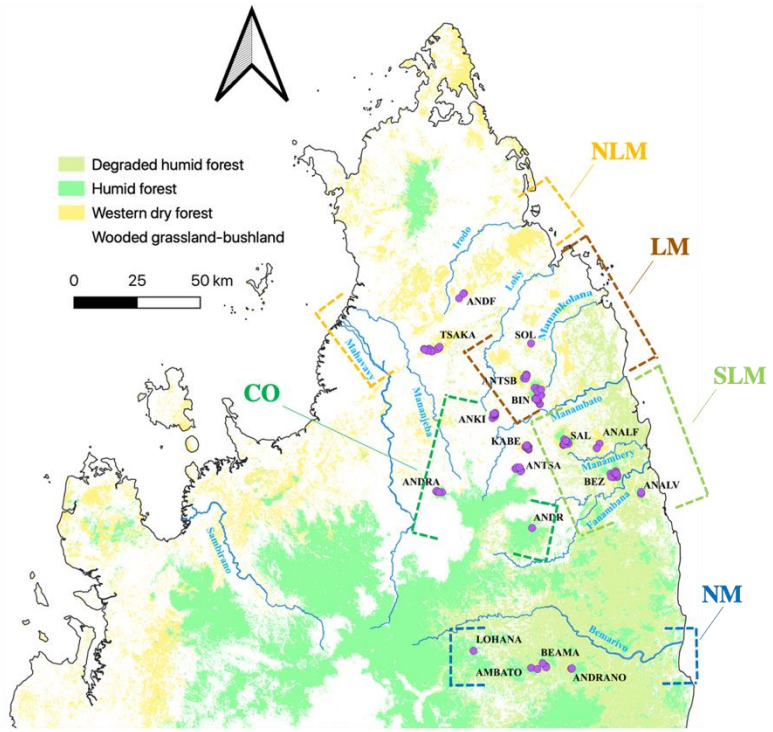


Figure S6. Sampling distribution *Eliurus myoxinus*.

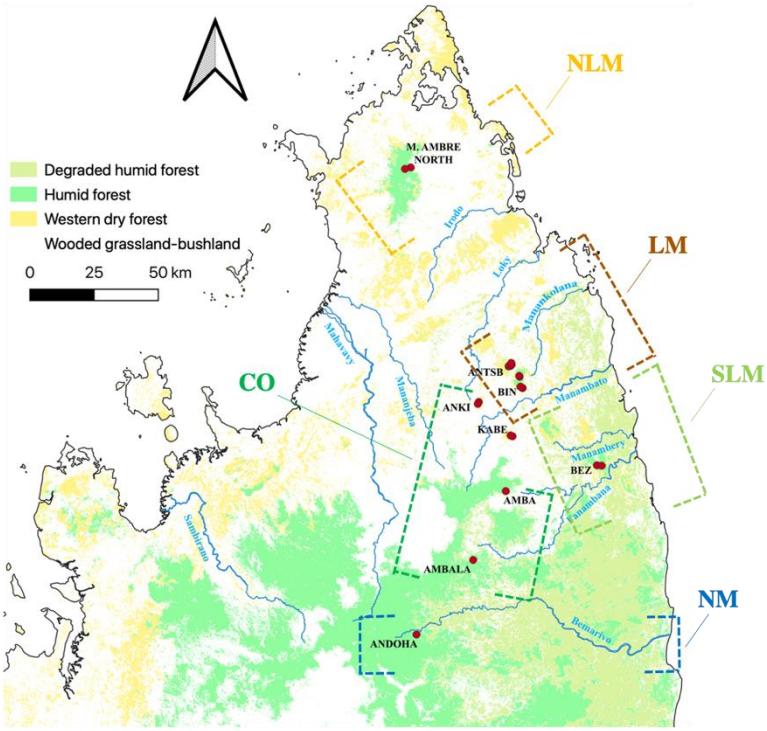


Figure S7. Sampling distribution *Eliurus ellermani.*

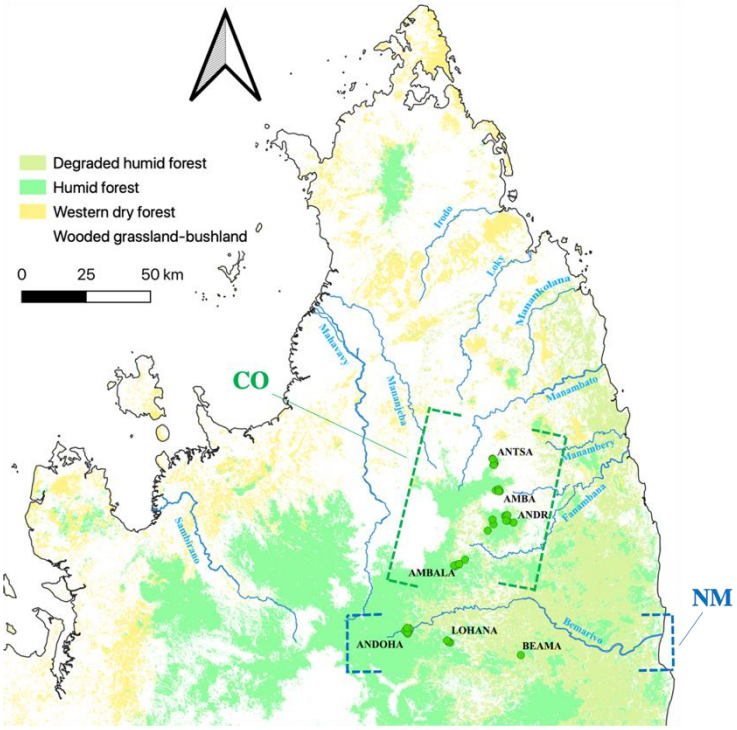


Figure S8. Sampling distribution *Eliurus minor.*

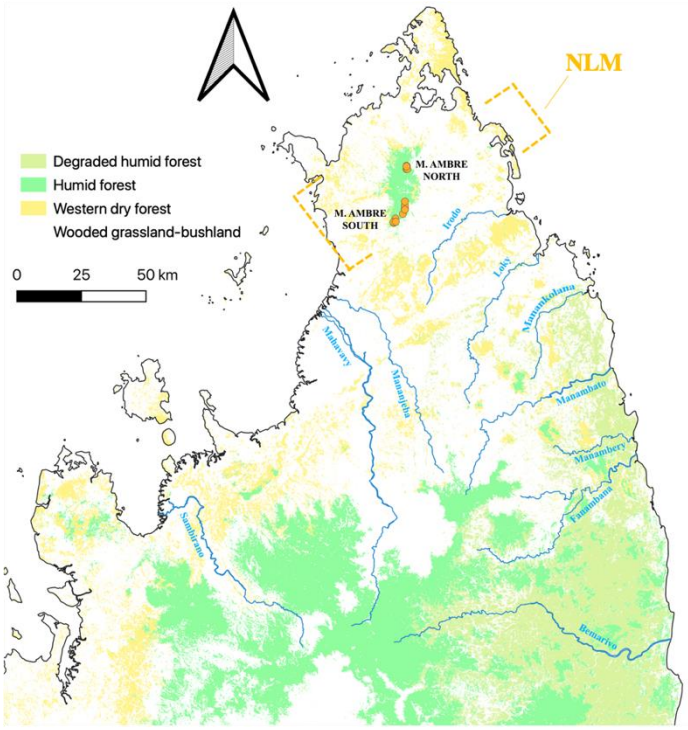


Figure S9. Sampling distribution *Eliurus sp. nova.*

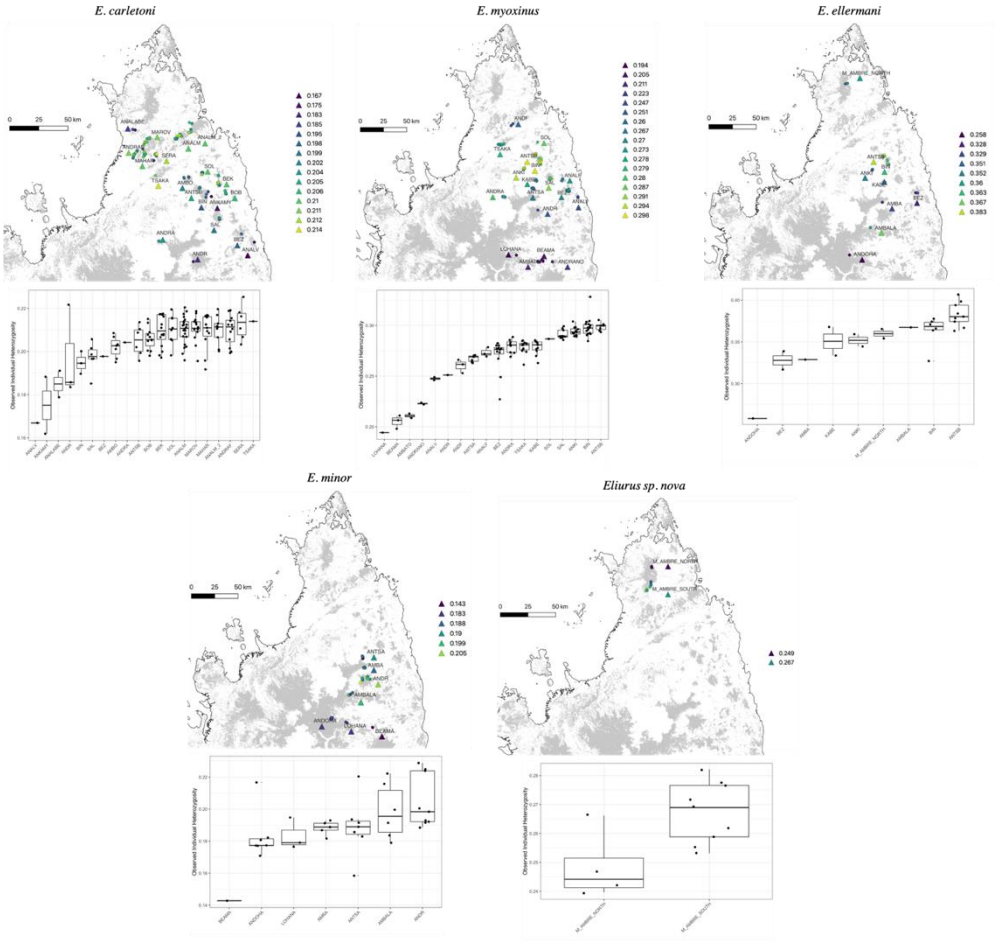


Figure S10. Genomic diversity of *Eliurus* species in northern Madagascar. Genomic diversity refers to individual (circles) or population (triangles) observed heterozygosity. Box plots shows the distribution of individuals observed heterozygosity within each population (sampling site). In maps, grey colour identifies both dry deciduous and humid forests, whereas white is mostly composed of grassland.

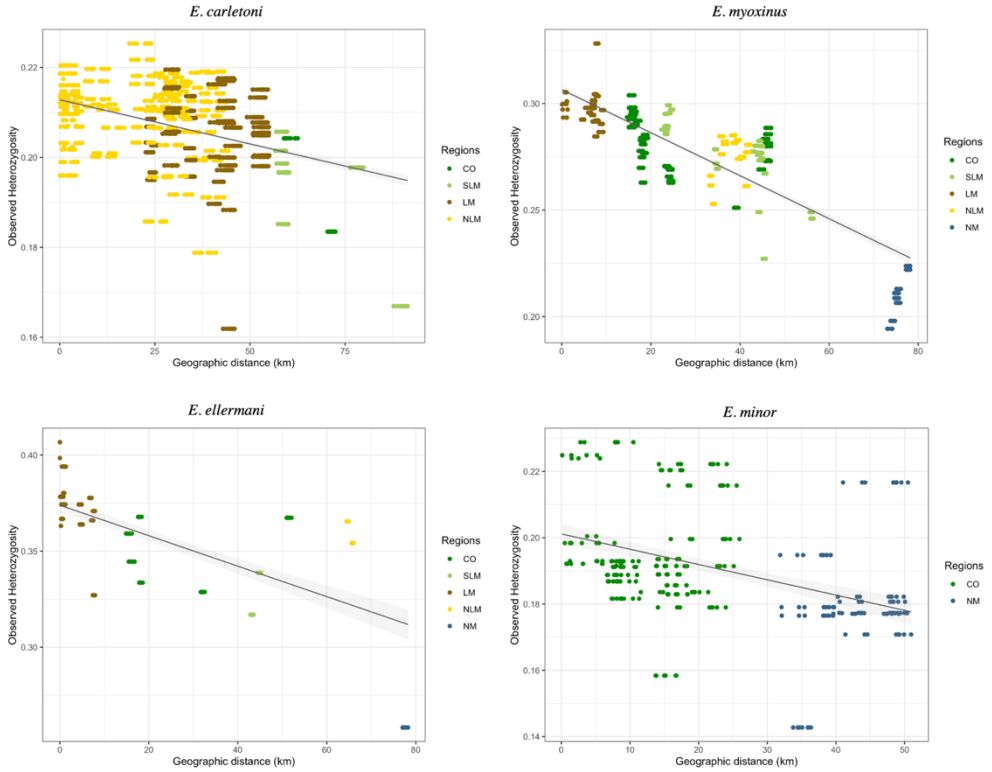


Figure S11. Decline in individual observed heterozygosity and origin of expansion in *E. carletoni*, *E. myoxinus*, *E. ellermani* and *E. minor*. Linear regression between geographic distance and observed heterozygosity from the site with highest correlation, namely ANALM for *E. carletoni* (Spearman ρ = -0.27, p-value = 0), ANTSB for *E. myoxinus* (Spearman ρ = -0.77, p-value = 0) and *E. ellermani* (Spearman ρ = -0.71, p-value = 0), and ANDR for *E. minor* (Spearman ρ = -0.5, p-value = 0).

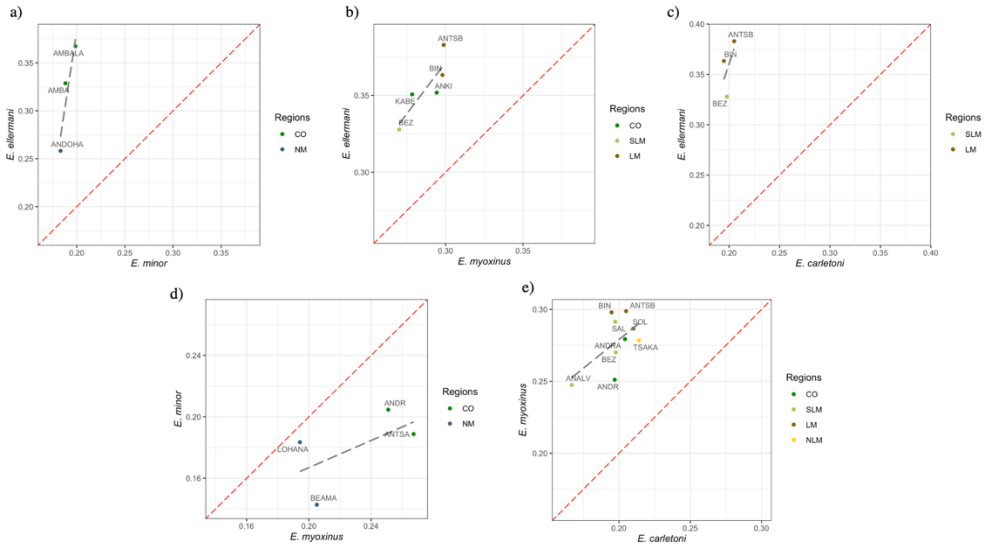


Figure S12. Comparison of population observed heterozygosity between co-occurring *Eliurus* species pairs. Population observed heterozygosity were correlated between pairs of *Eliurus* species, and the significance of such correlation assessed using nonparametric Spearman correlation. a) *E. minor* - *E. ellermani*, Spearman ρ= 1.0, p-value = 0.33; b) *E. myoxinus* - *E. ellermani*, Spearman ρ = 1.0, p-value = 0.016; c) *E. carletoni* - *E. ellermani*, Spearman ρ = 0.5, p-value = 1.0; d) *E. myoxinus* - *E. minor*, Spearman ρ = 0.6, p-value = 0.42; e) *E. carletoni* - *E. myoxinus*, Spearman ρ = 0.3, p-value = 0.44. The results show that only *E. myoxinus* and *E. ellermani* are significantly correlated.

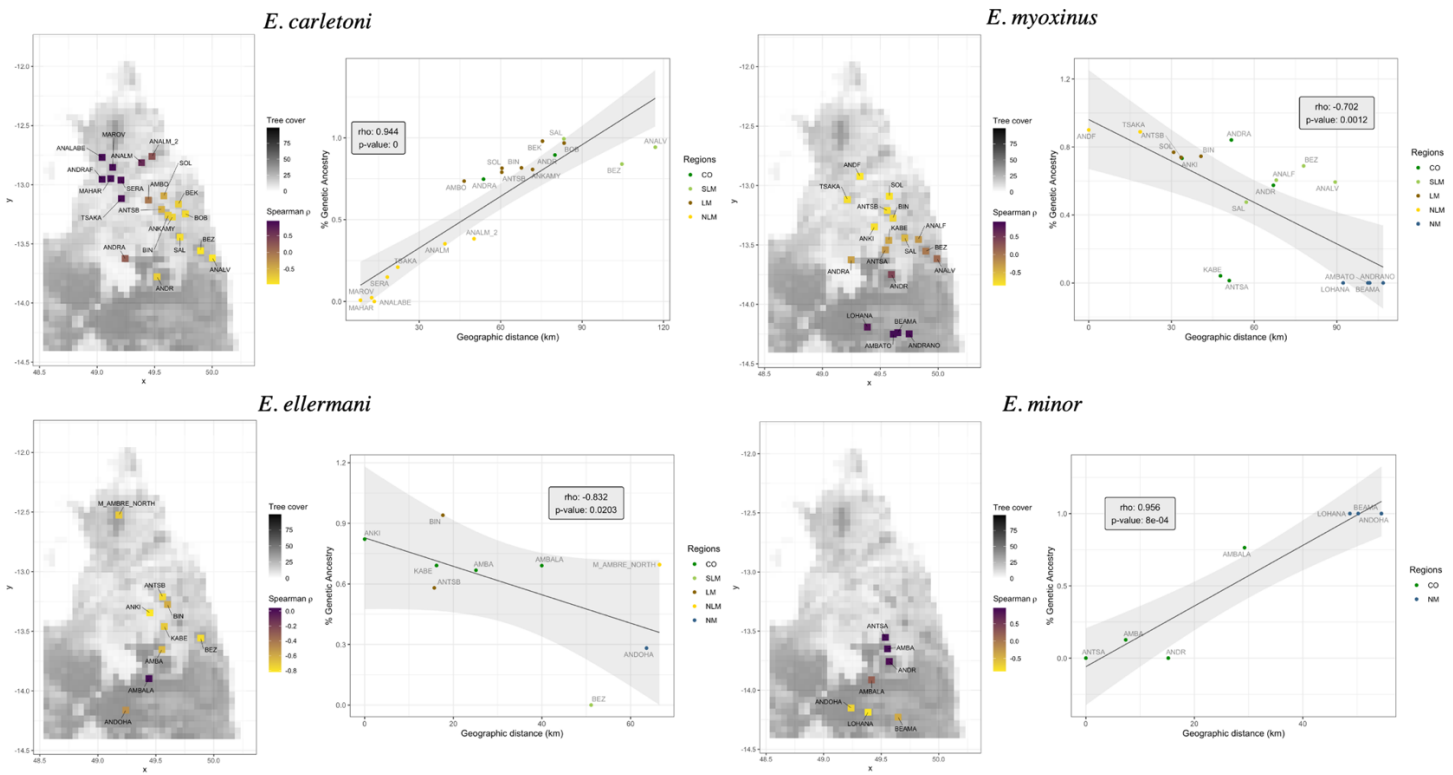


Figure S13. Linear regression of population genetic ancestry with geographical distance in *E. carletoni*, *E. myoxinus*, *E. ellermani* and *E. minor*. Map: squared cells indicate sampling sites, with colour code proportional to the amount of correlation (Pearson) between population genetic ancestry and geographical distance between the focal cell and the remaining sampling sites. Plot: Linear regression between geographic distance and population genetic ancestry from the site with highest correlation, namely MAHAR for *E. carletoni*, ANDF for *E. myoxinus*, ANKI for *E. ellermani* and ANTSA for *E. minor*.

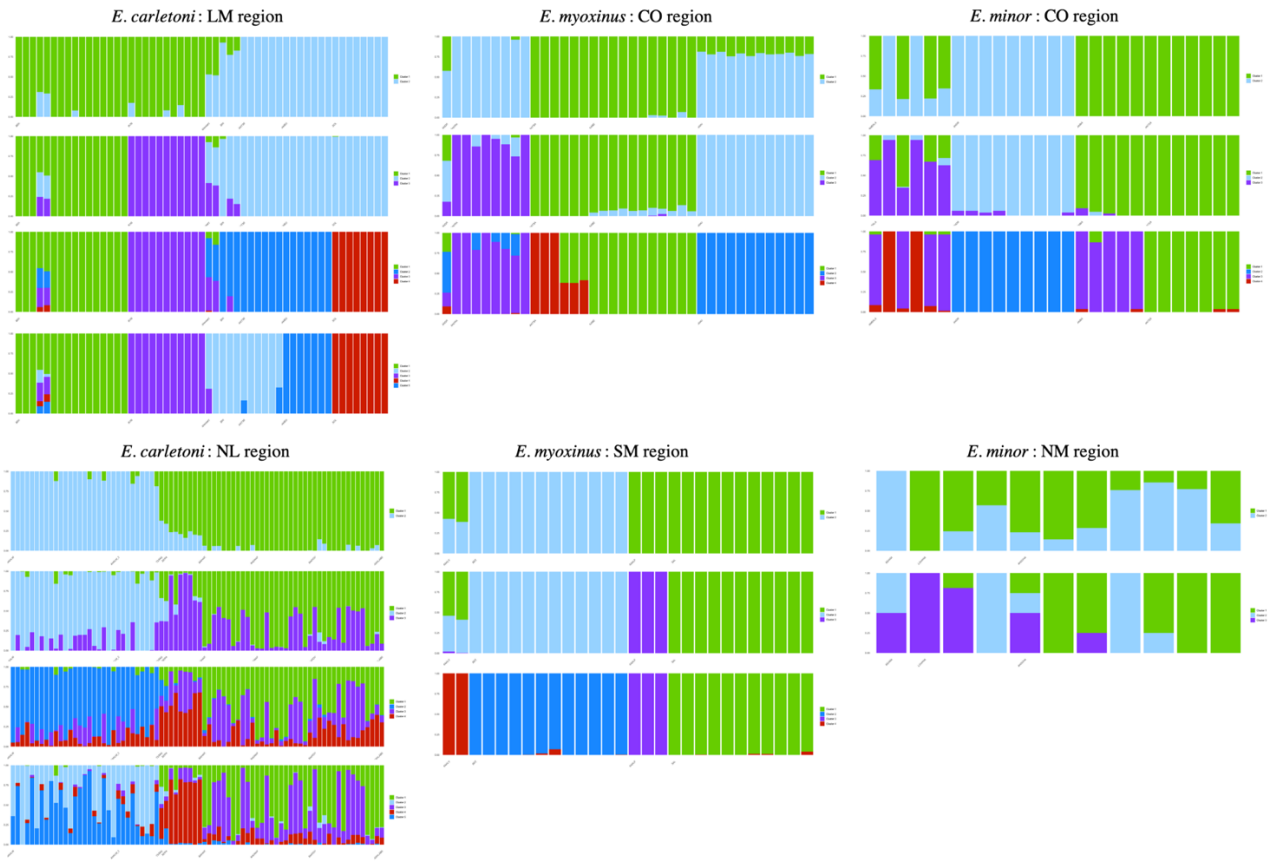


Figure S14. ADMIXTURE analysis within region for Eliurus species with large geographic distribution. Regional ADMIXTURE analysis was possible only for two regions within E. carletoni, E. myoxinus and E. minor distribution. LM: Loky-Manambato; NL: North Loky river; CO: COMATSA; SM: South Manambato; NM: North Marojejy.

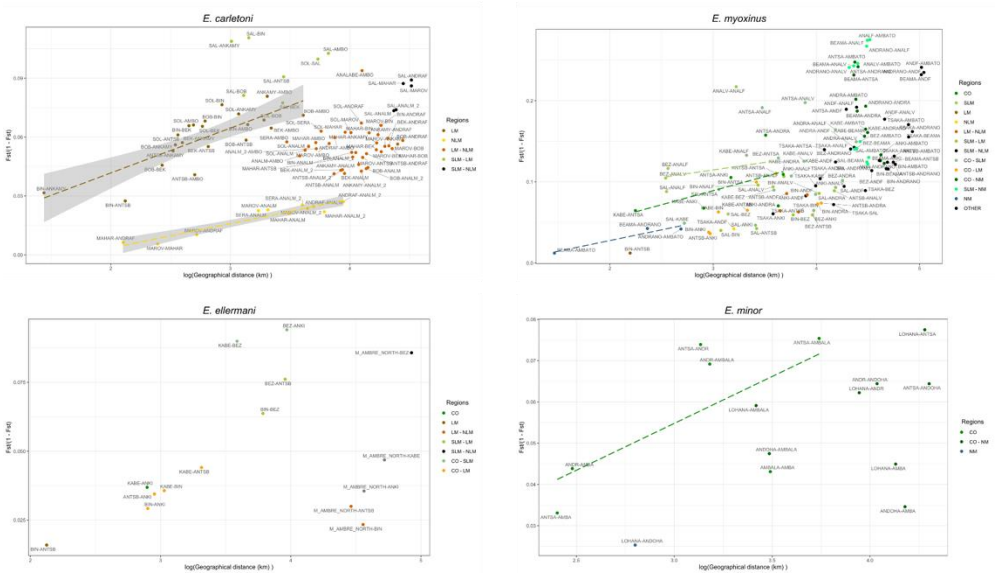


Figure S15. Isolation-by-distance across *Eliurus* species sampled in northern Madagascar. Isolation-by-distance patterns were computed by regressing pairwise genetic distance $F_{st}/(1-F_{st})$ between populations with the logarithm of geographical distance. Colours identify within or between-regions pairwise comparisons.

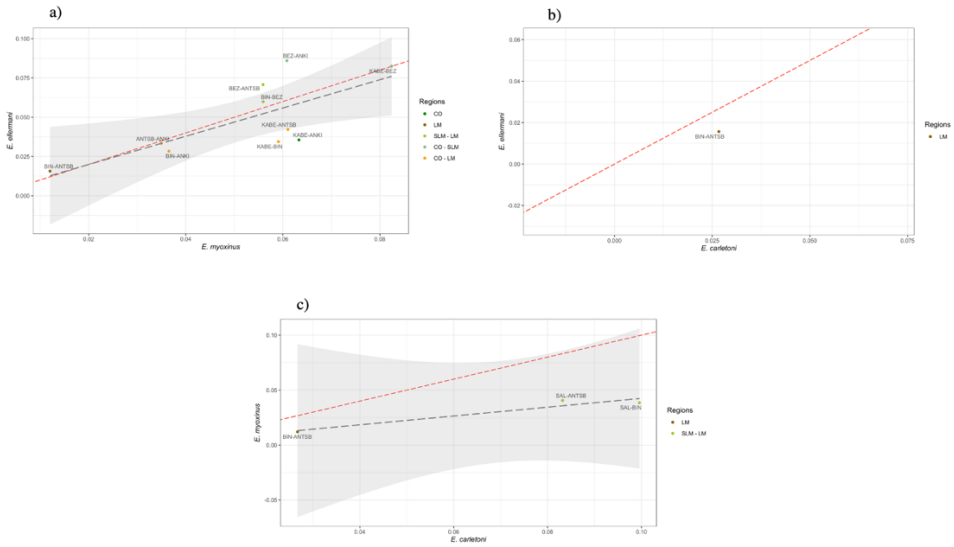


Figure S16. Comparison of population pairwise F_{st} between co-occurring Eliurus species pairs. a) E. myoxinus - E. ellermani: Spearman ρ = 1.0, p-value = 0.017; b) E. carletoni – E. ellermani: Spearman ρ = 0.5, p-value = 1.0; c) E. carletoni – E. myoxinus: Spearman ρ = 0.3, p-value = 0.44. Colours in dots refer to within-region or between-regions comparisons. CO: COMATSA region; SLM: Southern Loky-Manambato; LM: Loky-Manambato; NLM: North Loky-Manambato.

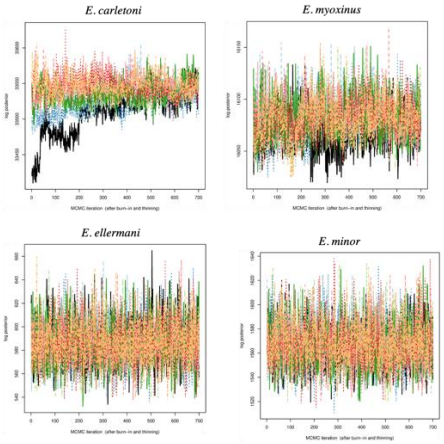


Figure S17. MCMC convergence in EEMS analysis across *Eliurus* species.

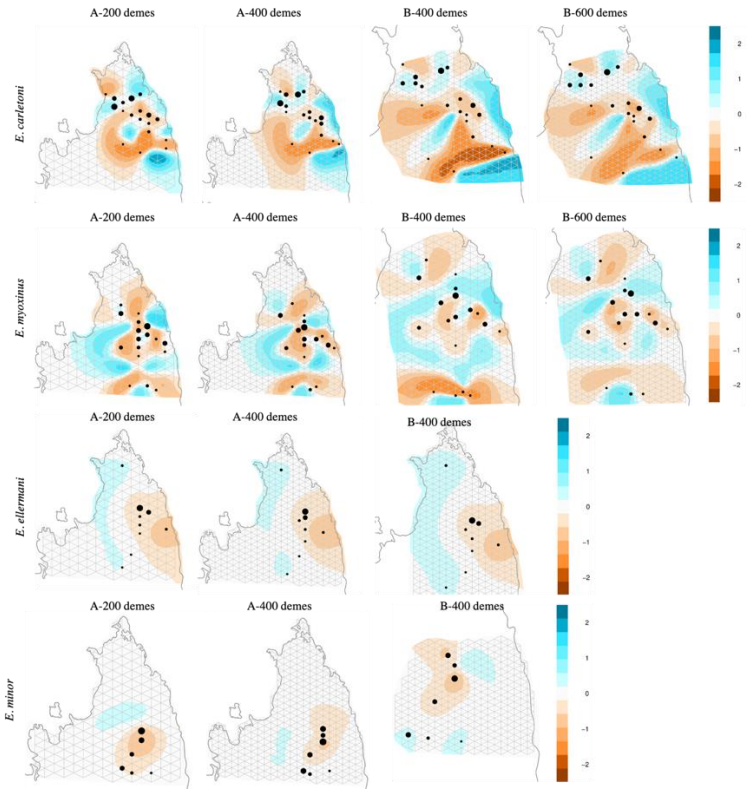


Figure S18. The effect of grid size on EEMS results. First row: *E. carletoni*; Second row: *E. myoxinus*; Third row: *E. ellermani*; Fourth row: *E. minor*. Colours represent relative rates of migration, ranging from low (orange) to high (blue). Type A- consists into a polygon area that is larger than type B-. For the same polygon type, EEMS analyses with different number of demes are performed. The results show no major change on spatial pattern with changes in grid density.

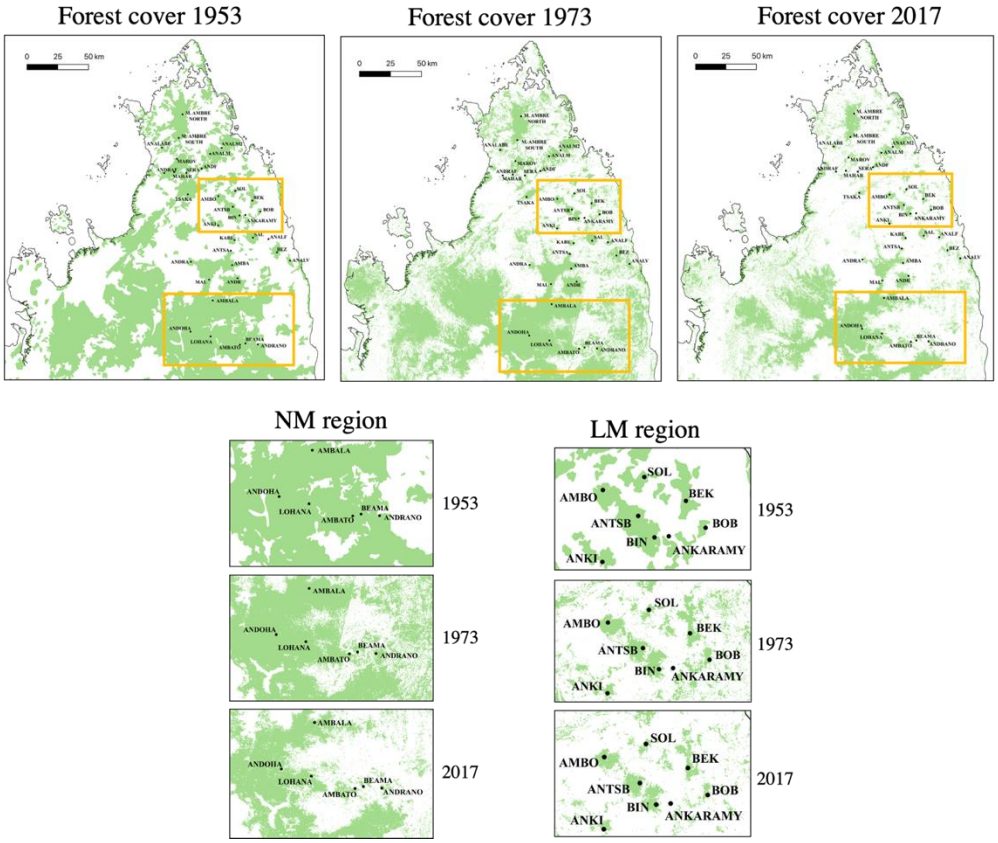


Figure S19. Forest cover in northern Madagascar in 1953, 1973 and 2017. NM: North Marojejy, LM: Loky-Manambato. Source: Viellident et al., 2018.

Table S1. List of individuals used for the construction of the *de-novo* catalog of each *Eliurus* species. A) *E. carletoni*; B) *E. myoxinus*; C) *E. minor*; D) *E. ellermani*; E) *Eliurussp. nova*. Forest: complete forest name; Site: forest abbreviation; ID: identifier of the individual; Nloci: number of genotyped loci; Altitude: altitude of the capture site; NFWReads: number of used forward reads; Coverage: mean coverage; WCoverage: weighted mean coverage calculated by giving more weight to loci that are more represented across the dataset.

a)

Forest	Site	ID	Latitude	Longitude	Altitude (m)	NFWReads	Nloci	Coverage	WCoverage	Species
Mahamasina/Ankarana	MAHAR	F19_2012	-12.95819	49.11944	110	3006660	149372	20.129	21.496	*E. carletoni*
Mahamasina/Ankarana	MAHAR	F26_2012	-12.94759	49.12521	127	2849431	149878	19.012	20.355	*E. carletoni*
Andrafiabe_tenan'ankarana	ANDRAF	F126_2012	-12.97395	49.03363	52	1969656	134537	14.64	15.734	*E. carletoni*
Solaniampilana	SOL	H07_2011	-13.08976	49.58438	175	1917519	142997	13.41	14.223	*E. carletoni*
Antsahabe	ANTSB	H37_2011	-13.21111	49.55832	737	1830037	137435	13.316	14.172	*E. carletoni*
Solaniampilana	SOL	I47_2011	-13.08869	49.58095	110	2061433	142981	14.418	15.279	*E. carletoni*
Ampasimaty	ANALM_2	P36_2012	-12.78281	49.4722	109	2292560	144833	15.829	16.957	*E. carletoni*
Analamerana	ANALM	C28_2012	-12.81092	49.41023	271	1087101	137056	7.932	8.196	*E. carletoni*
Salafaina	SAL	B62_2013	-13.43622	49.71942	380	2071012	142396	14.544	15.494	*E. carletoni*
Solaniampilana	SOL	I22_2011	-13.08855	49.57435	111	2734893	148470	18.421	19.545	*E. carletoni*
Ambohitsitondroina	AMBO	F57_2011	-13.13172	49.43522	132	2947989	148214	19.89	21.148	*E. carletoni*
Andrafiamena	ANDF	B45_2012	-12.91802	49.29342	496	2980611	149457	19.943	21.297	*E. carletoni*
Andrafiamena	ANDF	A11_2012	-12.90433	49.33725	408	5145364	157147	32.742	36.368	*E. carletoni*
Antserasera/Masiamposa	SERA	A42_2012	-12.95253	49.21246	298	4284234	154766	27.682	30.776	*E. carletoni*
Andoharano	ANDR	A46_2012	-13.79328	49.51964	1222	4155229	152396	27.266	29.778	*E. carletoni*
Bekaraoka	BEK	D21_2011	-13.17053	49.70282	150	2191605	144180	15.2	16.17	*E. carletoni*
Marovato/Ankarana	MAROV	H11_2012	-12.86177	49.12588	156	3896453	153865	25.324	28.185	*E. carletoni*
Marovato/Ankarana	MAROV	H7_2012	-12.85455	49.13514	195	4576695	152650	29.982	32.674	*E. carletoni*
Solaniampilana	SOL	K35_2011	-13.09551	49.57943	138	3514444	152504	23.045	24.936	*E. carletoni*
Ankaramy	ANKAMY	K43_2011	-13.27239	49.65454	439	3329347	150178	22.169	23.807	*E. carletoni*

b)

Forest	Site	ID	Latitude	Longitude	Altitude (m)	NFWReads	Nloci	Coverage	WCoverage	Species
Andavakoera	TSAKA	D79_2012	-13.11718	49.23765	474	5646287	152330	37.066	40.261	*E. myoxinus*
Antsahavary	ANTSA	F81_2015	-13.53363	49.54598	1036	2992206	149623	19.998	21.44	*E. myoxinus*
Andavakoera	TSAKA	P72_2012	-13.1085	49.24591	448	5005102	153180	32.675	35.469	*E. myoxinus*
Antsahabe	ANTSB	H19_2011	-13.21448	49.55575	737	2713374	151091	17.959	19.156	*E. myoxinus*
Antsahabe	ANTSB	H26_2011	-13.2163	49.56255	571	3974335	151985	26.15	28.179	*E. myoxinus*
Antsahabe	ANTSB	H49_2011	-13.20529	49.5632	693	3835420	149559	25.645	27.324	*E. myoxinus*
Analafiana	ANALF	B20_2013	-13.46023	49.82661	64	4048709	150599	26.884	28.882	*E. myoxinus*
Bezavona - Ankirendrina	BEZ	C10_2013	-13.55535	49.90294	330	3287358	148296	22.168	23.556	*E. myoxinus*
Anketrakabe	KABE	G6_2015	-13.46135	49.56617	818	4571883	151366	30.204	32.573	*E. myoxinus*
Andraharikely	ANDRA	I10_2015	-13.62444	49.24383	856	3821568	149644	25.538	27.297	*E. myoxinus*
Binara	BIN	J60_2011	-13.27434	49.61935	345	3972754	150323	26.428	28.271	*E. myoxinus*
Andrafiamena	ANDF	B51_2012	-12.91797	49.32937	475	3474441	147004	23.635	25.021	*E. myoxinus*
Analafiana	ANALF	B17_2013	-13.44332	49.83579	40	3851412	147455	26.119	27.738	*E. myoxinus*
Binara	BIN	BE78_2017	-13.262	49.60656	822	4563874	148769	30.678	32.966	*E. myoxinus*
Binara	BIN	BF20_2017	-13.27454	49.60973	596	4417265	146916	30.067	31.979	*E. myoxinus*
Salafaina	SAL	B81_2013	-13.44763	49.70726	669	3660798	147263	24.859	26.368	*E. myoxinus*
Andavakoera	TSAKA	D75_2012	-13.11674	49.23788	474	3655932	146317	24.986	26.655	*E. myoxinus*
Binara	BIN	BF42_2017	-13.28253	49.61072	507	2795339	147532	18.947	20.035	*E. myoxinus*
Ambatomainty	AMBATO	BK100_2018	-14.24838	49.60216	857	3047060	146089	20.858	21.921	*E. myoxinus*
Andoharano	ANDR	E11_2014	-13.74913	49.59581	937	2645416	145232	18.215	19.353	*E. myoxinus*

c)

Forest	Site	ID	Latitude	Longitude	Altitude (m)	NFWReads	Nloci	Coverage	WCoverage	Species
Maladialina	MAL	A69_2014	-13.75956	49.41195	1014	3426518	159873	21.433	23.399	*E. minor*
Andoharano	ANDR	E32_2014	-13.74079	49.58126	1025	2715449	164908	16.466	17.98	*E. minor*
Andoharano	ANDR	E25_2014	-13.74029	49.58227	1025	2623176	158304	16.57	17.865	*E. minor*
Lohananjialava	LOHANA	BK111_2018	-14.18806	49.39257	942	3158713	157708	20.029	22.226	*E. minor*
Andoharano	ANDR	A46_2014	-13.79328	49.51964	1222	2887289	160453	17.995	20.198	*E. minor*
Andoharano	ANDR	E17_2014	-13.75705	49.58968	991	1999289	155937	12.821	13.763	*E. minor*
Antsahavary	ANTSA	F37_2015	-13.55122	49.53839	1049	1986238	145542	13.647	14.921	*E. minor*
Ambalasoa	AMBALA	B72_2014	-13.91378	49.41881	1075	1820287	151162	12.042	13.309	*E. minor*
Andohanibemarivo	ANDOHA	BK143_2018	-14.16067	49.23928	990	2736654	157067	17.423	18.844	*E. minor*
Ambalasoa	AMBALA	B92_2014	-13.92058	49.4065	1148	1672877	133997	12.484	13.596	*E. minor*

d)

Forest	Site	ID	Latitude	Longitude	Altitude (m)	NFWReads	Nloci	Coverage	WCoverage	Species
Binara	BIN	K71_2011	-13.28993	49.60661	767	3753097	153373	24.47	26.257	*E. ellermani*
Montagne d'Ambre	M_AMBRE_NORTH	BC74_2017	-12.52726	49.17227	1118	3303123	148411	22.257	23.955	*E. ellermani*
Antsahabe	ANTSB	H21_2011	-13.21329	49.5569	737	3640943	156906	23.205	25.253	*E. ellermani*
Binara	BIN	K55_2011	-13.28992	49.60658	767	2497097	146505	17.044	18.221	*E. ellermani*
Ankinjanala	ANKI	G52_2015	-13.34159	49.45042	1084	2311782	147705	15.651	16.767	*E. ellermani*
Montagne d'Ambre	M_AMBRE_NORTH	BD44_2017	-12.52175	49.19254	938	2187080	150403	14.541	15.4	*E. ellermani*
Anketrakabe	KABE	G9_2015	-13.45726	49.56654	818	1883826	144687	13.02	13.611	*E. ellermani*
Ambalasoa	AMBALA	B25_2014	-13.89701	49.44081	918	1756733	134933	13.019	13.757	*E. ellermani*
Antsahabe	ANTSB	H22_2011	-13.21265	49.55706	737	1415416	130676	10.831	11.342	*E. ellermani*
Antsahabe	ANTSB	H41_2011	-13.21265	49.55706	737	1249548	127564	9.795	10.247	*E. ellermani*

e)

Forest	Site	ID	Latitude	Longitude	Altitude (m)	NFWReads	Nloci	Coverage	WCoverage	Species
Montagne d'Ambre	M_AMBRE_NORTH	BD29_2017	-12.5082	49.18153	923	1149823	147839	7.778	8.017	*Eliurus* sp. nova
Montagne d'Ambre	M_AMBRE_NORTH	BC51_2017	-12.51994	49.18104	1050	4236752	154050	27.502	28.875	*Eliurus* sp. nova
Montagne d'Ambre	M_AMBRE_SOUTH	BE35_2017	-12.67702	49.17084	1002	1822201	151723	12.01	12.466	*Eliurus* sp. nova
Montagne d'Ambre	M_AMBRE_SOUTH	BG72_2017	-12.69357	49.14442	946	2657125	156392	16.99	17.98	*Eliurus* sp. nova
Montagne d'Ambre	M_AMBRE_SOUTH	BD68_2017	-12.65919	49.17768	1006	3512318	154213	22.776	23.854	*Eliurus* sp. nova
Montagne d'Ambre	M_AMBRE_SOUTH	BG77_2017	-12.70927	49.13926	812	3008579	153590	19.588	20.471	*Eliurus* sp. nova
Montagne d'Ambre	M_AMBRE_SOUTH	BH10_2017	-12.70601	49.13793	903	1751145	151693	11.544	11.967	*Eliurus* sp. nova
Montagne d'Ambre	M_AMBRE_SOUTH	BG61_2017	-12.70442	49.14583	930	3004007	152987	19.636	20.512	*Eliurus* sp. nova

Chapter 5

Genetic diversity of mouse lemurs and tufted-tailed rats in northern Madagascar: new species and insights on past forest cover changes

5 Genetic and morphological diversity of mouse lemurs (*Microcebus* spp.) in northern Madagascar: The discovery of a putative new species?

Gabriele Maria Sgarlata[1*], Jordi Salmona[1,2*], Barbara Le Pors[1], Emmanuel Rasolondraibe[3], Fabien Jan[1], Tantely Ralantoharijaona[3], Ando Rakotonanahary[3], Jacquis Randriamaroson[3], Adam Joseph Marques[1], Isa Aleixo Pais[1,5,6], Tiago Zoeten[1], Dhurham Said Ali Ousseni[3], Simon Benjamin Knoop[1], Helena Teixeira[1,7], Vivien Gabillaud[1], Alexandra Miller[1,8], Mohamed Thani Ibouroi[3,4,9], Solofonirina Rasoloharijaona[3], John Rigobert Zaonarivelo[4], Nicole Volasoa Andriaholinirina[3], Lounès Chikhi[1,2]

1- Instituto Gulbenkian de Ciênca, Rua da Quinta Grande, 6, P-2780-156 Oeiras, Portugal
2- Laboratoire Évolution & Diversité Biologique (EDB UMR 5174), Université de Toulouse Midi-Pyrénées, CNRS, IRD, UPS. 118 route de Narbonne, Bât. 4R1, 31062 Toulouse cedex 9, France
3- Faculté des Sciences, Université de Mahajanga, BP 652 401, Mahajanga, Madagascar
4- Département des Sciences de la Nature et de l'Environnement, Université d'Antsiranana, 201 Antsiranana, Madagascar
5- Organisms and Environment Division, Cardiff School of Biosciences, Cardiff University, CF10 3AX, Cardiff, UK
6- Centre for Research in Anthropology (CRIA-FCSH/NOVA), 1069-061 Lisbon, Portugal
7- Institute of Zoology, University of Veterinary Medicine Hanover, Hanover, Germany
8- School of Human Sciences, The University of Western Australia, Crawley, WA 6009
9- Institut des Sciences et Industrie du Vivant et de l'environnement, AgroParisTech, 19 Avenue de Maine 75015 Paris, France

* Shared first co-authors

AUTHOR CONTRIBUTIONS: GMS, LC and JS designed the project. GMS collected part of the samples, performed DNA extraction and analyzed genetic and morphological data. BLP and FJ collected most of the samples. JS collected part of the samples and generated part of the genetic data. ER, TR, AR, JR contributed in the collection of samples. AJM generated part of

the genetic data. IAP extracted DNA from tissue samples and generated most of the mtDNA data. TZ, DSAO, SBK, HT, VG, AM, MTI contributed to the collection of samples. SR, JRZ and NVA contributed in the organization of the fieldwork. GMS wrote the manuscript and, LC and JS revised it.

Highlights

- Extend the known distribution of three endangered mouse lemurs' species (genus *Microcebus*)
- Report the discovery of a new putative mouse lemur species
- Report a new rare case of co-occurrence between two mouse lemur species
- Identify a simple morphological trait that discriminates between two mouse lemur species
- Provide genetic evidence suggesting that humid forests were covering present-day dry northern Madagascar, at some point in the recent past

5.1 Abstract

Tropical forests harbor extremely high levels of biological diversity and are quickly disappearing. Despite the increasingly recognized high rate of habitat loss, it is expected that new species will be discovered as more effort is put to document tropical biodiversity. Exploring under-studied regions is particularly urgent if we consider the rapid changes in habitat due to anthropogenic activities. Madagascar is known for its extraordinary biological diversity and endemicity. It is also threatened by habitat loss and fragmentation. It holds more than 100 endemic primate species (lemurs). Among these, *Microcebus* (mouse lemurs) is one of the more diverse genera. We sampled mouse lemurs from several sites across northern Madagascar, including forests never sampled before. We obtained morphological data from 99 *Microcebus* individuals; we extracted DNA from tissue samples of 42 individuals and amplified two mitochondrial loci (cytb and cox2) commonly used for species identification. Our findings update the distribution of three species (*Microcebus tavaratra*, *Microcebus arnholdi*, and *Microcebus mamiratra*), including a major increase in the distribution area of *M. arnholdi*. We also report the discovery of a new *Microcebus* lineage genetically related to *M. arnholdi*. Several complementary approaches suggest that the newly identified *Microcebus* lineage might correspond to a new putative species, to be confirmed or rejected with additional data. In addition, morphological analyses showed (a) clear phenotypic differences between *M. tavaratra* and *M. arnholdi*, but no clear differences between the new *Microcebus* lineage and the sister species *M. arnholdi*; and (b) a significant correlation between climatic variables and morphology, suggesting a possible relationship between species identity, morphology, and environment. By integrating morphological, climatic, genetic, and spatial data of two northern *Microcebus* species, we show that the spatial distribution of forest- dwelling species may be used as a proxy to reconstruct the past spatial changes in forest cover and vegetation type.

5.2 Introduction

Tropical forests hold an extraordinary amount of biological diversity and new species and taxa continue to be discovered at a high rate (Burgin, Colella, Kahn, & Upham, 2018). Nevertheless, tropical ecosystems are currently undergoing rapid environmental changes, mainly due to human activities (Venter et al., 2016), which pose a serious and concrete risk to the persistence of world biodiversity.

Madagascar is a recognized biodiversity hotspot (Myers, Mittermeier, Mittermeier, Fonseca, & Kent, 2000), whose forest-dwelling species are increasingly threatened by extensive and ongoing habitat loss and fragmentation (Harper, Steininger, Tucker, Juhn, & Hawkins, 2007; Schwitzer et al., 2014; Waeber, Schuurman, & Wilmé, 2018). This threat is especially serious for the species that are distributed in restricted ranges (i.e., microendemics), a category that includes a high percentage of Malagasy taxa (Wilmé, Goodman, & Ganzhorn, 2006).

In the present study, we focus on mouse lemurs (genus *Microcebus*), a small-sized nocturnal primate group that occurs in a wide variety of habitat types across Madagascar (Kappeler & Rasoloarison, 2003). The number of recognized Microcebus species has greatly increased from two species, more than 20 years ago, to more than 20 putative species today (Ross and Kappeler, 2006; Andriantompohavana et al., 2006; Hotaling et al., 2016; Louis et al., 2006, 2008; Olivieri et al., 2007; Radespiel et al., 2008, 2012; Rasoloarison, Weisrock, Yoder, Rakotondravony, & Kappeler, 2013). Although earlier studies applied morphological data as criteria to identify new *Microcebus* species (e.g., Rasoloarison, Goodman, & Ganzhorn, 2000; Zimmermann, Cepok, Rakotoarison, Zietemann, & Radespiel, 1998), more recently the use of genetic data has played a major role in the discovery of new *Microcebus* taxa in Madagascar. Genetic data, 15–20 years ago, in

particularmitochondrial DNA (mtDNA), appeared to supersede morphological data, as they were not always sufficient to discriminate *Microcebus* species (Yoder, Burns, & Génin, 2002; Yoder et al., 2005). More recently, Weisrock et al. (2010) used mtDNA and four nuclear loci to provide a more robust estimate of *Microcebus* species delimitations. Although the effective population size of mtDNA is smaller than that of nuclear DNA, Weisrock et al. (2010) observed a noticeable concordance between mtDNA lineages and patterns of nuclear genetic structure, hence suggesting that mtDNA was often sufficient to identify different species. In the most recent comprehensive study on *Microcebus* phylogenetics, Hotaling et al. (2016) used a multilocus genetic dataset (mtDNA and nuclear loci) and applied a Bayesian approach based on the multispecies coalescent (MSC) model (Rannala & Yang, 2003; Yang, 2002) to assess the validity of species delimitation. However, Sukumaran and Knowles (2017) have shown that the MSC model cannot always separate species divergence from intraspecies population structure, and they concluded that a rigorous approach able to distinguish between population- and species-level structure still needs to be developed. Altogether, this suggests that the identification of species should not rely only on genetic or genomic data but should integrate other types of data, including morphological or behavioral data, when available.

In this work, we present the results of surveys across northern Madagascar, where five *Microcebus* species have been described: *Microcebus arnholdi*, *Microcebus tavaratra*, *Microcebus mamiratra*, *Microcebus margotmarshae*, and *Microcebus sambiranensis*. We collected Microcebus samples from several forest fragments for which nearly no information is currently available. We obtained morphological data from 99 new Microcebus individuals, and we genotyped 42 individuals for two mtDNA loci (cytb and cox2) commonly used for species identification and representing currently the main reference for species identification in the *Microcebus* genus (see however, Louis & Lei, 2016 for a recent use of

mitogenome for 23 species). Mitochondrial genetic data were used to construct phylogenetic trees using both maximum-likelihood (ML) and Bayesian approaches. We used the MSC model to test species delimitation hypotheses and compare our results with *Microcebus* species previously defined or confirmed using this approach (see Hotaling et al., 2016). Considering the limitations of the MSC model in distinguishing population structure from species divergence (Sukumaran & Knowles, 2017), we also used an ad hoc analysis that compares phylogenetic and geographic distances within and between sister-groups.

Most *Microcebus* species have been described based on the phylogenetic species concept, which considers genetic monophyly as evidence for species delimitation (Cracraft, 1983). However, addi- tional data, such as morphology and climate, can provide further lines of evidence to support species delimitation (De Queiroz, 2005, 2007) by identifying clear morphological differences and habitat prefer- ences among the hypothesized species. Therefore, we used georeferenced climatic and morphological data of the sampled individuals (a) to assess whether habitat preferences and phenotypic differences could be detected among the four *Microcebus* taxa studied here; and (b) to quantify the contribution of the climatic factors measured on morphological variability. According to the “recent divergence” hypothesis, we expect that taxa genetically more divergent will show stronger morphological and habitat differences compared to taxa showing lower genetic divergence (see, for instance, Struck et al., 2018; Chenuil et al., 2019).

Finally, we consider (a) the “niche conservatism” hypothesis, defined as the tendency of species to retain aspects of their fundamental niche over time (Wiens et al., 2010); and (b) the fact that *Microcebus* spp. are mostly forest-dependent species, to qualitatively reconstruct the forest cover changes that occurred in northern Madagascar during the late Quaternary. In

particular, we integrate information on habitat preferences, morphological differences, and geographical location of two northern mouse lemur's species typically associated with either dry (*M. tavaratra*) or humid (*M. arnholdi*) habitat and discuss the biogeographical changes that could have led to their present fragmented distribution.

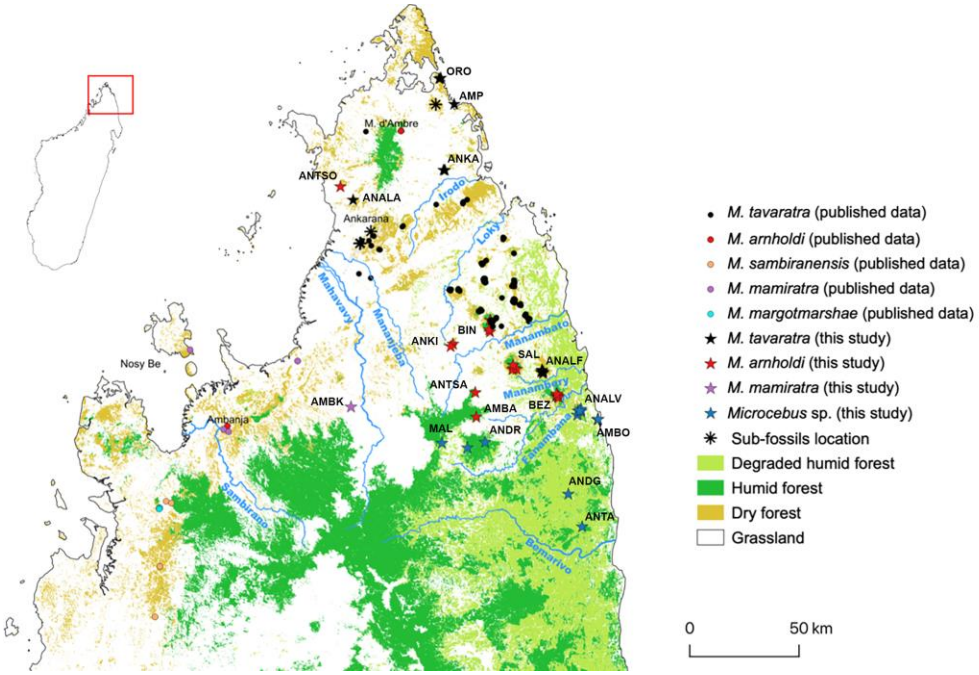


Figure 1. Sampling distribution of *Microcebus* spp. across northern Madagascar. Black labels represent forest fragments surveyed between 2010 and 2017. AMBA, Ambaliha; AMBK, Ambakirano; AMBO, Ambohitrandrina; AMP, Ampiho; ANALA, Analabe; ANALF, Analafiana; ANALV, Analalava; ANDG, Andrangotra; ANDR, Andravory; ANKA, Ankarongana; ANKI, Ankinjanala; ANTA, Antananivo; ANTSA, Antsahavary; ANTSO, Antsoroby; BEZ, Bezavona; BIN, Binara; MAL, Maladialina; ORO, Oronjia; SAL, Salafaina. Stars indicate the individuals sampled in the present study. Circles indicate published data. Asterisks indicate subfossil localities for *Indri indri*, *Prolemur simus* and *Propithecus* cf. *diadema*.

5.3 Material and Methods

5.3.1 Study species

In the present study, we focus on three of the five *Microcebus* species occurring in northern Madagascar (Figure 1): *M. arnholdi*, *M. tavaratra* and *M. mamiratra*.

M. arnholdi (head-body length: 10–12.6 cm) has been described using two mtDNA loci from a population sampled at Montagne d'Ambre (AMBRE), northern Madagascar (Figure 1; Louis et al., 2008). Little is known about this species, except that it occurs in humid habitats of northern Madagascar (Mittermeier et al., 2010). It was thought to be exclusively present in AMBRE until Weisrock et al. (2010) found that an individual sampled at Ambanja, more than 150 km south of AMBRE (Figure 1), was genetically clustering with individuals from AMBRE. *M. arnholdi* is currently classified as Endangered (Andriaholinirina et al., 2014).

M. tavaratra (head-body length: 12–14 cm) inhabits dry deciduous and transition forests of northern Madagascar, between 50–670 m of altitude (Mittermeier et al., 2010; Rasoloarison et al., 2000; Sgarlata et al., 2018). Originally described from morphological and genetic data by Rasoloarison et al. (2000) and Yoder et al. (2000), it is currently known to occur in the regions of the Ankarana, Analamerana, Andavakoera, Andrafiamena, and Loky-Manambato, and probably in the lowland forests of AMBRE (Figure 1). *M. tavaratra* is classified as Vulnerable (Andriaholinirina et al., 2014).

M. mamiratra (head-body length: 10.5–12.7cm) occurs in the humid primary and secondary forests of the small Lokobe Strict Nature Reserve of Nosy Be Island (Andriantompohavana et al., 2006; Olivieri et al., 2007). It was first described by Andriantompohavana et al. (2006) and later independently reported synonymous with *Microcebus lokobensis* by Olivieri et al., 2007

before being recognized as one single species. It is classified as Critically Endangered (Andriaholinirina et al., 2014).

5.3.2 Study sites

Nineteen sites were visited between 2010 and 2017 (Figure 1 and Table 1). Fieldwork varied across years but was typically conducted between May and October to avoid the rainy season. We conducted our sampling in several dry deciduous forest fragments of the far north: the Oronjia (or Orangéa, abbreviated as ORO) new protected area (NAP for Nouvelle Aire Protégée), the Ampiho (AMP) forest part of the Ambodivahibe NAP, the Ankarongana forest (ANKA), the Analabe (ANALA), and Antsoroby forests (ANTSO). We collected samples from one dry forest located south of the Manambato River, Analafiana (ANALF), and a littoral forest south of Fanambana River, Ambohitrandrina (AMBO; Figure 1). We also surveyed several humid forest fragments: the Binara (BIN) mountain forest located within the Loky-Manambato NAP; the Salafaina and Bezavona-Ankirendrina forests, located between the Manambato and Fanambana Rivers; the Ankinjanala forest located at the headwater of the Loky River; the Antsahavary and Ambaliha (AMBA) forests within the Corridor of Marojejy-Anjanaharibe Sud-Tsaratanana (COMATSA) NAP; Analalava (ANALV) forest, south of the Fanambana River; the Andravory humid massif between north COMATSA and Fanambana River; the Andrangotra (ANDG) forest, a small fragment close to Antsirabe-Nord village, and the Antananivo (ANTA) forest, part of the Makirovana-Tsihomanaomby forest complex NAP, located south of the Fanambana River. We also visited a small forest fragment close to the Ambakirano village, at the east of the Mahavavy River (AMBK).

Table 1. Collecting sites and genetic data of the *Microcebus* individuals sampled across northern Madagascar.

Forest	Abbreviation	Year	N	cytb	cox2	Morpho (N)	Altitude (m)	Latitude	Longitude	Forest type	Clade
Oronjia	ORO	2015	2	1	1	2	15, 82	−12.255	49.378	D	*Microcebus tavaratra*
Ampiho	AMP	2017	1	1	–	1	9	−12.361	49.440	D	*M. tavaratra*
Ankarongana	ANKA	2013	2	2	2	2	194, 254	−12.634	49.403	D	*M. tavaratra*
Analabe	ANALA	2017	1	1	–	10	101	−12.763	49.021	D	*M. tavaratra*
Analafiana	ANALF	2013	5	5	5	5	41–90	−13.456	49.831	D	*M. tavaratra*
Antsoroby	ANTSO	2017	1	1	–	1	56	−12.709	48.966	D	*Microcebus arnholdi*
Binara	BIN	2010	3	3	3	27	526–781	−13.278	49.606	T-H	*M. arnholdi*
Salafaina	SAL	2013	4	4	4	4	309–653	−13.444	49.711	T-H	*M. arnholdi*
Bezavona	BEZ	2013	5	5	5	5	231–600	−13.555	49.897	T-H	*M. arnholdi*
Ankinjanala	ANKI	2015	2	2	2	1	714–993	−13.355	49.447	T-H	*M. arnholdi*
Antsahavary	ANTSA	2015	1	1	–	1	951	−13.548	49.550	H	*M. arnholdi*
Ambaliha	AMBA	2015	1	1	1	1	972	−13.649	49.556	H	*M. arnholdi*
Analalava	ANALV	2013	5	5	5	5	63–899	−13.617	49.991	T-H	*Microcebus* sp.
Ambohitrandrina	AMBO	2013	3	3	3	3	15–17	−13.649	50.069	D	*Microcebus* sp.
Andrangotra	ANDG	2015	1	1	1	1	252	−13.959	49.951	H	*Microcebus* sp.
Maladialina	MAL	2014	1	1	1	1	986	−13.759	49.413	H	*Microcebus* sp.
Andravory	ANDR	2014	2	2	2	2	899, 954	−13.764	49.558	H	*Microcebus* sp.
Antananivo	ANTA	2014	1	1	1	1	259	−14.093	50.012	H	*Microcebus* sp.
Ambakirano	AMBK	2015	1	–	1	–	253	−13.616	49.026	H	*Microcebus mamiratra*
Total	–	–	42	40	37	73	–	–	–	–	–

Note: N = n° of individuals; cytb = n° of individuals genotyped for cytb locus; cox2 = n° of individuals genotyped for cox2 locus; Morpho (N) = sample size for morphological data; Altitude = altitude of the capture site; D = dry; H = humid; T-H = transition-humid; Clade: the mtDNA clade to which the genotyped individuals cluster. "–" = information not available.
Abbreviation: mtDNA, mitochondrial DNA.

5.3.3 Sampling and laboratory procedures

We captured mouse lemur individuals at night (between 18:00 and 07:00) using Sherman traps (H. B. Sherman Traps®), following established field procedures (Rakotondravony & Radespiel, 2009). We collected ear biopsies and 12 morphological measurements (Olivieri et al., 2007; see Supporting Information B), and released individuals at their capture site later the same day. Our handling protocol for biopsy sampling and morphological measurements does not require anesthetization of the captured individuals. We recorded the global positioning system (GPS) coordinates of the capture site and acquired the altitude information from a digital elevation model database using GPS Visualizer (http://www.gpsvisualizer. com/elevation). We stored the biopsies in Queen's lysis buffer (Seutin, White, & Boag, 1991) at ambient temperature in the field (between 2 and 6 months, depending on when the individual was caught) and at −20°C without changing buffer solution, once the samples are shipped to our laboratory. We extracted total

genomic DNA from 42 new samples using a Qiagen modified protocol described in Sgarlata et al. (2018) and Aleixo-Pais et al. (2018). We amplified and sequenced two mitochondrial (mtDNA) loci using published primers: cytochrome b (cytb; 1,140 bp; Irwin, Kocher, & Wilson, 1991) and the cytochrome c oxidase subunit II (cox2; 684 bp; Adkins & Honeycutt, 1994). Details on the polymerase chain reaction (PCR) amplification can be found in Sgarlata et al. (2018). We sequenced the PCR products in an ABI 3130 XL Genetic Analyzer (Applied Biosystems, Foster City, CA) and edited the sequences using GENEIOUS PRO v10.2.3 (http://www.geneious.com, Kearse et al., 2012). Data for both mtDNA loci were available for 36 out of 42 new *Microcebus* samples.

We retrieved additional mtDNA sequences from Weisrock et al. (2010) and Hotaling et al. (2016). The complete dataset does not include data *for M. margotmarshae*, given that sequences on the same mtDNA loci were not available. The complete datasets consisted of (a) cytb sequences for 313 individuals, representing 22 described *Microcebus* species; and (b) cox2 sequences for 335 individuals, representing 24 described *Microcebus* species. We built an additional dataset by combining cytb and cox2 sequences. We aligned and visually checked the sequences using the Clustal Omega method (Sievers et al., 2011). We performed a format conversion of DNA alignments using ALTER (Glez-Peña, Gómez-Blanco, Reboiro-Jato, Fdez-Riverola, & Posada, 2010).

5.3.4 Phylogenetic analyses

We used two statistical methods for phylogenetic analysis: a Bayesian Markov Chain Monte Carlo (MCMC) approach implemen- ted in MRBAYES v3.2.1 (Ronquist et al., 2012) and an ML approach implemented in RAXML v8.2.X (Stamatakis, 2014; see Appendix for more details).

5.3.5 Species validation

Phylogenetic analyses showed the presence of a monophyletic clade never reported before, closely related to *M. sambiranensis* and *M. arnholdi* (Figure 2). For the sake of simplicity, we will refer to the new mtDNA clade as "*Microcebus* sp." in the rest of the manuscript. We tested the species delimitation hypothesis between the three sister-groups, *M. sambiranensis*, *M. arnholdi* and the new *Microcebus* sp. clade, using the BPP v3.4 software (Yang, 2015), a Bayesian MCMC program for analyzing DNA sequence alignments under the MSC model (Rannala & Yang, 2003; Yang, 2002). While the method is usually used with multilocus data, it has also been shown that such a model can accurately assign species identity with even one locus (Yang & Rannala, 2017; see Appendix for further details).

Methods based on the MSC model, as BPP software, under some conditions, cannot efficiently distinguish species and population genetic structure, thus leading to the identification of spurious species (Sukumaran & Knowles, 2017). Intraspecific genetic structure arises when there is reduced gene-flow among populations, which can be due, for instance, to (a) geographical distance (i.e., isolation by distance; Wright, 1943); and/or (b) geographical isolation, as a consequence of the presence of a partial barrier between populations (Avise, 2000). To assess whether *M. arnholdi* and the new *Microcebus* sp. clade represent intraspecific lineages isolated due to geographical distance, we compared phylogenetic and geographical distance among individuals (within and between clades). This is very similar to the commonly performed test of "isolation by distance," and the assumption we make is that individuals belonging to the same species should show a similar level of genetic differentiation when separated by similar geographic distances. If the two mtDNA lineages relate to intraspecific population structure, it would be more likely to observe, for the same geographic distance, similar values of phylogenetic distance between interclades and

intraclade comparisons. Of course, this cannot be an absolute test but a reasonable one at this stage. We performed this analysis with the cytb, cox2 and combined cytb+cox2 dataset in R (R Core Team, 2018). We estimated phylogenetic distances between individuals with the adephylo R package (Jombart, Balloux, & Dray, 2010), using the patristic distance, that is the sum of branch lengths on the path between individuals. We used pairwise.wilcox.test function in R to assess significant differences in the distribution of phylogenetic distance values between intraclade and interclades groups. Last, we compared the proportion of nucleotide differences of recognized *Microcebus* species pairs with that of *Microcebus* sp.–*M. arnholdi* clade pair. We computed the proportion of nucleotide differences on the cytb and cox2 datasets using the software MEGA v7 (Kumar, Stecher, & Tamura, 2016).

5.3.6 Morphological analyses

We used 12 morphological measurements to assess whether morphological differences could be detected between M. arnholdi, *M. tavaratra*, *Microcebus* sp., and *M. mamiratra* clades. We performed the analysis on four groups of individuals: (a) *M. tavaratra–M. arnholdi*; (b) *M. tavaratra–M. arnholdi* in BIN forest; and (c) *M. arnholdi–Microcebus* sp. clade; (d) *M. arnholdi–Microcebus* sp. clade-*M. mamiratra* (Table S3). For comparison purposes, we also used morphological data of individuals from AMBRE (Station de Roussettes; N = 18) and Ankarana (ANKAR; N = 7), for which no genetic data are presented here, and BIN (N = 24; genetically identified as *M. tavaratra* in Sgarlata et al., 2018). AMBRE is inhabited by *M. arnholdi* (Louis et al., 2008), while Ankarana and BIN by *M. tavaratra* (Andriantompohavana et al., 2006); this is why data from these sites were used as reference for species morphological assignment. In addition, we conducted the same analysis on Microcebus bongolavensis–*Microcebus ravelobensis*, well-known *Microcebus* sister species for which similar morphological measures were available (Olivieri et al., 2007). We used the K-

means clustering algorithm to identify the most likely number of divergent groups based on morphological information (see Appendix for more details). We also used the Discriminant Analysis of Principal Component (DAPC) to identify which morphological variables contribute most significantly to the morphological structure. DAPC analysis was performed using the dapc function in adegenet R package (Jombart, 2008; Jombart & Ahmed, 2011).

Finally, we examined whether a certain percentage of individual morphological variability could be explained by the effect of abiotic factors. We used the individual coordinates of the first or second dimensions of the K-means clustering analysis and 19 bioclimatic variables extracted from the WorldClim v2 dataset (Fick & Hijmans, 2017). Bioclimatic information for the georeferenced locations of the sampled individuals was obtained using the raster and sp R packages (Bivand, Pebesma, & Gómez-Rubio, 2013; Hijmans, 2017; Pebesma & Bivand, 2010). To disentangle the effects of either geography (as proxy for genetic relatedness) or climate on morphological variability, we used variance-partitioning analysis for estimating the proportions of variation explained by (a) climate controlling for geography (Climate | Geography); (b) geography controlling for climate (Geography | Climate); and (c) the joint contribution of geography and climate (Geography AND Climate). We performed the same analyses using as a response variable “mtDNA identity”, instead of “morphology”. We then selected, among the 19 bioclimatic variables, those that showed low levels of collinearity, measured as Pearson correlation ($R < 0.2$) and assessed differences of those variables on the distribution of the Microcebus species studied here. Species pairwise comparisons were assessed for significance using the nonparametric Wilcoxon test.

5.4 Results

From an extensive survey effort including forests never sampled before, we present new mitochondrial genetic data for 42 *Microcebus* individuals and new morphological data from 99 individuals. We used these data to clarify the taxonomy and phylogeography of mouse lemurs in northern Madagascar. Using ML and Bayesian coalescent-based methods to reconstruct Microcebus phylogeny, we propose to update the distribution of three species (*M. tavaratra*, *M. arnholdi*, and *M. mamiratra*). Moreover, our phylogenetic analyses reveal a novel mtDNA lineage which, based on several complementary approaches, is supported to be a possible new *Microcebus* species.

5.4.1 Phylogenetic analyses

The phylogenetic analyses performed with the combined mtDNA loci (i.e., cox2 + cytb) were consistent with previously published *Microcebus* phylogenies (Figure 2; Hotaling et al., 2016; Louis & Lei, 2016; Weisrock et al., 2010). Both ML (RAXML) and Bayesian (MRBAYES) methods produced similar trees, with few differences in clade topology (Figure 2; Figures S1 and S2). In particular, the cytb + cox2 RAXML phylogenetic tree showed discordance for dichotomies between *Microcebus tanosi* and *M. tavaratra*, inconsistency on the sister-species of *Microcebus rufus* and lack of monophyly for *Microcebus gerpi* (Figure S2c). In addition, we noticed that, in both RAXML and MRBAYES trees, three *Microcebus murinus* individuals (top of Figure 2 and Figure S2c) could not be associated unambiguously to the *M. murinus* clade, likely due to the shorter mtDNA sequences of these individuals compared to the whole dataset. All previously described *Microcebus* species (N = 24) showed monophyly (Figure 2). Besides the three *M. murinus* individuals, the only exception is a divergent individual of *M. rufus*, which was also reported in Hotaling et al. (2016).

Among the 42 newly collected samples, 11 clustered with the *M. tavaratra* clade, therefore, extending its distribution range to forests north of the Irodo River (ORO, AMP, and ANKA), south-west of the M. d'Ambre (ANALA) and south of the Manambato River (ANALF; Figure 1). One additional individual clustered with the *M. mamiratra* clade broadening its distribution to the gallery forest neighboring the Ambakirano village, along the Mahavavy River in northwestern Madagascar. Moreover, our phylogenetic analyses showed that one individual from the ANTSO forest, south-west of AMBRE, and 16 individuals from humid forests south of the Loky River were genetically associated with the *M. arnholdi* clade, greatly extending its distribution. One of the most remarkable results was that the remaining 13 samples, collected south of the Manambato River, formed a new monophyletic clade (*Microcebus* sp.), sister to *M. arnholdi*. The monophyly of the new clade was well supported in all analyses except when using the cox2 dataset alone. In the following sections, we assess whether the new mtDNA clade would represent a new *Microcebus* species.

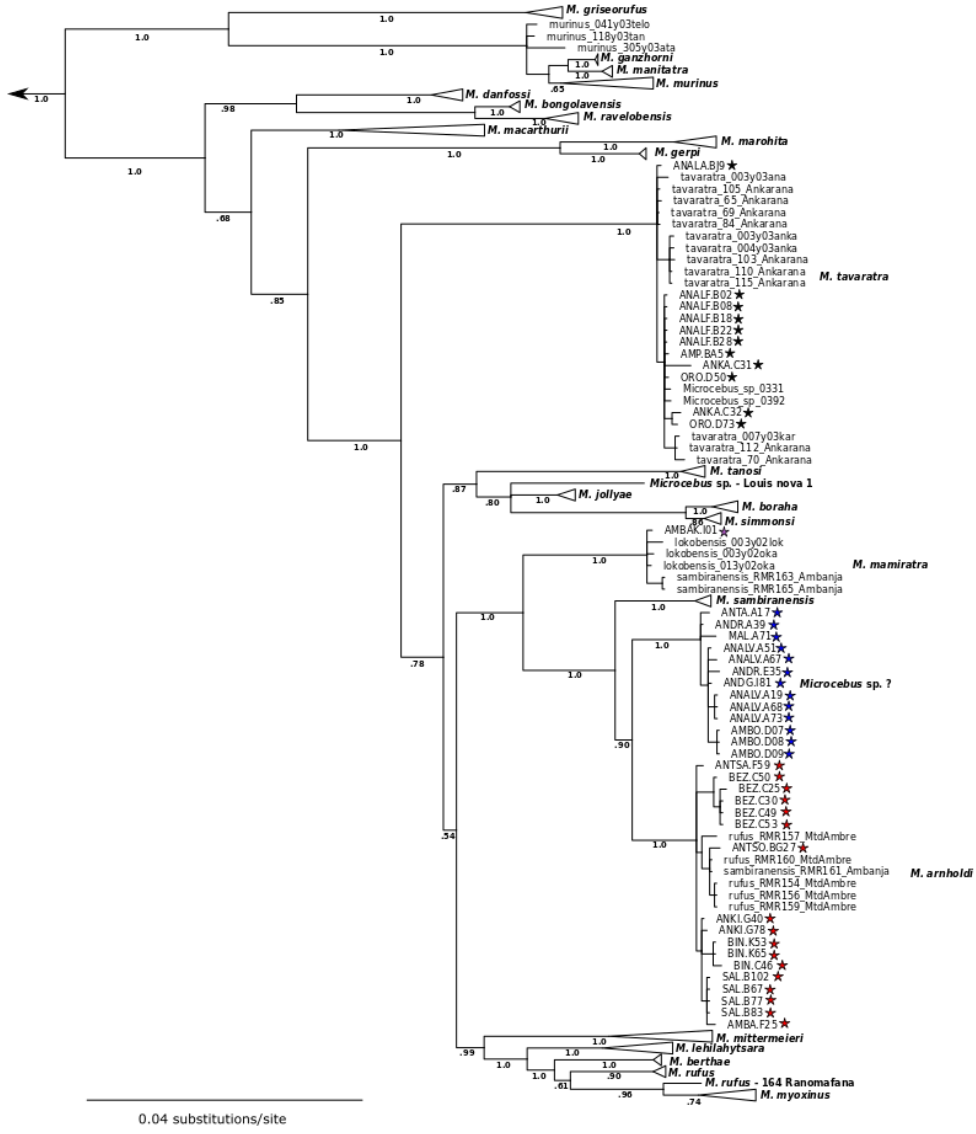


Figure 2. Bayesian phylogenetic tree of the *Microcebus* genus. We analyzed the combined mtDNA dataset (cytb + cox2) using the Bayesian Markov Chain Monte Carlo approach implemented in MRBAYES v3.2.1. Stars show the 42 new sampled individuals. Data from other *Microcebus* species were retrieved from Weisrock et al. (2010) and Hotaling et al. (2016). The numbers below branches are posterior probabilities. The sample ID does not refer necessarily to species identity, but does correspond to the sample ID present in the NCBI database.

5.4.2 Species validation analyses

The mean proportion of pairwise differences (p-distance) of cytb among all recognized *Microcebus* species ranged from 0.9% and 15.3% (Figure S3a). The p-distance between *M. arnholdi* and the new *Microcebus* sp. clade (1.5%) appeared greater than the p-distance between several currently recognized species pairs (N = 4), namely *Microcebus simmonsi–Microcebus boraha* (0.9%), *Microcebus manitatra–Microcebus ganzhorni* (1%), *M. ganzhorni–M. murinus* (1.2%) and *M. ravelobensis–M. bongolavensis* (1.2%). For cox2 we found that *Microcebus* species diverged between 0.7% and 16.3% (Figure S3b). Again, several recognized species pairs (N = 3) exhibited p-distances smaller than *M. arnholdi*-new *Microcebus* sp. clade pair (1.3%): *M. simmonsi–M. boraha* (0.7%), M. manitatra–*M. ganzhorni* (1%), *Microcebus* sp.–*M. sambiranensis* (1.2%).

When we used the BPP species delimitation method to test whether *M. arnholdi*, *M. sambiranensis*, and *Microcebus* sp. represented distinct species, we found strong support of species divergence (Table 2), robust to all combinations of priors (minimum posterior probability = 0.9997). The results of the randomized datasets showed (as expected) zero support for species divergence among the three clades. Finally, we performed BPP tests on the species pairs that showed lower p-distance than *Microcebus* sp.–*M. arnholdi* (Table S4). Overall, we found the support of species delimitation also for these species' pairs, although for some prior combinations posterior probability values were very low (Table S4a).

We further performed an ad hoc test that compares phylogenetic and geographical distance among individuals (within and between taxa), to assess whether *Microcebus* sp. would correspond to a new *Microcebus* species or to intraspecific genetic structure within *M. arnholdi*. For all mtDNA datasets (Figure 3 and Figure S4), we found, for similar values of

geographical distance, overall higher inter than intra clades phylogenetic distances (Wilcoxon signed-ranks test: χ^2 = 278.43; df = 2; p < .001). We performed the same analysis on *M. bongolavensis–M. ravelobensis*, one of the species pairs with lower p-distance than *Microcebus* sp.–*M. arnholdi* (Figure 3b). Again, we observed overall higher inter than intraclades phylogenetic distances (Wilcoxon signed-ranks test: χ^2 = 143.78; df = 2; p < .001), although at a lower extent than *Microcebus* sp.–*M. arnholdi* comparison (Figure 3; Figures S4 and S5).

Table 2. Posterior probabilities from the species delimitation test for *Microcebus* sp., *Microcebus arnholdi*, and *Microcebus sambiranensis*.

	alg0	alg0-random	alg1	alg1-random
Small pop. & deep div.	1	0	1	0
Small pop. & recent div.	1	0	1	0
Medium pop. & deep div.	1	0	1	0
Medium pop. & recent div.	1	0	1	0
Large pop. & deep div.	0.9998	0	1	0
Large pop. & recent div.	0.9998	0	0.9997	0

Note: alg0 and alg1: results using either algorithm0 or algorithm1; alg0-random and alg1-random present the results for the randomized dataset. We performed the BPP analyses on the combined mtDNA dataset (cytb + cox2). The first column describes the combination of priors (effective population size and time of divergence) used in the analyses. The species delimitation test suggests with high probability that the three clades correspond to distinct species.
Abbreviation: mtDNA, mitochondrial DNA.

5.4.3 Morphological analyses

We used 12 morphological variables across all captured *Microcebus* species. Summary statistics for each species are presented in Table 3. To assess whether species differ morphologically we performed K-means clustering analyses on the morphological data (Table S3 and Figure 4a–c). The analyses carried out with *M. arnholdi* and *M. tavaratra* (Groups I and II; Table S3), identified K = 2 as the most likely number of clusters and allowed discriminating both species (Figure S6a,b and Table S5a,b), with clusters partitioned along the dimension explaining most variation (36.4% for Group I and 55.1% for Group II). The individual from ANTSO forest clustered with AMBRE samples, wherein *M. arnholdi* is known to occur (Louis et al., 2008).

Most BIN individuals clustered with *M. tavaratra*, except three (BIN.C46, BIN.K53, and BIN.K65) that were associated with *M. arnholdi* (Figure 4b and Table S5b). The analysis carried out with *M. arnholdi* and *Microcebus* sp. showed K = 3 as the most likely number of clusters (Figure S6c). However, one of the three clusters included only one individual (Figure 4c). The other individuals were scattered in two clusters that were not representing species/clade identity. Similarly, the K-means analysis performed on the three genetically close clades (*M. arnholdi*, *Microcebus* sp., and *M. mamiratra*; Table S3, Group IV) did not identify morphological differences (data not shown). For comparison, we also assessed whether morphological differences would be detectable in another well-known group of sister species (viz., *M. bongolavensis–M. ravelobensis*; Olivieri et al., 2007). The K-means analysis determined that K = 2 was the best number of clusters (Figure S7); however, *M. ravelobensis* individuals were distributed across the two clusters, not showing a clear morphological divergence between the two species.

We first performed a DAPC analysis on the species comparison that showed clear morphological differences, that is, *M. arnholdi– M. tavaratra* (Group I dataset; Table S3). We used species identity as prior groups, to identify the morphological variables that significantly contribute to the morphological structure between the two species (Figure S8c). Our results show that “Tail length” is the morphological variable that has the largest effect on phenotypic structure, being on average longer in M. tavaratra than in *M. arnholdi* (Figure S8d). We also performed DAPC analysis on the “cryptic” sister-taxa *M. arnholdi–Microcebus* sp. and M. bongolavensis–*M. ravelobensis*, considering that morphological differences might be constrained to a few variables instead of the whole organism. For *M. arnholdi–Microcebus* sp., DAPC analysis showed that “Ear length,” “Ear width,” and “Tail length” were the morphological variables that have the largest effect on phenotypic structure between *M. arnholdi* and *Microcebus* sp.; however, differences for each variable were not significant (Figure S9a–c). Whereas,

M. bongolavensis–M. ravelobensis showed “Intraorbital distance” and “Hind foot length” as the significant discriminant morphological variables, being, for both variables, longer in *M. bongolavensis* than in *M. ravelobensis* (Figure S9d–f). We investigated the effects of climate and geography on morphology by conducting a partial redundancy analysis, first on the continuous morphological coordinates, and then on the discrete morphological clusters identified with the K-means approach. For *M. tavaratra* and *M. arnholdi* (Group I; Table S3), we detected a significant association between morphology and climate (adjusted R^2 = 0.56; Table 4a). Accordingly, partial redundancy analyses on clade identity suggest a significant relationship between species identity and climate (adjusted R^2 = 0.63; Table 4b). However, when we performed the same analyses only on the *M. tavaratra* and *M. arnholdi* individuals occurring in BIN (Group II, removing AMBRE and ANKAR; Table S3), we could not detect any significant relationship (Table 4a,b). We assessed which variable was contributing to climatic divergence in *M. tavaratra* and *M. arnholdi* distributions and we found that “Maximum temperature of the warmest month” and “Precipitation of Coldest Quarter” were the two climatic variables showing significant differences between the two species (Figure 5).

For the *M. arnholdi–Microcebus* sp. comparison (Group III in Table 4a), we detected a positive association with geography (adjusted R^2 = 0.21; $p < .001$) and the joint contribution of geography and climate (adjusted R^2 = 0.41; p = not testable). Also, the partial redundancy analysis performed on cluster identity did not show correspondence between clusters and abiotic factors (Group III in Figure 4c and Table 4b).

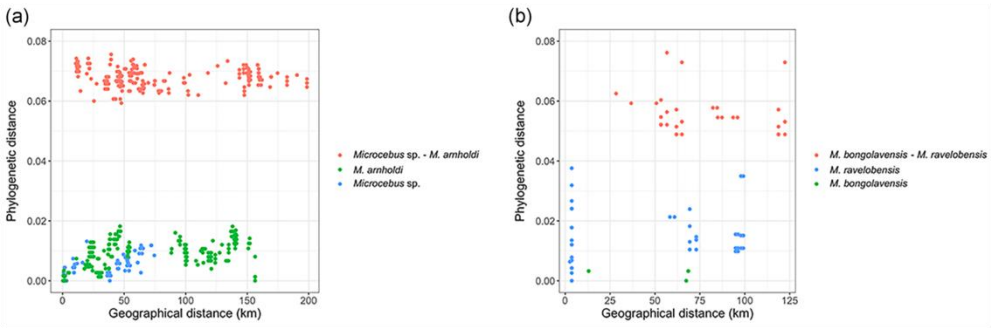


Figure 3. Relationship between phylogenetic and geographic distance. These results refer to the analyses performed with the combined mtDNA dataset (cytb + cox2). Phylogenetic distance is measured as patristic distance, which is the sum of branch lengths on the path between individuals. (a) *Microcebus* sp.–*Microcebus arnholdi*, and (b) *Microcebus bongolavensis* – *Microcebus ravelobensis* comparisons showed, for the same geographic distance, higher interclade than intraclade phylogenetic distance. We expect higher values of phylogenetic distance with respect to geography for individuals belonging to different species than for those of the same species. Abbreviation: mtDNA, mitochondrial DNA

Table 3. Mean and standard deviation of the morphometric variables for each *Microcebus* taxon.

	Microcebus arnholdi (N = 34)	*Microcebus tavaratra* (N = 51)	*Microcebus* sp. (N = 13)	*Microcebus mamiratra* (N = 1)
Ear length	20.5 ± 2.7	21.4 ± 2.5	17.8 ± 2.8	21.6
Ear width	12.6 ± 1.5	13.3 ± 1.2	12 ± 1.9	11.7
Head length	33.8 ± 2.2	34.1 ± 2	33.5 ± 3.1	30.7
Head width	20.5 ± 1.4	20.6 ± 1.4	19.9 ± 1.7	22.2
Snout length	7.9 ± 1.3	7.9 ± 1.2	7.9 ± 1.9	11.4
Interorbital dist.	21.5 ± 1.5	21.5 ± 1.3	20.9 ± 1.7	25.0
Intraorbital dist.	7 ± 1.4	7.1 ± 1.5	6.5 ± 0.8	6.8
Lower leg length	38.2 ± 4	39.8 ± 2.9	34.8 ± 4.8	40.1
Hind foot length	21.2 ± 1.9	21.9 ± 1.7	19.7 ± 2.4	21.6
Body length	79.7 ± 11.4	81.8 ± 10.2	76.3 ± 18.2	96.0
Tail length	145.5 ± 19.7	153.3 ± 18.8	134.3 ± 19.6	149.0
Tail circumference	19.7 ± 2.4	19.9 ± 2.2	19.8 ± 3.3	19.0

Note: All variables are expressed in mm.

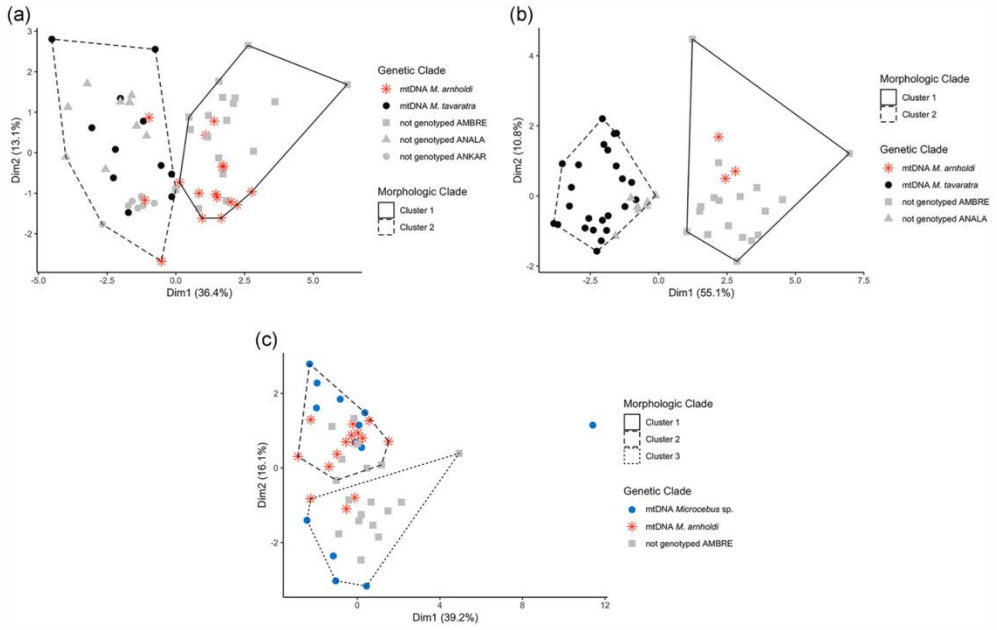


Figure 4. K-means analyses on *Microcebus tavaratra*, *M. arnholdi*, and *Microcebus* sp. morphological data. The K-means analyses show morphological differences between *M. tavaratra* and M. arnholdi, but none between M. arnholdi and Microcebus sp. (a) Group I: *M. arnholdi* (Cluster 1) versus *M. tavaratra* (Cluster 1) from Northern Madagascar; (b) Group II: *M. arnholdi* (Cluster 1) versus *M. tavaratra* (Cluster 2) from Binara forest; (c) Group III: *M. arnholdi* versus *Microcebus* sp. from northern Madagascar. AMBRE, Montagne d'Ambre; ANALA, Analabe; ANKAR, Ankarana.

Table 4. Partial redundancy analyses.

a)				
Morphology		**Adj.R^2**	**% Variance**	***p* Value**
Group I	Climate | Geography	0.56	48.92	.001
	Climate AND Geography	0.01	51.03	NA
	Geography | Climate	0.00	0.05	.65
Group II (only BIN)	Climate | Geography	0.00	4.29	.420
	Climate AND Geography	0.32	55.99	NA
	Geography | Climate	0.06	39.72	.386
Group III	Climate | Geography	0.00	0.46	.75
	Climate AND Geography	0.41	73.33	NA
	Geography | Climate	0.21	26.21	.00
b)				
Clade Identity		**Adj.R^2**	**% Variance**	***p* Value**
Group I	Climate | Geography	0.63	53.92	.001
	Climate AND Geography	0.14	38.25	NA
	Geography | Climate	0.05	7.83	.001
Group II (only BIN)	Climate | Geography	0	0.55	.883
	Climate AND Geography	0.4	61.77	NA
	Geography | Climate	0	37.68	.52
	Climate | Geography	0.04	4.36	.024
Group III	Climate AND Geography	0.47	70.21	NA
	Geography | Climate	0.22	25.43	.001

Note: Group I: *M. arnholdi* versus *M. tavaratra* in northern Madagascar; Group II: *M. arnholdi* versus *M. tavaratra* in BIN forest; Group III: *Microcebus* sp. versus *M. arnholdi*. Climate | Geography: effect of climate controlling for spatial autocorrelation; Climate AND Geography: joint effect of climate and geography; Geography | Climate: effect of geography controlling for climate; Adj.R^2: adjusted R^2, a measure of the predictive power; % Variance: the proportion of the variance. NA: significance cannot be tested. We assessed the effect of climate and geography on (a) morphological variability and (b) mtDNA clade identity of the studied mouse lemur species. The results suggest a significant relationship between morphology, clade identity, and climate for *M. tavaratra* and *M. arnholdi* (Group I).

Abbreviations: BIN, Binara forest; mtDNA, mitochondrial DNA.

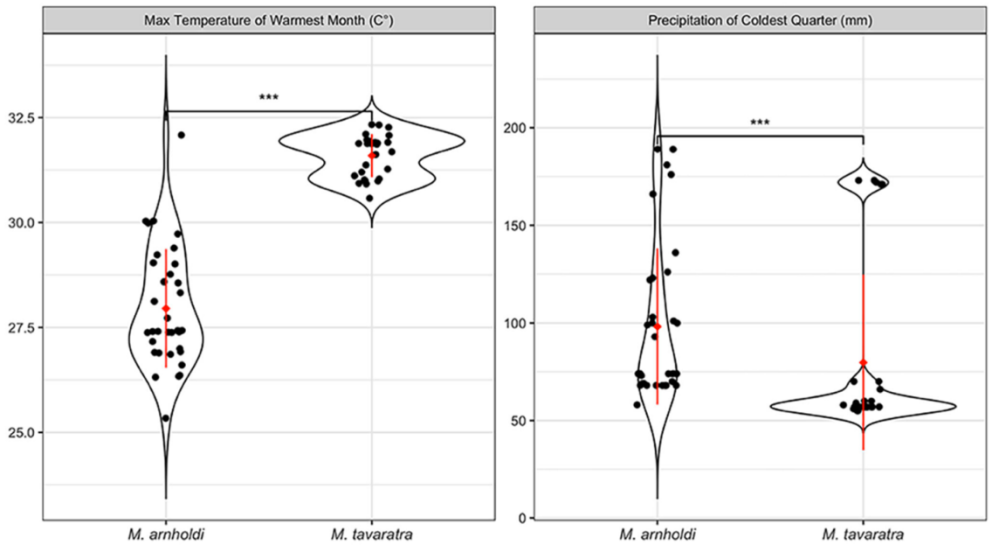


Figure 5. Bioclimatic differences between *M. tavaratra* and *M. arnholdi*. Among the 19 bioclimatic variables tested in the presented study, we selected those that showed a level of colinearity (measured as correlation) <0.2. Each dot represents the climatic value extracted from the site in which each *Microcebus* individual was sampled. We observed significant climatic differences between *M. tavaratra* and *M. arnholdi* sampling sites. ***p < .01

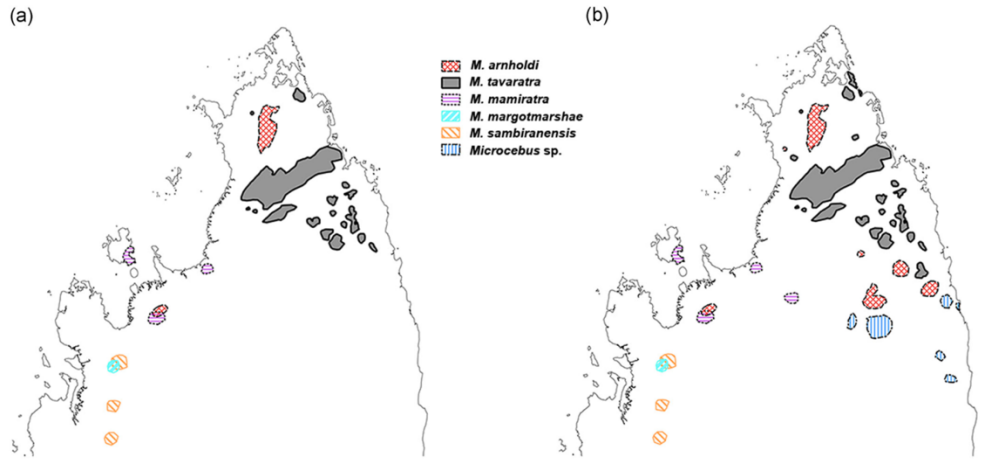


Figure 6. Species distribution of northern *Microcebus*. (a) Species distributions before the present study; (b) species distributions after the present study. Our results support an increase of the area of occurrence for M. tavaratra (6.6%), M. arnholdi (114%), and *Microcebus mamiratra* (29%)

5.5 Discussion

5.5.1 Update of species distributions

M. tavaratra was known to occur between forests north of the Irodo River and the Manambato River, including also the Ankarana Special Reserve (Figure 1). Our phylogenetic analyses not only confirm that *M. tavaratra* occurs in forests located north of its previous putative distribution but is also present south of the Manambato River, in the ANALF forest (Figure 1). These results lead to a total increase of the species-area distribution of 6.6% (Figure 6) and include locations that were previously predicted by ecological niche modeling (Kamilar, Blanco, & Muldoon, 2016). Overall, we observed that current *M. tavaratra* populations are scattered across the region in forests that are separated by nonforested habitat (grasslands, agricultural land, etc.), suggesting past connectivity between present-day dry forest fragments.

M. mamiratra was first reported in the Nosy Be Island and later along the north-west coast (Figure 1; Andriantompohavana et al., 2006; Olivieri et al., 2007). Our results provide evidence of its presence in the riparian forest neighboring Ambakirano, along the Mahavavy River (AMBK, Figure 1), therefore, widening its distribution of 29% to the east and suggesting ancient forest connectivity among these three localities (Figure 6).

Our results also greatly extend *M. arnholdi*'s distribution area (114%). The species was thought to be mostly limited to the mountainous forest of Montagne d'Ambre until it was described in Ambanja, 130 km south-west of Montagne d'Ambre, by Weisrock et al. (2010; Figure 6). Our results suggest that the distribution of *M. arnholdi* also extends south-east but appears to be restricted to the humid forest patches of northern Madagascar (Figure 1). As for the other species, this scattered distribution suggests that these forests may represent relict habitats of an ancient continuous humid forest. The

scattered distribution of *M. arnholdi* in humid forests is surprisingly located within a well-recognized dry biogeographic zone (BIN and SAL; Figure S11), and its presence in the small humid area within the dry ANTSO forest (Figure 1; Figures S10 and S11), points to humid vegetation as a key factor for *M. arnholdi* resilience. For BIN and SAL forests, altitude likely has a major role in maintaining a humid microclimate, as it has been demonstrated for other tropical regions (e.g., Anderson et al., 2018). The case of the ANTSO forest is different as this forest is located at a low altitude and mainly composed of dry vegetation. Interestingly, we sampled one *M. arnholdi* individual in the only part of that fragment that exhibited moist vegetation. Physical characteristics of the soil (e.g., geomorphology) are a known factor affecting the tree access to water, and, therefore, the type of vegetation at the land surface (Chase, Johnson, & Martin, 2012). Thus, we hypothesize that particular geomorphological characteristics can contribute to the persistence of humid microclimate suitable to the survival of *M. arnholdi*, even at a small spatial scale such as that of the ANTSO forest.

5.5.2 A new location of *Microcebus* sympatry

Importantly, we discovered the presence of *M. arnholdi* in the BIN mountainous forest, previously thought to be only inhabited by *M. tavaratra*. Our results represent the first demonstration of co- occurrence of *M. tavaratra* and *M. arnholdi* (see BIN in Figure 1). Several *Microcebus* species pairs occur in sympatry in western Madagascar, which generally include the widely distributed grey mouse lemur (*M. murinus*). *M. murinus* has a very wide distribution, likely due to a recent expansion across western Madagascar (Schneider, Chikhi, Currat, & Radespiel, 2010). As a consequence, it is co-distributed with three locally restricted congeneric species: *M. ravelobensis*; *Microcebus berthae*; *Microcebus myoxinus* (Rasoloarison et al., 2000; Schmid & Kappeler, 1994; Zimmermann et al., 1998). The cooccurrence we describe may also be the result of a recent expansion of one of the two

species (*M. arnholdi* or *M. tavaratra*). Our result raises to four the number of published co-occurrences of unrelated *Microcebus* species (not including contact zones between sister species). Further genetic analyses are required to determine whether interbreeding occurs between *M. tavaratra* and *M. arnholdi* in BIN.

5.5.3 Insights on recent forest changes from forest-dwelling species distributions

Our findings confirm that *M. tavaratra* and *M. arnholdi* are distributed in areas characterized by divergent climatic conditions (Groups I and II in Table 4a), which can be represented by the two bioclimatic variables showing low levels of collinearity (Figure 5). Unlike *M. arnholdi*, *M. tavaratra* seems to be more adapted to dry habitats (Figure 5 and Table 4b). This explains its broad presence in the mostly dry northern Madagascar (Figure 1). Differences in climatic preferences are also associated with clear differences in morphology between *M. tavaratra* and *M. arnholdi* individuals (Figure 4a,b), supported by the significant contribution of climate on morphology (Groups I and II in Table 4a). Interestingly, we found that “Tail length” is the morphological variable that contributes the most to the phenotypic divergence between *M. tavaratra* and *M. arnholdi* (Figure S8c), being shorter in the *Microcebus* species commonly associated to colder habitat (*M. arnholdi*; Figure S8d and Figure 5). This is consistent with previous studies showing that mice raised at warm temperatures have a significantly longer tail than those raised in cold temperatures (Serrat, 2014). The mechanism behind this phenotype needs to be explored, but according to Allen’s rule (Allen, 1877) this could represent an adaptive trait that, under cold temperatures, reduces heat loss through body extremities (e.g., tail).

Our results suggest a possible link between taxonomy, morphological variability, and environmental factors for these two species. Thus,

assuming that *M. tavaratra* and *M. arnholdi* have retained the ecological traits related to their current niche (viz., “niche conservatism”; Wiens et al., 2010) during the late Quaternary, we can use present distribution of the dry-associated (*M. tavaratra*) and humid-associated (*M. arnholdi*) mouse lemurs to reconstruct the recent history of dry and humid forests in northern Madagascar. Looking at the large distribution and presence of *M. tavaratra* in northern dry forests and the broad scattered distribution of *M. arnholdi* in humid forests, we speculate that current species distributions are the result of a recent shift from humid to dry forest cover in northern Madagascar, with consequent range expansion of *M. tavaratra* and contraction of *M. arnholdi*. In fact, the finding in the currently dry forests of the Ankarana and Montagne des Français of *Propithecus cf. diadema*, *Indri indri*, and *Prolemur simus* subfossils, is particularly notable. Indeed, these larger body-size lemurs currently occur only in the humid forests of eastern Madagascar, which suggests that during early-mid Holocene, northern Madagascar was more humid than today (Godfrey, Jungers, Simons, Chatrath, & Rakotosamimanana, 1999; Jungers, Godfrey, Simons, & Chatrath, 1995; Simons et al., 1995; Figure 1). Therefore, we propose that when humid habitats were more continuous than today, *M. arnholdi* would have had a larger and continuous distribution than *M. tavaratra*. Then, the increase of dry conditions during the Holocene may have caused the contraction and disconnection of humid forests and an increase of dry habitats (Gasse & Van Campo, 1998; Virah-Sawmy, Willis, Gillson, & Williams, 2010), confining *M. arnholdi* populations in the remaining humid forest fragments, while *M. tavaratra* populations would have been able to increase their range in the expanding dry forests. Subsequent dry events and possible human activities may have further fragmented dry forests to the extent that we observe today. Similar scenarios have been proposed to explain the genetic diversity, demographic history, and distribution of northern Madagascar’s tuft- tailed rats (*Eliurus carletoni* and *Eliurus tanala*; Rakotoarisoa, Raheriarisena, & Goodman, 2013a, 2013b), sifaka (Propithecus tattersalli; Quéméré, Amelot,

Pierson, Crouau-Roy, & Chikhi, 2012; Salmona, Heller, Quéméré, & Chikhi, 2017) and the leaf chameleons (e.g., Brookesia ebenaui and Brookesia minima; Raxworthy & Nussbaum, 1995). More genetic data coupled with population divergence time estimates and/or spatial simulations would help infer which humid to dry climatic oscillation likely led to the current distribution of mouse lemurs in northern Madagascar.

5.5.4 New mtDNA lineage

Our phylogenetic analyses identified a novel mtDNA clade (*Microcebus* sp., Figure 2) closely related to the *M. arnholdi* clade. The proportion of pairwise differences between *M. arnholdi* and *Microcebus* sp. was larger than the differences observed between well accepted pairs of species (N = 3 or 4; Figure S3a,b). Using a Bayesian coalescent-based species delimitation test (BPP software) we found support to consider *M. sambiranensis*, *M. arnholdi* and *Microcebus* sp. as distinct species (Table 2). The BPP results for *M. ganzhorni*, *M manitatra*, and *M. murinus* and those for *Microcebus danfossi*, *M. bongolavensis* (Table S4), were much weaker under several parameter combinations. Although our analyses are limited to the use of two mitochondrial loci, our BPP results for *M. arnholdi* and *M. sambiranensis* are in agreement with those of Hotaling et al. (2016), which used a combined dataset of two mtDNA and four nuclear loci.

The BPP software uses an implementation of the MSC model, and recent work (Sukumaran & Knowles, 2017) suggested that, under certain conditions, BPP results can be considered reliable for species delimitation only if other types of data are used (e.g., ecological, ethological). Here we have used a simple approach, which may help to distinguish intra and interspecies genetic structure. We compared the individual phylogenetic and geographic distances within and among clades (Figure 3; Figures S4 and S5). Overall, we observed that phylogenetic distances were larger between *M.*

arnholdi and *Microcebus* sp. (red dots in Figure 3a) than within each clade, whichever geographic distance we considered (blue and green dots in Figure 3a). This suggests that geographic distance may not be the factor limiting gene flow among the populations of these two clades. Furthermore, we may exclude that *M. arnholdi* and the new monophyletic clade are the results of present physical barriers to gene-flow, given that *M. arnholdi* individuals sampled in forests separated by large open habitat and several permanent rivers still cluster within the same clade (Figure 1). Conversely, the MAL and AMBA samples are located within the same continuous forest but belong to different clusters.

Despite the genetic evidence of species delimitation for *M. arnholdi* and *Microcebus* sp., morphological analysis with K-means approach did not show clear differences between these two taxa (Figure 4c), similarly to another group of sister species (*M. bongolavensis–M. ravelobensis*) for which similar morphological variables were available (Figure S7). When we used DAPC analysis on morphological data, however, both *M. arnholdi–Microcebus* sp. and *M. bongolavensis–M. ravelobensis* comparisons identified three morphological variables that most differentiate each taxon from its pair. But only *M. bongolavensis–M. ravelobensis* comparison showed significant differences between the distribution of two of the three identified morphological variables (Figure S9). Following these results, we attempted to assess whether the morphological clusters identified in *M. arnholdi–Microcebus* sp. comparison would reflect differences in abiotic factors rather than species identity. We did not find support for this hypothesis; however, we found evidence for the contribution of geography on morphological variability (Group III in Table 4a). Early studies reported that morphological traits could be effective for species recognition in the *Microcebus* genus (Yoder et al., 2000); however, this is not always the case (Rasoloarison et al., 2013; Yoder et al., 2005). Hence, the lack of morphological differences between *M. arnholdi* and the individuals of the new mtDNA clade does not

necessarily reject the species hypothesis for the new mtDNA lineage. Moreover, according to the “recent divergence hypothesis”, it is expected that more genetically divergent taxa (*M. tavaratra–M. arnholdi*) would show stronger phenotypic differences than less genetically divergent one (*Microcebus* sp.–*M. arnholdi*).

Altogether, these results support the hypothesis that the new mtDNA clade may represent a distinct species. However, several *Microcebus* species are female philopatric (Radespiel, Lutermann, Schmelting, Bruford, & Zimmermann, 2003), resulting in strong mitochondrial genetic structure which may not necessarily relate to species delimitation. Therefore, we propose that the new mtDNA clade is considered at best a tentative hypothesis of distinct species, to be confirmed or rejected through the application of nuclear genomic data and/or other ecological information.

5.5.5 Conservation implications

These results have major implications for the conservation of northern Madagascar Microcebus species. Based on the preliminary results of this study, most distribution updates of the species studied here were considered in the recent IUCN Red List assessments conducted in May 2018, Antananarivo, Madagascar. However, we have to highlight that deforestation rates in northern Madagascar have been rising in the last decade, sustained by strong demands for charcoal and hardwood by neighboring major towns, and international demand for precious woods and vanilla (e.g., Vieilledent et al., 2018, Waeber, Wilmé, Mercier, Camara & Lowry, 2016). Therefore, there is a real danger that, for instance, the populations in lowland forests of the new *Microcebus* sp. (i.e., ANALV, AMBO, ANDG, and ANTA) will not last long. Indeed, the two unprotected forests surveyed in 2013 (ANALV, AMBO) had almost entirely disappeared by 2018 (JS, BLP, ER pers. obs.). This points

out the urgency to further extend the Malagasy protected area system to remote locations, to save the yet unrevealed diversity it hosts.

Acknowledgments

We thank the Direction Générale du Ministère de l'Environnement et des Forêts de Madagascar, Direction Régionale de l'Environnement, Ecologie et des Forêts (DREEF DIANA and SAVA), Madagascar's Ad Hoc Committee for Fauna and Flora and Organizational Committee for Environmental Research (CAFF/CORE), Conservation International, WWF, Missouri Botanical Garden and the Fanamby NGO. This study benefited from the continuous support of the Department of Animal Biology and Ecology, University of Mahajanga, the Department of Animal Biology, University of Antananarivo, Department of Natural and envornmental Sciences at the University of Antsiranana and from the participation of Malagasy MSc students, field assistants, volunteers, and many great local guides and cooks which we warmly thank for their help in the field and for sharing their incomparable knowledge of the forest, misaotra anareo jiaby. This research was funded through the 2015–2016 BiodivERsA COFUND call for research proposals, with the national funders ANR (grant number ANR-16-EBI3–0014), FCT (grant number Biodiversa/0003/2015) and PT-DLR (grant number 01LC1617A). It was also partly funded by the FCT (grant numbers PTDC/BIA-BEC/100176/ 2008, PTDC/BIA-BIC/4476/2012, and PTDC-BIA-EVL/30815/2017 to L.C., SFRH/BD/64875/2009 to J.S., PD/BD/114343/2016 to G.S), the LabEx entitled TULIP (grant numbers ANR-10LABX-41, ANR-11-IDEX- 0002-02), the LIA BEEG-B (Laboratoire International Associé-Bioinformatics, Ecology, Evolution, Genomics and Behaviour, CNRS) between the CNRS and IGC, the Investissement d'Avenir grant of the Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01), the Rufford Small Grant Foundation (grant number 10941-1 to JS and 12973-1 to MTI) and the Groupement de Recherche International (GDRI) Biodiversité et

developpement durable-Madagascar. In addition, we thank the ITI facility of Instituto Gulbenkian de Ciencia, in particular, João Garcia, for its logistical support in the computational analyses.

Conflict of interests

The authors declare that there are no conflicts of interest.

Data availability statement

The generated genetic datasets have been deposited in the NCBI repository, including the GPS coordinates of the sampled individuals (MN699484 - MN699560). Morphological data are provided in Supporting Information B.

Ethics statement

We conducted this study in agreement with the laws of the countries of Portugal, France, and Madagascar (CITES; permit IDs 485C-EA09/ MG10, 559C-EA10/MG13, 782C-EA12/MG14, 1035C-EA12/MG15, 994C-EA11/MG17). Our research complied with the ASP Principles for the Ethical Treatment of Nonhuman Primates and the International Primatological Society Code of Best Practices for Field Primatology.

5.6 Appendix: detailed methods

5.6.1 Phylogenetic analyses

In MRBAYES v3.2.1, we used four MCMC chains, each at the default temperature and running for 15,000,000 generations. Trees were sampled every 15,000 generations and 25% of total sampled trees were discarded as the “burn-in.” To assess MCMC convergence, we checked for stationarity of the log probability of the data over time, we verified that the potential scale

reduction factor was close to 1 and we used Tracer v1.6 (Rambaut, Drummond, & Suchard, 2014) to confirm that all estimated sample size values were >200. In RAXML v8.2.X, we used the rapid Bootstrap option with automatic detection of the sufficient number of bootstrap replicates (option: “autoMR”). In both approaches, we estimated the phylogeny for each mtDNA dataset (cox2, cytb, and cox2_cytb). We used sequences of *Cheirogaleus major* (EU825335), *Cheirogaleus medius* (EU825333), *Cheirogaleus crossleyi* (EU825354) and *Mirza coquereli* (U53571) as outgroups (retrieved from GeneBank). Retained posterior distributions of trees are summarized in MRBAYES to build a consensus tree, while in RaxML majority-rule consensus tree is estimated from the Bootstrap replicates. Phylogenetic trees were visualized in FigTree (http://tree.bio.ed.ac.uk/software/figtree/) and modified in Inkscape (https://inkscape.org/).

We estimated the best partitioning scheme and best-fit model of evolution for each mtDNA dataset in PARTITIONFINDER v2.1.1 (Guindon et al., 2010; Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2017; Stamatakis, 2006) using the greedy algorithm and selecting the best model according to AICc criteria. We performed a separate analysis for each phylogenetic method, given that MRBAYES and RAXML accept only some of the models of evolution that can be estimated in PARTITIONFINDER. Details on the selected model of evolution for each mtDNA dataset and phylogenetic method are presented in Supporting Information A (Table S1).

5.6.2 Species validation

We carried out the species delimitation analysis using a user-specified guide tree (option “A10”; Rannala & Yang, 2013; Yang & Rannala, 2010) that defines the *M. sambiranensis*, *M. arnholdi* and *Microcebus* sp. clades as different species. The BPP v3.4 software requires prior information on ancestral population size (θ) and the age of the root of the species tree (τ),

and as in Hotaling et al. (2016) we tested a wide range of prior values for a total of six combinations of θ and τ prior values (Supporting Information A, Table S2). We ran BPP on the combined mtDNA dataset (cytb + cox2), with four repetitions for each combination of prior values, using alternatively rjMCMC algorithm0 or algorithm1, a burn-in of 100,000 generations and a total of 10,000 MCMC thinning samples. We assess BPP robustness by performing the same analyses on datasets in which individuals were randomly assigned to one of the three clades.

5.6.3 Morphological analysis

We used K-means clustering, an approach that uses an unsupervised machine learning algorithm for partitioning individuals in a set of predefined K-groups by minimizing differences within-group and maximizing differences between groups. We estimated the optimal number of predefined K-groups using NbClust function in the NbClust R package (Charrad, Ghazzali, Boiteau, & Niknafs, 2014). This function computes 26 indices, each proposing the best number of K-groups, and the K value with higher support among the 26 indices is selected as the best clustering scheme. After performing the K-means clustering analysis using the K-means function in R, we validated the quality of the clustering using the average silhouette approach, implemented in the factoextra R package (Kassambara & Mundt, 2017).

We performed partial redundancy analyses to test the partial contribution of geography (as a proxy for genetic relatedness) and climate on morphological variability (Borcard & Legendre, 1994; Borcard, Legendre, & Drapeau, 1992). Significance was assessed using 1,000 permutations (analysis of variance function in R). In this analysis, we reduced the 19 bioclimatic variables to their most relevant and significant components by performing PCA analyses. Geographic distances were transformed in principal coordinates of neighbor matrices to account for positive spatial

autocorrelation among individuals (Borcard & Legendre, 2002; Dray, Legendre, & Peres-Neto, 2006). Partial redundancy analyses and tests for significance were performed in R using the vegan package (Oksanen et al., 2018).

5.7 References

Adkins, R. M., & Honeycutt, R. L. (1994). Evolution of the primate cytochrome c oxidase subunit II gene. Journal of Molecular Evolution. 38, 215–231.

Aleixo-Pais, I., Salmona, J., Sgarlata, G. M., Rakotonanahary, A., Sousa, A. P., Parreira, B., ... Chikhi, L. (2018). The genetic structure of a mouse lemur living in a fragmented habitat in Northern Madagascar. Conservation Genetics. 20, 229–243.

Allen, J. A. (1877). The influence of physical conditions in the genesis of species. Radical Review, 1, 107–140.

Anderson, J., Keppel, G., Thomson, S., Randell, A., Raituva, J., Koroi, I., ... Kleindorfer, S. (2018). Changes in climate and vegetation with altitude on Mount Batilamu, Viti Levu, Fiji. Journal of Tropical Ecology, 34(5), 316–325.

Andriaholinirina, N., Baden, A., Blanco, M., Chikhi, L., Cooke, A., Davies, N., & Zaramody, A. (2014). Microcebus tavaratraMicrocebus tavaratra. The IUCN Red List of Threatened Species 2014: e.T41571A16113861.

Andriantompohavana, R., Zaonarivelo, J. R., Engberg, S. E., Randriamampionona, R., Mcguire, S. M., Shore, G. D., ... Louis, E. E. (2006). Mouse lemurs of northwestern Madagascar with a description of a new species at Lokobe Special Reserve. Occasional papers/Museum of Texas Tech University. 259, 1–24.

Avise, J. C. (2000). Phylogeography: The History and Formation of Species. Cambridge, MA: Harvard University Press.

Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2013). Applied Spatial Data Analysis with R. New York: Springer-Verlag.

Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecological Modelling, 153, 51–68.

Borcard, D., & Legendre, P. (1994). Environmental control and spatial structure in ecological communities: An example using oribatid mites (Acari, Oribatei). Environmental and Ecological Statistics, 1, 37–61.

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. Ecology, 73, 1045–1055.

Burgin, C. J., Colella, J. P., Kahn, P. L., & Upham, N. S. (2018). How many species of mammals are there? Journal of Mammalogy. 9, 1–14.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for determining the relevant number of clusters in a dataset. Journal of Statistical Software. 61, 10828.

Chase, M. N., Johnson, E. A., & Martin, Y. E. (2012). The influence of geomorphic processes on plant distribution and abundance as reflected in plant tolerance curves. Ecological Monographs, 82, 429–447.

Chenuil, A., Cahill, A. E., Délémontey, N., Du Salliant Du Luc, E., & Fanton, A. (2019). Problems and questions posed by cryptic species. A framework to guide future studies. In E. Casetta, J. Marques da Silva & D. Vecchi (Eds.), From Assessing to Conserving Biodiversity. History, Philosophy and Theory of the Life Sciences (24, pp. 77–106). Switzerland: Springer Nature.

Cracraft, J. (1983). Species concepts and speciation analysis. Current Ornithology. 1, 159–187.

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecological Modelling, 196, 483–493.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology. 37, 4302–4315.

Gasse, F., & Van Campo, E. (1998). A 40,000-yr pollen and diatom record from Lake Tritrivakely, Madagascar, in the Southern Tropics. Quaternary Research. 49, 299–311.

Glez-Peña, D., Gómez-Blanco, D., Reboiro-Jato, M., Fdez-Riverola, F., & Posada, D. (2010). ALTER: Program-oriented conversion of DNA and protein alignments. Nucleic Acids Research. 38, W14–W18.

Godfrey, L. R., Jungers, W. L., Simons, E. L., Chatrath, P. S., & Rakotosamimanana, B. (1999). Past and Present Distributions of Lemurs in Madagascar (pp. 19–53). New York: Kluwer Academic/Plenum.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology. 59, 307–321.

Harper, G. J., Steininger, M. K., Tucker, C. J., Juhn, D., & Hawkins, F. (2007). Fifty years of deforestation and forest fragmentation in Madagascar. Environment Conservation Journal. 34, 325–333.

Hijmans, R. J. (2017). raster: Geographic data analysis and modeling. R package version, 2, 6–7.

Hotaling, S., Foley, M. E., Lawrence, N. M., Bocanegra, J., Blanco, M. B., Rasoloarison, R., & Weisrock, D. W. (2016). Species discovery and validation in a cryptic radiation of endangered primates: Coalescent- based species delimitation in Madagascar's mouse lemurs. Molecular Ecology. 25, 2029–2045.

Irwin, D. M., Kocher, T. D., & Wilson, A. C. (1991). Evolution of the cytochrome b gene of mammals. Journal of Molecular Evolution. 32, 128–144.

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics, 24, 1403–1405.

Jombart, T., & Ahmed, I. (2011). Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics. 27, 3070–3071.

Jombart, T., Balloux, F., & Dray, S. (2010). Adephylo: New tools for investigating the phylogenetic signal in biological traits. Bioinformatics. 26, 1907–1909.

Jungers, W. L., Godfrey, L. R., Simons, E. L., & Chatrath, P. S. (1995). Subfossil Indri indri from the Ankarana Massif of northern Madagascar. American Journal of Physical Anthropology. 97, 357–366.

Kamilar, J. M., Blanco, M. B., & Muldoon, K. M. (2016). Ecological niche modeling of mouse lemurs (Microcebus sp.) and its implications for their species diversity and biogeography. In S. M. Lehman, U. Radespiel & E. Zimmermann (Eds.), The Dwarf and Mouse Lemurs of Madagascar: Biology, Behavior, and Conservation Biogeography of the Cheirogaleidae (pp. 449–618). Cambridge, United Kingdom: Cambridge University Press.

Kappeler, P. M., & Rasoloarison, R. M. (2003). Microcebus, mouse lemurs, Tsidy. In S. M. Goodman & J. P. Benstead (Eds.), The Natural History of Madagascar (pp. 1310–1315). Chicago, USA: University of Chicago Press.

Kassambara, A., & Mundt, F. (2017). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.5.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., & Drummond, A. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28, 1647–1649.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for Bigger datasets. Molecular Biology and Evolution. 33, 1870–1874.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution. 34, 772–773.

Louis, E. E., Coles, M. S., Andriantompohavana, R., Sommer, J. A., Engberg, S. E., Zaonarivelo, J. R., & Brenneman, R. A. (2006). Revision of the mouse lemurs (Microcebus) of eastern Madagascar. International Journal of Primatology. 27, 347–389.

Louis, E. E., Engberg, S. E., Mcguire, S. M., Mccormick, M. J., Randria-mampionona, R., Ranaivoarisoa, J. F., & Lei, R. (2008). Revision of the mouse lemurs, Microcebus (primates, Lemuriformes), of Northern and Northwestern Madagascar with descriptions of two new species at Montagne d'Ambre National Park and Antafondro Classified Forest. Primate Conservation, 23, 19–38.

Louis, E. E., & Lei, R. (2016). Mitogenomics of the family Cheirogaleidae and relationships to taxonomy and biogeography in Madagascar. In U. Radespiel, E. Zimmermann & S. M. Lehman (Eds.), Dwarf and mouse lemurs of Madagascar: biology, behavior, and conservation biogeography of the Cheirogaleidae. Cambridge (UK): Cambridge Universtiy Press.

Mittermeier, R.A., Hawkins, F., Kappeler, P.M., Langrand, O., Louis, E.E., Mackinnon, J., & Mackinnon, J. (2010). Lemurs of Madagascar (Conservation International).

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Fonseca, G. A. B. da, & Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature, 403, 853–858.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., & Wagner, H. (2018) vegan: Community Ecology PackageR package version 2.5-2https://CRAN.R-project.org/package=vegan

Olivieri, G., Zimmermann, E., Randrianambinina, B., Rasoloharijaona, S., Rakotondravony, D., Guschanski, K., & Radespiel, U. (2007). The ever-increasing diversity in mouse lemurs: Three new species in north and northwestern Madagascar. Molecular Phylogenetics and Evolution, 43, 309–327.

Pebesma, E. J., & Bivand, R. S. (2010). Using rainfall radar data to improve interpolated maps of dose rate in the Netherlands. The Science of the total environment, 409(2), 123–133.

De Queiroz, K. (2005). A unified species concept and its consequences for the future of taxonomy. Proceedings of the California Academy of Sciences, 56(June), 196–215.

De Queiroz, K. (2007). Species concepts and species delimitation. Systematic Biology, 56(6), 879–886.

Quéméré, E., Amelot, X., Pierson, J., Crouau-Roy, B., & Chikhi, L. (2012). Genetic data suggest a natural prehuman origin of open habitats in northern Madagascar and question the deforestation narrative in this region. Proceedings of the National Academy of Sciences, 109, 13028–13033.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Radespiel, U., Lutermann, H., Schmelting, B., Bruford, M. W., & Zimmermann, E. (2003). Patterns and dynamics of sex-biased dispersal in a nocturnal primate, the grey mouse lemur, Microcebus murinus. Animal Behaviour, 65, 709–719.

Radespiel, U., Olivieri, G., Rasolofoson, D. W., Rakotondratsimba, G., Rakotonirainy, O., Rasoloharijaona, S., & Randrianarison, R. M. (2008). Exceptional diversity of mouse lemurs (Microcebus spp.) in the Makira region with the description of one new species. American Journal of Primatology. 70, 1033–1046.

Radespiel, U., Ratsimbazafy, J. H., Rasoloharijaona, S., Raveloson, H., Andriaholinirina, N., Rakotondravony, R., & Randrianambinina, B. (2012). First indications of a highland specialist among mouse lemurs (Microcebus spp.) and evidence for a new mouse lemur species from eastern Madagascar. Primates, 53, 157–170.

Rakotoarisoa, J.-E., Raheriarisena, M., & Goodman, S. M. (2013a). Late quaternary climatic vegetational shifts in an ecological transition zone of

northern Madagascar: Insights from genetic analyses of two endemic rodent species. Journal of Evolutionary Biology, 26(5), 1019–1034.

Rakotoarisoa, J.-E., Raheriarisena, M., & Goodman, S. M. (2013b). A phylogeographic study of the endemic rodent Eliurus carletoni (Rodentia: Nesomyinae) in an ecological transition zone of northern Madagascar. Journal of Heredity, 104(1), 23–35.

Rakotondravony, R., & Radespiel, U. (2009). Varying patterns of coexistence of two mouse lemur species (Microcebus ravelobensis and M. murinus) in a heterogeneous landscape. American Journal of Primatology. 71, 928–938.

Rambaut, A., Drummond, A. J., & Suchard, M. (2014). Tracer, v1.6. Rannala, B., & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics, 164, 1645–1656.

Rannala, B., & Yang, Z. (2013). Improved reversible jump algorithms for Bayesian species delimitation. Genetics, 194, 245–253.

Rasoloarison, R. M., Goodman, S. M., & Ganzhorn, J. U. (2000). Taxonomic revision of mouse lemurs (Microcebus) in the western portions of Madagascar. International Journal of Primatology. 21, 963–1019.

Rasoloarison, R. M., Weisrock, D. W., Yoder, A. D., Rakotondravony, D., & Kappeler, P. M. (2013). Two new species of mouse lemurs (Cheirogaleidae: Microcebus) from eastern Madagascar. International Journal of Primatology. 34, 455–469.

Raxworthy, C. J., & Nussbaum, R. A. (1995). Systematics, speciation and biogeography of the dwarf chameleons (Brookesia; Reptilia, Squamata,

Chamaeleontidae) of northern Madagascar. Journal of Zoology. 235, 525–558.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology. 61, 539–542.

Salmona, J., Heller, R., Quéméré, E., & Chikhi, L. (2017). Climate change and human colonization triggered habitat loss and fragmentation in Madagascar. Molecular Ecology, 26(19), 5203–5222.

Schmid, J., & Kappeler, P. M. (1994). Sympatric mouse lemurs (Microcebus spp.) in western Madagascar. Folia Primatol. 63, 162–170.

Schneider, N., Chikhi, L., Currat, M., & Radespiel, U. (2010). Signals of recent spatial expansions in the grey mouse lemur (Microcebus murinus). BMC Evolutionary Biology. 10, 105.

Schwitzer, C., Mittermeier, R., Johnson, S., Donati, G., Irwin, M., Peacock, H., ... Chikhi, L. (2014). Averting lemur extinctions amid Madagascar's political crisis. Science, 343, 842–843.

Serrat, M. A. (2014). Environmental temperature impact on bone and cartilage growth. Comprehensive Physiology. 4, 621–655.

Seutin, G., White, B. N., & Boag, P. T. (1991). Preservation of avian blood and tissue samples for DNA analyses. Canadian Journal of Zoology. 69, 82–90.

Sgarlata, G. M., Salmona, J., Aleixo-Pais, I., Rakotonanahary, A., Sousa, A. P., Kun-Rodrigues, C., ... Chikhi, L. (2018). Genetic differentiation and demographic history of the northern rufous mouse lemur (Microcebus

tavaratra) across a fragmented landscape in northern Madagascar. International Journal of Primatology. 1–25.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., & Li, W. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology.7, 539.

Simons, E. L., Burney, D. A., Chatrath, P. S., Godfrey, L. R., Jungers, W. L., & Rakotosamimanana, B. (1995). AMS 14C dates for extinct lemurs from caves in the Ankarana Massif, northern Madagascar. Quaternary Research. 43, 249–254.

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22, 2688–2690.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and postanalysis of large phylogenies. Bioinformatics. 30, 1312–1313.

Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., ... Dimitrov, D. (2018). Finding Evolutionary Processes Hidden in Cryptic Species. Trends in Ecology & Evolution, 33(3), 153–163.

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. Proceedings of the National Academy of Sciences. 114, 1607–1612.

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., & Watson, J. E. M. (2016). Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. Nature Communications. 7, 12558.

Vieilledent, G., Grinand, C., Rakotomalala, F. A., Ranaivosoa, R., Rakotoarijaona, J.-R., Allnutt, T. F., & Achard, F. (2018). Combining global tree cover loss data with historical national forest-cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. Biological Conservation, 222, 189–197.

Virah-Sawmy, M., Willis, K. J., Gillson, L., & Williams, J. (2010). Evidence for drought and forest declines during the recent megafaunal extinctions in Madagascar. Journal of Biogeography. 37, 506–519.

Waeber, P.O., Schuurman, D., & Wilmé, L. (2018). Madagascar's rosewood (Dalbergia sp.) stocks as a political challenge (PeerJ Inc.).

Waeber, P. O., Wilmé, L., Mercier, J.-R., Camara, C., & Lowry, P. P., II (2016). How effective have thirty years of internationally driven conservation and development efforts been in Madagascar? PLOS ONE, 11(8):e0161115.

Weisrock, D. W., Rasoloarison, R. M., Fiorentino, I., Ralison, J. M., Goodman, S. M., Kappeler, P. M., & Yoder, A. D. (2010). Delimiting species without nuclear monophyly in Madagascar's mouse lemurs. PLOS ONE, 5):e9883.

Wiens, J. J., Ackerly, D. D., Allen, A. P., Anacker, B. L., Buckley, L. B., Cornell, H. V., ... Stephens, P. R. (2010). Niche conservatism as an emerging principle in ecology and conserva- tion biology. Ecology Letters, 13(10), 1310–1324.

Wilmé, L., Goodman, S. M., & Ganzhorn, J. U. (2006). Biogeographic evolution of Madagascar's Microendemic Biota. Science, 312, 1063–1065.

Wright, S. (1943). Isolation by distance. Genetics, 28, 114–138.

Yang, Z. (2002). Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. Genetics, 162, 1811–1823.

Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. Current Zoology. 61, 854–865.

Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. Proceedings of the National Academy of Sciences, 107, 9264–9269.

Yang, Z., & Rannala, B. (2017). Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses. Molecular Ecology. 26, 3028–3036.

Yoder, A. D., Burns, M. M., & Génin, F. (2002). Molecular evidence of reproductive isolation in sympatric sibling species of mouse lemurs. International Journal of Primatology. 23, 1335–1343.

Yoder, A. D., Olson, L. E., Hanley, C., Heckman, K. L., Rasoloarison, R., Russell, A. L., ... Goodman, S. M. (2005). A multidimensional approach for detecting species patterns in Malagasy vertebrates. Proceedings of the National Academy of Sciences, 102, 6587–6594.

Yoder, A. D., Rasoloarison, R. M., Goodman, S. M., Irwin, J. A., Atsalis, S., Ravosa, M. J., & Ganzhorn, J. U. (2000). Remarkable species diversity in Malagasy mouse lemurs (primates, Microcebus). Proceedings of the National Academy of Sciences, 97, 11325–11330.

Zimmermann, E., Cepok, S., Rakotoarison, N., Zietemann, V., & Radespiel, U. (1998). Sympatric mouse lemurs in north-west Madagascar: A new rufous mouse lemur species (Microcebus ravelobensis). Folia Primatologica, 69, 106–114.

5.8 Supplementary Information

Table S1. Best-fit partitioning scheme and nucleotide substitution model for each mtDNA dataset. RaxML accepts one unique model of evolution for all partitions, whereas a model of evolution can be specified for each partition in MrBayes.

	MrBayes		RaxML	
	partitions	model	partitions	model
cytb	1-1140; 2-1140; 3-1140	SYM+I+G; HKY+I+G; GTR+I+G	1-1140, 2-1140; 3-1140	GTRGAMMAI
cox2	1-684, 2-684, 3-684	GTR+I+G	1-684, 2-684, 3-684	GTRCAT
cox2+cytb	1-684, 685-1824; 686-1824, 2-684; 3-684; 687-1824	GTR+I+G; GTR+I+G; HKY+G; GTR+G	1-684, 685-1824; 686-1824, 2-684; 3-684; 687-1824	GTRGAMMAI

Table S2. Combinations of prior values used in BPP analyses. Θ = effective population size; τ = divergence time at the root of the species tree. α and β represent the parameters of the inverse gamma prior distribution.

θ(α,β)			τ(α,β)		
3	0.002	Small population size	3	0.02	deep divergence
3	0.002	Small population size	3	0.002	recent divergence
3	0.02	Medium population size	3	0.02	deep divergence
3	0.02	Medium population size	3	0.002	recent divergence
3	0.2	Large population size	3	0.02	deep divergence
3	0.2	Large population size	3	0.002	recent divergence

Table S3. Description of the grouping tests for morphological analyses. AMBRE (Montagne d'ambre) and ANKAR (Ankarana) individuals are considered the morphological reference for, respectively, *Microcebus arnholdi* and *Microcebus tavaratra*. Only some of the individuals used for morphological analyses were genotyped for cytb and cox2.

Group	Species Comparison	Sites	N
I	*M. tavaratra* - *M. arnholdi*	ANTSO (1), ANALA (10), ANALF (5), AMBRE (18), ANKA (2), ORO (2), AMP (1), ANKAR (7), AMBA (1), SAL (4), BIN (3), ANKI (1), ANTSA (1), BEZ (5)	61
II	*M. tavaratra* - *M. arnholdi*	BIN (27), AMBRE (18), ANKAR (7)	52
III	*M. arnholdi* - *Microcebus* sp.	ANTSO (1), AMBRE (18), BIN (3), SAL (4), BEZ (5), ANKI (1), ANTSA (1), AMBA (1), ANALV (5), AMBO (3), ANDG (1), MAL (1), ANDR (2), ANTA (1)	47
IV	*M. arnholdi* - *Microcebus* sp. - *M. mamiratra*	ANTSO (1), AMBRE (18), BIN (3), SAL (4), BEZ (5), ANKI (1), ANTSA (1), AMBA (1), ANALV (5), AMBO (3), ANDG (1), MAL (1), ANDR (2), ANTA (1), AMBK (1)	48

Table S4. Results of species delimitation tests for species pairs that showed lower p-distance than *Microcebus* sp.– *M. arnholdi*. a) *M. murinus*, *M. ganzhorni* and *M. manitatra*; b) *M. jollye*, *M. simmonsis* and *M. boraha*; c) *M. danfossi*, *M. bongolavensis* and *M. ravelobensis*. We performed the BPP analyses on the combined mtDNA dataset (cytb+cox2). The first column describes the combination of priors (effective population size and time of divergence) used in the analyses. alg0 and alg1 show the results obtained, respectively, with either algorithm0 or algorithm1. alg0-random and alg1-random present the results for the randomized dataset. The values are the posterior probability of the three clades being distinct species.

a)

	alg0	alg0-random	alg1	alg1-random
small pop. & deep div.	1	0	1	0
small pop. & recent div.	0.99	0	0.99	0
medium pop. & deep div.	1	0	1	0
medium pop. & recent div.	0.99	0	0.99	0
large pop. & deep div.	0.05	0	0.99	0
large pop. & recent div.	0.07	0	0.07	0

b)

	alg0	alg0-random	alg1	alg1-random
small pop. & deep div.	1	0	1	0
small pop. & recent div.	1	0	1	0
medium pop. & deep div.	1	0	1	0
medium pop. & recent div.	1	0	1	0
large pop. & deep div.	0.99	0	1	0
large pop. & recent div.	0.99	0	0.99	0

c)

	alg0	alg0-random	alg1	alg1-random
small pop. & deep div.	1	0	1	0
small pop. & recent div.	1	0	1	0
medium pop. & deep div.	1	0	1	0
medium pop. & recent div.	1	0.66	1	0.66
large pop. & deep div.	0.77	0	1	0
large pop. & recent div.	0.69	0.07	0.70	0.08

Table S5. Cluster assignment of *Microcebus* individuals based on morphological data. In Group I (a) and II (b), cluster 1 and 2 refer, respectively, to *M. arnholdi* and *M. tavaratra* individuals. In Group III (c), there is no match between morphological clusters and clade identity. N: number of individuals per site.

a)

Sites	Cluster 1	Cluster 2	N
AMBA	0	1	1
AMBRE	0.94	0.06	18
AMP	0	1	1
ANALA	0	1	10
ANALF	0	1	5
ANKA	0	1	2
ANKAR	0	1	7
ANKI	1	0	1
ANTSO	1	0	1
BEZ	0.6	0.4	5
BIN	1	0	3
ORO	0	1	2
SAL	1	0	4

b)

Sites	Cluster 1	Cluster 2	N. tot
AMBRE	1	0	18
ANKAR	0	1	7
BIN	0.11	0.89	27

c)

Sites	Cluster 1	Cluster 2	Cluster 3	N. tot
AMBA	1	0	0	1
AMBO	0	0	1	3
AMBRE	0.5	0	0.5	18
ANALV	0	0	1	5
ANDR	1	0	0	2
ANKI	0	0	1	1
ANTA	1	0	0	1
ANTSA	0	0	1	1
ANDG	0	1	0	1
ANTSO	1	0	0	1
BEZ	0	0	1	5
BIN	0	0	1	3
MAL	1	0	0	1
SAL	0	0	1	4

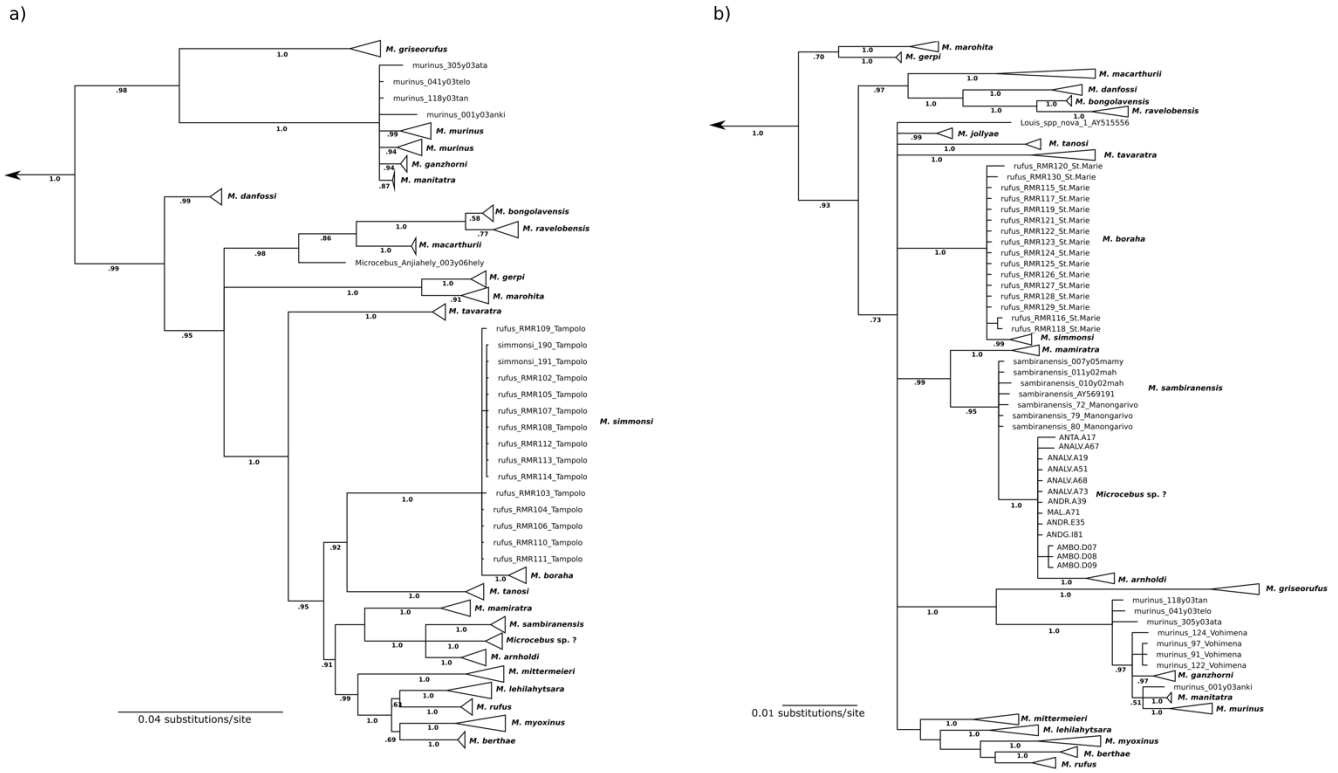


Figure S1. Bayesian phylogenetic analyses of *Microcebus* genus using mtDNA data. a) cytb dataset; b) cox2 dataset. We constructed the phylogenetic trees using MrBayes. Numbers below branches show posterior probabilities. The arrow refers to the outgroups: *Cheirogaleus major*, *Cheirogaleus medius*, *Cheirogaleus crossleyi* and *Mirza coquereli*. Cox2 phylogenetic tree showed less resolving power as observed, for instance, in the case of *M. sambiranensis*, *M. arnholdi* and *Microcebus* sp.

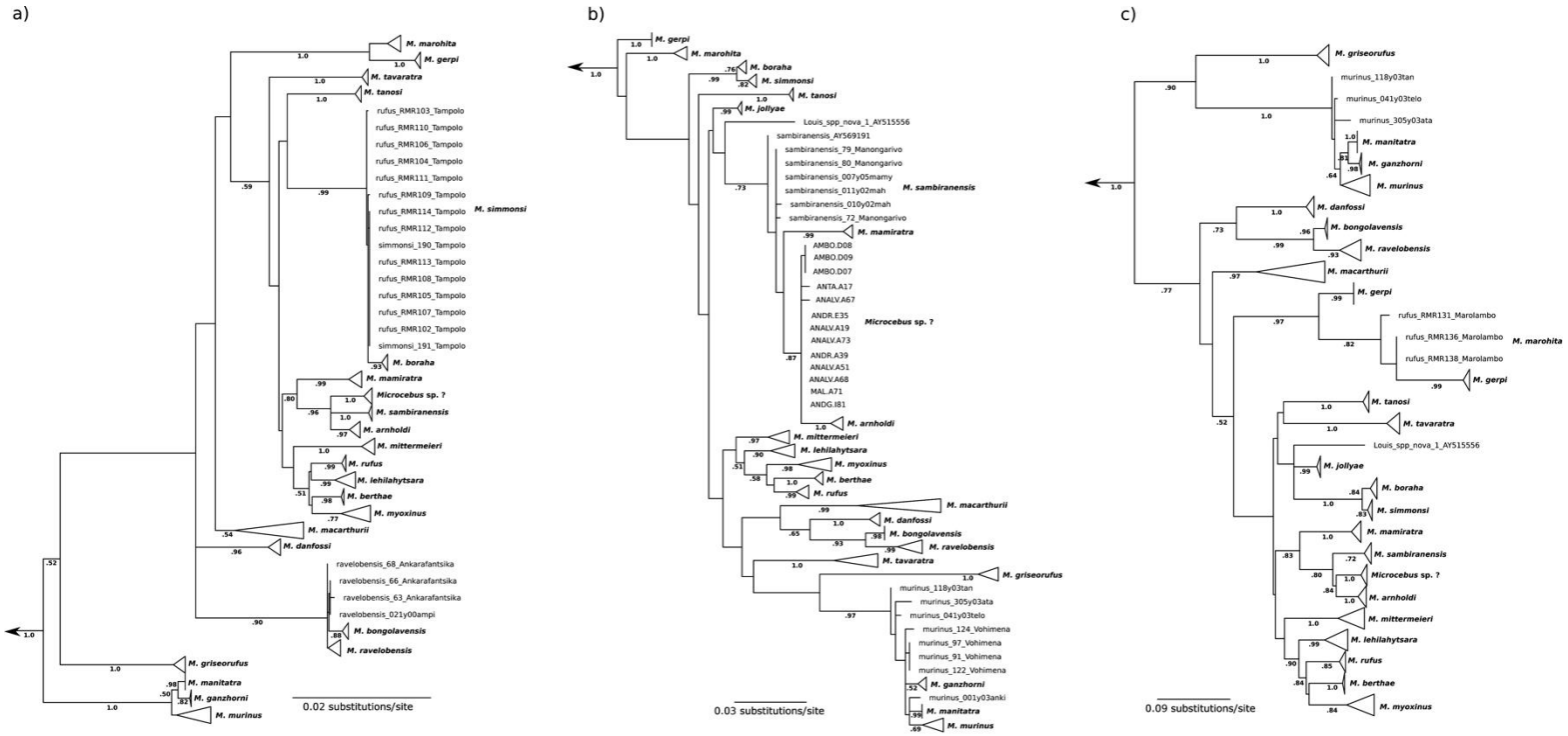


Figure S2. Maximum likelihood phylogenetic analyses of *Microcebus* genus using mtDNA data. a) cytb dataset; b) cox2 dataset; c) cox2+cytb dataset. We constructed phylogenetic trees in RaxML. Numbers are branch support values. The arrow refers to the outgroups: *Cheirogaleus major*, *Cheirogaleus medius*, *Cheirogaleus crossleyi* and *Mirza coquereli*.

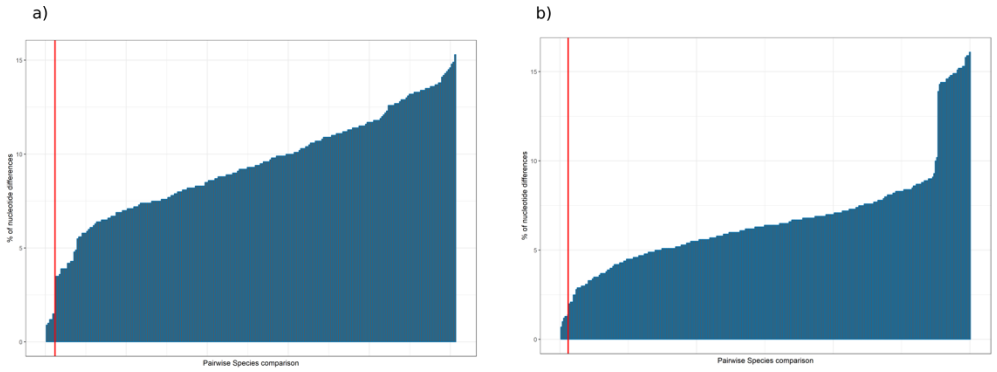


Figure 8 Pairwise proportion of nucleotide differences between *Microcebus* species pairs. a) cytb dataset; b) cox2 dataset. Each bar represents the average % of nucleotide differences between two *Microcebus* species. The vertical red line marks the *Microcebus* sp. – *M. arnholdi* comparison. The plot shows that several currently recognized species comparisons have lower genetic divergence than the new *Microcebus* clade and *M. arnholdi*.

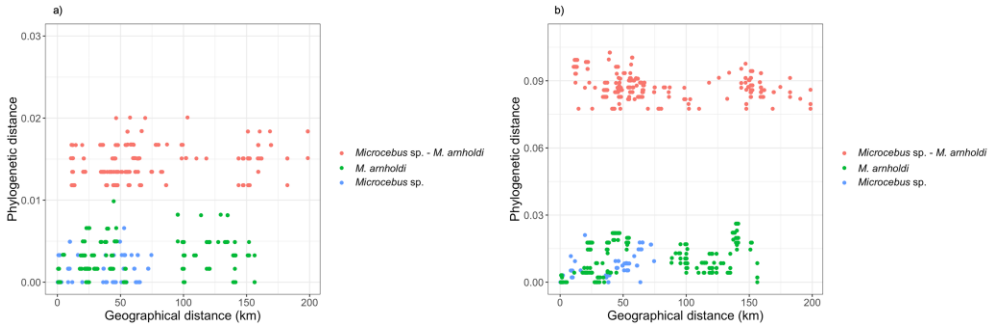


Figure S3. Relationship between phylogenetic and geographic distance for *M. arnholdi* – *Microcebus* sp. comparison. a) cox2 dataset; b) cytb dataset. Phylogenetic distance is measured as patristic distance, which is the sum of branch lengths on the path between individuals. Overall, we observed higher inter- than intra-clades phylogenetic distances (Kruskal-Wallis rank sum test: X2=244.09, df=2, P<0.001; X2=272.7, df=2, P<0.001).

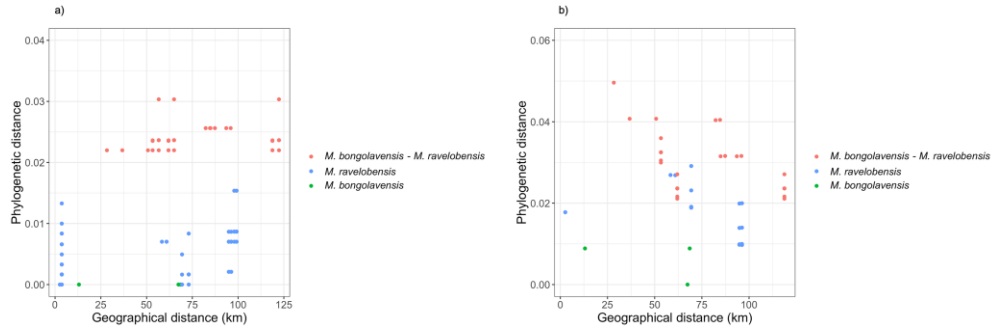


Figure S5. Relationship between phylogenetic and geographic distance for *M. bongolavensis*. – *M. ravelobensis* comparison. a) cox2 dataset; b) cytb dataset. Phylogenetic distance is measured as patristic distance, which is the sum of branch lengths on the path between individuals. Overall, we observed inter- than intra-clades phylogenetic distances (Kruskal-Wallis rank sum test: X2=88.15, df=2, P<0.001; X2=76.413, df=2, P<0.001).

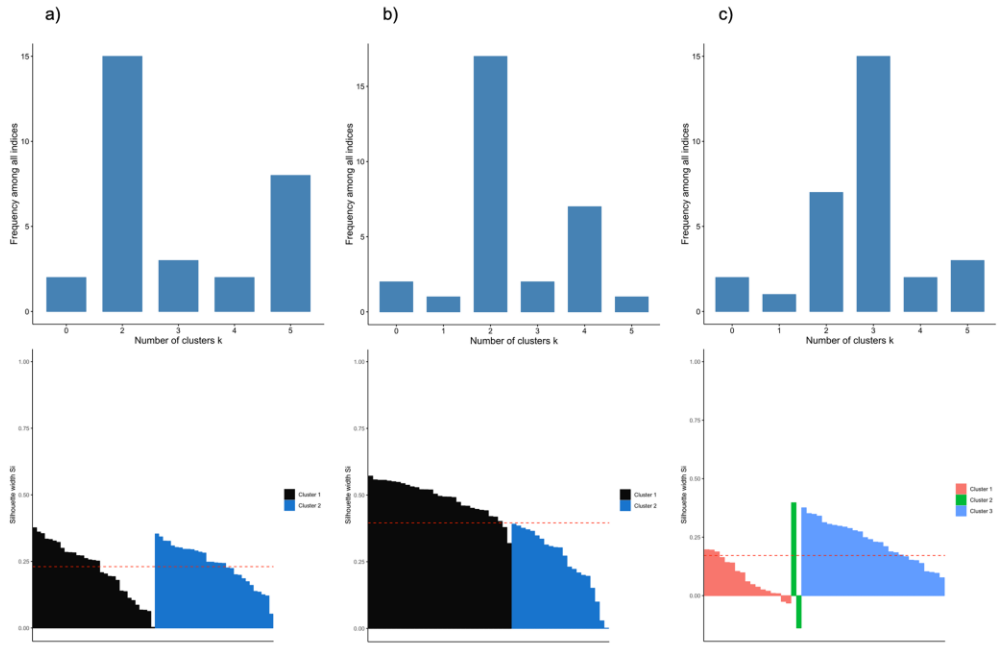


Figure S6. Selecting the number of clusters for K-means analysis on *Microcebus* morphological data. a) Group I; b) Group II; c) Group III. The first row shows the frequency with which each K value has been selected as the best number of clusters. The second row shows the results of the silhouette validation analysis. The red dashed line shows the average silhouette value.

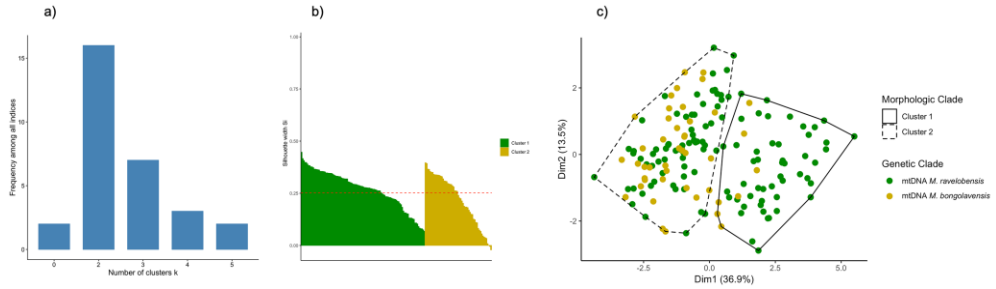


Figure S7. K-means analysis on *Microcebus bongolavensis* – *Microcebus ravelobensis*. a) Frequency with which each K value was selected as the best number of clusters; b) silhouette analysis; c) K-means plot setting K=2 as the best n° of cluster. The analysis does not show a clear morphological divergence between the two sister species.

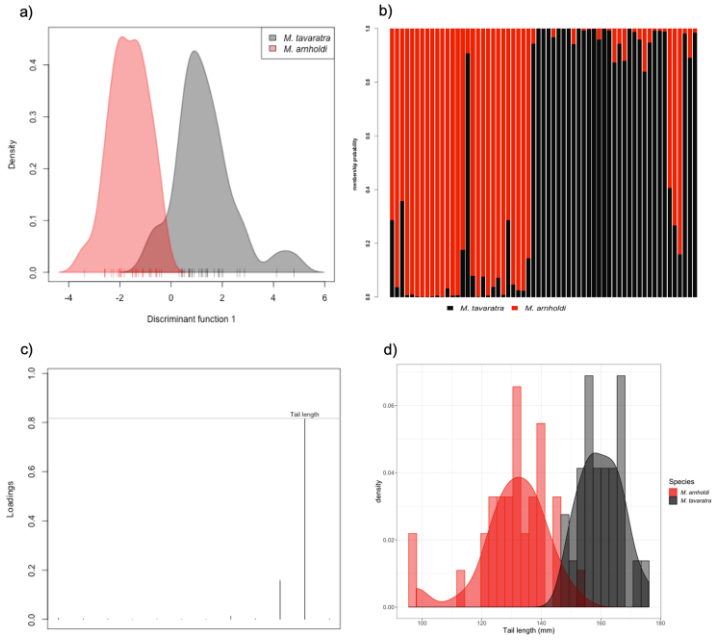


Figure S8. Discriminant Analysis of Principal Components between *Microcebus tavaratra* and *Microcebus arnholdi* from Northern Madagascar. a) Density distributions of the first discriminant component show clear morphological differences between the two species; b) barplot shows the individual assignment probability (based on morphology) to groups defined a priori (i.e. mtDNA clade); c) loading plot show that "Tail length" is the morphological variable that more significantly contribute to the morphological structure of *M. tavaratra* and *M. arnholdi*; d) Tail length distributions of *M. tavaratra* and *M. arnholdi* are significantly different (Two-sample Kilgomorov-Smirnov test: D = 0.97, P<0.001).

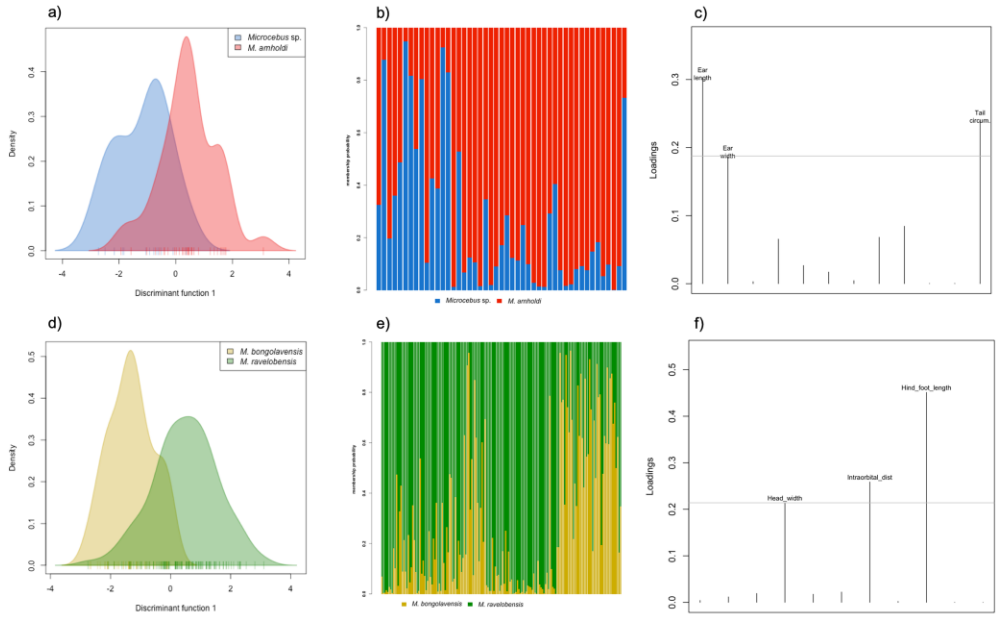


Figure S9. Discriminant Analysis of Principal Components for *Microcebus* sp.-*M. arnholdi* and *M. bongolavensis-M. ravelobensis*. a) and d) Density distributions of the first discriminant component; b) and e) barplots show the individual assignment probability (based on morphology) to groups defined a priori; c) and f) loading plots showing the morphological variable that more significantly contribute to the morphological structure. Any of the morphological variables identified in *Microcebus* sp.-*M. arnholdi* comparison are significantly different. Instead, Intraorbital distance (Two-sample Kolmogorov-Smirnov test: D= 0.5, P<0.001) and Hind foot length (Two-sample Kolmogorov-Smirnov test: D= 0.5, P<0.001) are significantly different in the *M. bongolavensis-M. ravelobensis* comparison.

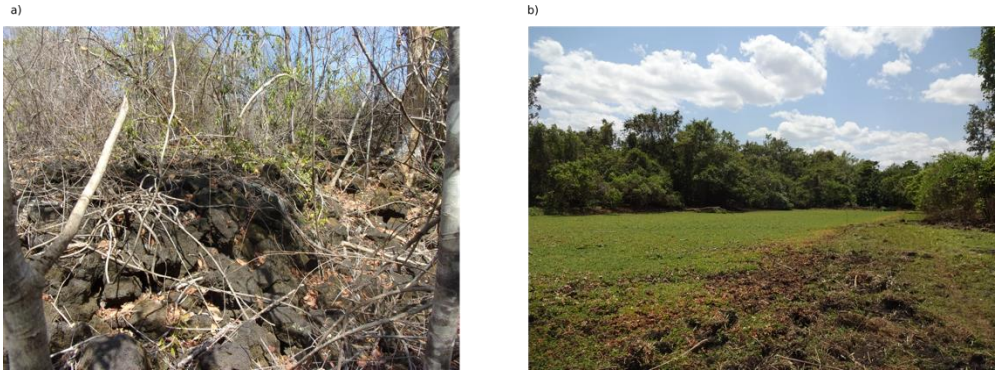


Figure S10. Vegetation diversity in Antsoroby forest (ANTSO), Northern Madagascar. a) dry vegetation, covering most of the forest; b) moist vegetation, close to the site in which *Microcebus arnholdi* individual was sampled.

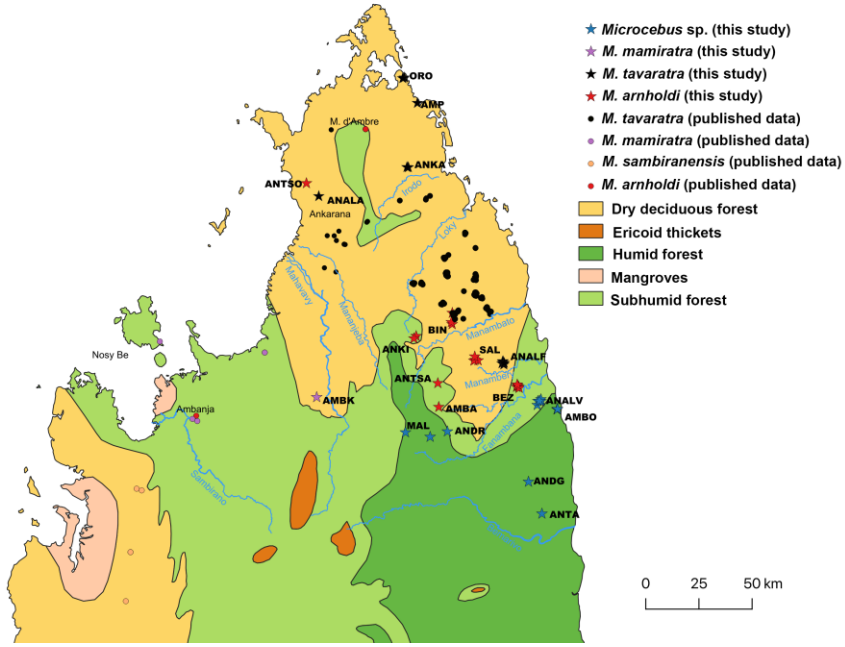


Figure S11. Northern *Microcebus* spp. distribution compared with Northern bioclimatic regions. The northern bioclimatic zones recognized today are not sufficient to comprehensively explain the current distribution of Northern *Microcebus* species. This is the case for the humid associated *M. arnholdi* populations located in BIN, SAL and ANTSO forests, which are placed in the dry bioclimatic zone. Ecoregions are defined according to Schatz G.E., 2000

The genomic diversity of the *Eliurus* genus in northern Madagascar with a putative new species

Gabriele Maria Sgarlata[1], Emmanuel Rasolondraibe[3], Barbara Le Pors[1], Jordi Salmona[1,2], Tantely Ralantoharijaona[3], Ando Rakotonanahary[3], Sophie Manzi[2], Amaya Iribar-Pelozuelo, John Rigobert Zaonarivelo[4], Nicole Volasoa Andriaholinirina[3], Solofonirina Rasoloharijaona[3], Lounès Chikhi[1, 2]

1. Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, 2780-156 Oeiras, Portugal;
2. Laboratoire Evolution & Diversité Biologique (EDB UMR 5174), CNRS, IRD, UPS, Université de Toulouse Midi-Pyrénées, Toulouse, France;
3. Département de Biologie Animale et Ecologie, Faculté des Sciences, Université de Mahajanga, Mahajanga, Madagascar
4. Département des Sciences de la Nature et de l'Environnement, Université d'Antsiranana, 201 Antsiranana, Madagascar

AUTHOR CONTRIBUTIONS: GMS, LC and JS designed the project. GMS collected part of the samples, performed DNA extraction for most samples and analyzed genomic data. BLP and ER led the fieldwork and collected most of the samples. JS collected part of the samples. TR, AR contributed in the collection of samples. SM and AI-P prepared the library for genomic sequencing. SR, NVA and JRZ contributed to the organization of the fieldwork. GMS wrote the manuscript and LC revised it.

Highlights

- Validate and clarify tuft-tailed rat species (*Eliurus* spp.) phylogeny with the use of genome-wide data
- Report the discovery of a new putative tuft-tailed rat species in northern Madagascar
- Document frequent occurrences of the typically dry-associated *E. myoxinus* in the humid forests of northern Madagascar. Since most *Eliurus* species are adapted to humid habitats, and given that the two *Eliurus* lineages closely-related to *E. myoxinus* are humid-associated species, we propose that *E. myoxinus* was originally adapted to humid habitat and that, later on, the increasing aridification of forests during the late Pleistocene – mid Holocene, promoted the adaptation of southern *E. myoxinus* populations to dry habitats.
- Document frequent cases of *Eliurus* species co-occurrence (sympatry and sintopy). The percentage of surveyed sites with *Eliurus* species co-occurrence ranged between 38% and 90%.
- Confirm previous scenarios on the history of humid forest cover and connectivity in northern Madagascar using genetic data.

5.9 Abstract

Madagascar is known to be a hotspot of biodiversity, meaning that it contains extraordinarily high level of species richness and endemism, while being severely threatened by habitat loss and fragmentation (HL&F). In fact, several studies have estimated that, in the last 60 years, Madagascar has lost about 50% of forest cover. In front of such threat to biodiversity, conservation effort can be directed, for instance, in the documentation of species that are still unknown to science, or in investigating how species respond to HL&F. The tufted-tail rats genus (*Eliurus* spp.) is the most speciose genus of endemic rodents in Madagascar, with 13 recognized taxa, broadly distinguished in dry versus humid associated species. The large species diversity of the genus and their association to specific habitat types make them a suitable model for investigating species evolution, as well as response to HL&F. In the present study, we sampled 125 *Eliurus* individuals across the forests of northern Madagascar, a region where dry and humid forest types can be found. We analysed mitochondrial DNA (mtDNA) and nuclear genomic (RAD-seq) diversity, and identified an undescribed *Eliurus* taxon (*Eliurus sp. nova*). We tested the hypothesis of a new *Eliurus* species using several approaches: i) DNA barcoding; ii) phylogenetic inferences; iii) species delimitation tests based on the Multi-Species Coalescent (MSC) model, iv) genealogical discordance (*gdi)* index; v) the *ad-hoc* test of isolation-by-distance within *versus* between sister-taxa and vi) comparisons of %GC content. All analyses support the recognition of the presently undescribed lineage as a separate species. In addition, we have also shown that *Eliurus myoxinus*, a species known from the dry forests of the west, is mostly found in humid forests in northern Madagascar. In conclusion, we discuss the implications of such findings in the context of *Eliurus* species evolution and diversification, and use the distribution of northern *Eliurus* species as a proxy for reconstructing past changes in forest cover and vegetation type in northern Madagascar.

5.10 Introduction

The endemic rodents of Madagascar (subfamily *Nesomyinae*) represent a diverse taxonomic group of the family Muridae, whose morphological diversity has led to the recognition of nine genera (Musser and Carleton, 2005) characterised by diverse life-history traits, such as scansorial, semi-fossorial or saltatorial behaviours. Their monophyly is now recognised, although debated over decades due to the lack of sufficiently powerful data and the morphological similarity with non-Malagasy rodents (Jansa et al., 1999; Michaux et al., 2001; Jansa and Weksler, 2004). Genetic data have indeed established that Malagasy rodents came from Africa during a single colonization event (Steppan and Schenk, 2017; Jansa and Carleton, in press), which has been dated between 15 and 30 Ma by (Poux et al., 2005) and between 12.8 and 15.6 Ma by (Steppan and Schenk, 2017). Since then, the *Nesomyinae* have diversified across the diverse ecotypes of Madagascar.

Most *Nesomyinae* genera are composed of few species (either one or four), with the exception of the *Eliurus* genus which, with 13 currently recognised species, is by far the most speciose genus of *Nesomyinae* (Jansa et al., 2019). The *Eliurus* genus was initially characterised by Alphonse Milne-Edwards in 1885 (Milne Edwards, 1885), with the description of *E. myoxinus* in western Madagascar. During the following decades, new *Eliurus* species were mainly described from the humid forests of eastern Madagascar (Carleton, 1994; Carleton and Goodman, 1998; Carleton, 2003). Recently, new *Eliurus* species from the west and north of Madagascar have been added to the genus (Carleton et al., 2001; Goodman et al., 2009; Jansa et al., 2019). In the last twenty-five years, field surveys and museum-specimen based research have greatly contributed to a better understanding of endemic rodent biodiversity in Madagascar (Carleton, 1994; Jansa et al., 1999; Goodman and Rasolonandrasana, 1999; Carleton et al., 2001; Jansa and Carleton, 2003; Carleton and Goodman, 2007; Goodman et al., 2007, 2009;

S. M. Goodman et al., 2013), leading to a seven-fold increase in *Eliurus* specimen collection and the description of five new species (see Fig.1 in Jansa et al., 2019). The systematic growth of the *Eliurus* genus over the last two decades is to a large extent related to the growing use of genetic data, mostly few mitochondrial and nuclear markers (e.g., Jansa et al., 1999; Goodman et al., 2009; Jansa et al., 2019).

While the use of genetic data may be problematic, as it has led to a likely inflation in the number of species on other Malagasy mammals, such as mouse lemurs and sportive lemurs (e.g., Tattersall, 2007; Tattersall and Cuozzo, 2019), morphological data may not always be sufficient to identify species, if their phenotypic differences are small, difficult to quantify or the result of limited sampling, even when divergence between sister-species is not recent (Bickford et al., 2007; Struck et al., 2018; Chenuil et al., 2019). For instance, one of the diagnostic morphological characters that Carleton, (1994) proposed for distinguishing the then newly discovered *E. ellermani* species from the closely-related *E. tanala* was the possession of a 'tail brush completely dark to the tip'. The study of Jansa et al., (2019), which integrated morphological and genetic data from a larger number of specimens, showed that individuals genetically identified as *E. ellermani* had in most cases a white tail tuft, and rarely a dark tipped tail. One could also use information on geographic distributions for taxonomic identification of observed and sampled individuals. However, recognised *Eliurus* species are i) characterised by large and scattered species distributions, and ii) often occurs in co-occurrence or syntopy with one or more *Eliurus* species (Carleton, 1994; Marquart, 2014; Jansa et al., 2019).

In the present study, we collected samples from 125 *Eliurus* individuals in several forest fragments of northern Madagascar, a transition region between the humid ecoregion in the east and the dry ecoregion in the west of the island, and assessed their mitochondrial (mtDNA) and genomic diversity. Up to now, genetic data on *Eliurus* were limited to two mtDNA and four nuclear loci (Jansa et al., 1999; Goodman et al., 2009; Tysdal and Jansa,

2014; Jansa et al., 2019). With the use of a Restriction site-associated DNA sequencing (RAD-seq) approach, which generates a reduced but wide representation of genomic diversity, we also aimed at validating and clarifying the *Eliurus* species tree topology. We performed several phylogenetic inferences and species delimitation tests, taking into account genomic and geographic variability within each taxon, as well as gene tree variation. We identify a new monophyletic and divergent lineage which we hypothesize to be a new *Eliurus* species.

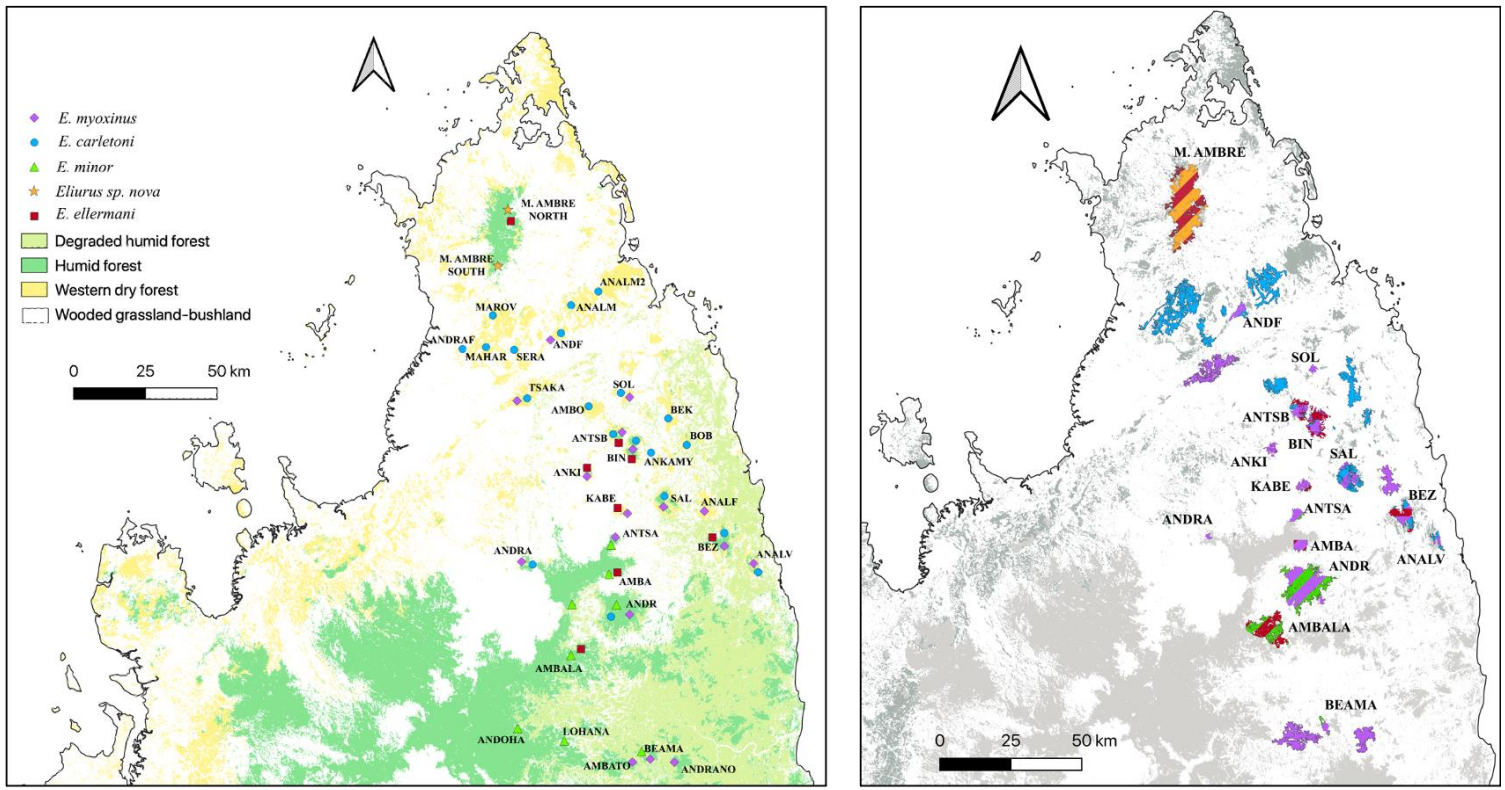


Figure 1. *Eliurus* individuals sampled in northern Madagascar. In this figure we present the sampling sites of the five *Eliurus* taxa identified in the present study: *E. myoxinus*; *E. carletoni*; *E. minor*; *E. ellermani* and *Eliurus sp.* nova. Detailed map for each species can be found in Fig. S18-22. The map on the right shows the localities in which each species has been identified, highlighting the sites of co-occurrence between several *Eliurus* species (see Table S5). Yellow: dry forests; Dark green: humid forests; Light green: degraded humid forests.

5.11 Materials and Methods

5.11.1 Study region

We surveyed 35 forest sites between 2010 and 2018 in northern Madagascar, a region considered to be a transition zone between the humid and dry ecoregions of eastern and western Madagascar, respectively. The landscape of northern Madagascar is characterized by forest fragments of diverse habitat types (dry deciduous, transition, humid evergreen, mountainous) separated by a matrix of savannah, grassland and agricultural lands. This region is crossed by six main rivers which further divide northern Madagascar in inter-river systems: the Irodo river, the Loky river, the Manambato river, the Manambery river, the Fanambana river and the Bemarivo river (Fig. 1).

North of the Loky river

At the north of the Loky river, there are mainly fragmented dry forests, frequently surrounded by riparian corridors along narrow canyons or by more extended riparian forests in proximity of the Irodo river. In this region, we mainly visited forest sites within three major massifs: Ankarana, Analamerana and Andrafiamena-Andavakoeira.

The Special Reserve (SR) of Ankarana (IUCN category IV) is a limestone massif of (~250 km^2) covered mostly by pinnacle karst (“tsingy”) and dry forests. However, a large diversity of vegetation types can be also observed, for instance, moist primary seasonal forest and specialized xerophytic scrub (Rossi, 1974; Fowler et al., 1989). Surface water is rare and only present along canyons, where underground rivers get to the surface. The thick soil on the basalt formation of the massif, the underground water and the protection from drying winds create optimal conditions for the persistence

of moist and humid vegetation within canyons (Fowler et al., 1989; Nicoll and Legrand, 1989). There are also humid deciduous forests on the basaltic soil of the northern side of massif, extending down from Montagne d'Ambre (Goodman et al., 2021, pp. 558–572). The surface area of the moist forests is limited to 20 – 50 km^2, representing a 'refugium' during the dry season for some animal taxa (e.g., lemurs) that, then during the wet season, also use more extensively the deciduous forests (Hawkins et al., 1990). The SR of Ankarana is currently managed by Madagascar National Parks (Goodman et al., 2021, pp. 558–572).

The SR of Analamerana (~470 km^2) is characterised by the same vegetational types as Ankarana Special Reserve. It is drier than Ankarana forests and lies entirely on limestone. The dominant vegetation is dry deciduous forest. Moist semi-deciduous vegetation can be found near river streams, with tree species typical from rainforest habitats (Hawkins et al., 1990). Two major blocks of forests can be identified within the Analamerana SR, one in the northeast that is well preserved, and one in the southwest which has been heavily fragmented by fires (Goodman et al., 2021, pp. 515–527). The Analamerana SR is currently managed by Madagascar National Parks.

The Andrafiamena-Andavakoeira Protected Harmonious Landscape (IUCN category V; ~730 km^2) is composed of two large forests, Andrafiamena and Andavakoeira differing in geological substrate types: limestone karst and sandstone, respectively (Buřivalová, 2011). As a consequence, also the vegetation occurring in the two forests is different. In fact, (Buřivalová, 2011) sampled large trees in forested habitats found on both substrates and determined that only 4% of the species were shared between the two forests. In particular, Andrafiamena forest has an alkaline soil layer, and is more water deprived and support a more xeric and deciduous vegetation than Andavakoeira forest (Fowler et al., 1989), which occurs on sandstone, where trees are less water deprived and often closely related to the tree species

occurring in the eastern rainforest. The Andrafiamena-Andavakoeira Protected area is managed by the NGO Fanamby.

Loky-Manambato region

The Loky-Manambato (LM) region is a biogeographical transition zone between dry deciduous and humid forests (Goodman and Wilmé, 2006), which is delimited by the Loky and Manambato rivers in the north and south, respectively. This region is crossed by the relatively shallow Manankolana River, bordered by riparian forests along most of its course, and by a national dirt road. It consists of an area of ~2,500 km^2 covered by ~360 km^2 of forests (Goodman et al., 2021, pp. 528–545), fragmented into a dozen major forest patches surrounded by human-altered grasslands, dry scrub and agricultural lands. Most forests are situated at low- to mid-elevations and mostly consist of dry deciduous vegetation. In contrast, some mountain forests on the southern edge (Binara and Antsahabe, plus Bobankora to a lower extent) are covered by a gradient of dry deciduous, transition, humid and ericoid vegetation (Gautier et al., 2006). The LM region is protected as a Protected Harmonious Landscape (IUCN category V) and managed by the Malagasy NGO "Fanamby" since 2005 (Goodman et al., 2021, pp. 528–545).

South Manambato region

In this region, forest fragments are currently un-protected (Goodman et al., 2021). Between the Manambato and the Manambery rivers, there are the lowland dry deciduous forest of Analafiana and the mountain humid forest of Salafaina, which topology and vegetation are similar to the Binara massif of the LM region. Between the Manambery and the Fanambana rivers, there is the evergreen humid forest of Bezavona-Ankirendrina. Ultimately, the southernmost visited forest of Analalava, south of the Fanambana River, is an evergreen humid forest. Little detailed ecological and vegetation information on these forest fragments can be found in the literature. However,

recurrent survey by members of our research team have observed over years, a decrease in habitat amount and quality (Salmona J. and Le Pors B. pers. comm.).

COMATSA region

Within what we call 'COMATSA region', we surveyed sites that are located in the continuous forest of the Corridor of Marojejy- Anjanaharibe Sud-Tsaratanana Nord protected area (COMATSA Nord) and those lying in the small unprotected forest fragments, at the north-east of the protected area (Ankinjanala and Anketrakabe). The vegetation composition of the COMATSA region is poorly studied. What is currently known is that it is mostly covered by continuous medium altitude evergreen forests, which only in few locations of the north-central portion transit to an ericoid vegetation. The forest of the Andravory Massif is isolated from the continuous forest by a 3-4 km stretch of secondary grassland along the Fanambana River (Goodman et al., 2021, pp. 597–611). The Ankinjanala and Anketrakabe forests are disconnected from the north-east portion of the COMATSA protected area by secondary grassland, with only some riparian forests along river streams. The COMATSA Nord protected area is classified as a Natural Resources Reserve (VI IUCN category) and is currently managed by WWF (Goodman et al., 2021, pp. 597–611).

North Marojejy

We visited three sites that are located in the unprotected forests at the north of the Anakarangona river and Marojejy National Park: Lohananjialava, Beamalonkely, Andranomenabe, Ambatomainty, Andohanibemarivo. They are moist semi-deciduous forest fragments separated by secondary grassland. Deforestation in this area is most likely recent.

Montagne d'Ambre

Montagne d'Ambre is an isolated volcanic massif with a peak of altitude at 1475 m. It is placed in a dry ecoregion, but due to its exposition to the trade winds, it receives more precipitation than neighbouring areas. Most of the massif is covered by medium-altitude moist evergreen forest, and only close to the summit the moist vegetation tends to transit to an ericoid vegetation (Nicoll and Legrand, 1989; Raxworthy and Nussbaum, 1994). Montagne d'Ambre is protected under the status of National Park (IUCN category II) and is managed by Madagascar National Parks (Goodman et al., 2021, pp. 500–514).

5.11.2 Study species

The tuft-tailed rats belong to the most speciose genus (*Eliurus*) of Malagasy endemic rodent (subfamily *Nesomyinae*), currently comprising 13 recognised species (see Jansa et al., 2019 for a recent update). They are nocturnal and easily recognizable by a noticeable tuft of long hairs over the distal portion of the tail. They are associated with forested habitats, but some species show tolerance to mild forest disturbance. In particular, trap captures data show the arboreal and scansorial behaviour of this group of rodents, since individuals can be both collected on lianas, trees and on ground (e.g., Marquart, 2014). However, there are some distinctions that can be made among species. For instance, Webb, (1954) observed that western *Eliurus* species are more arboreal than eastern species, rarely descending to the ground.

In the present work, we focus on four *Eliurus* species that have been described in northern and north-eastern Madagascar: *E. carletoni*, *E. ellermani*, *E. minor* and *E. myoxinus* (Fig. 1).

E. carletoni (Goodman et al., 2009) is a relatively large member of the *Eliurus* genus (head-body length: 143 – 150 mm; Goodman et al., 2009) known to occur in the extreme north of Madagascar (Fig. S17). It inhabits the

relatively mesic forest conditions between the pinnacle karsts (called in Malagasy *tsingy*) (Bésairie, 1964; Puy and Moat, 1996; Veress et al., 2008) found at the SR of Ankarana, SR of Analamerana and a nearby zone of metamorphic rock in the PHL-V of Loky-Manambato (Daraina region) (Rakotoarisoa et al., 2010; Rakotoarisoa et al., 2013a). *E. carletoni* can be considered a scansorial forest-dwelling species of dry deciduous habitats and it has been caught in both undisturbed and disturbed forests. Little is known on the ecology of this species. Documented records report an elevation range of occurrence between 20 m and 500 m (Jansa et al., 2019). *E. carletoni*, as also its sister-species *E. antsingy*, is characterised by a tuft that cover 45% – 55% of tail length. The tuft is dark blackish – brown and monochromatic over the entire tail (Goodman et al., 2009; Jansa et al., 2019). The dorsal pelage is typically dark-brown, which contrasts with the grayish-white ventral pelage. The tail length and pilosity distinguish *E. carletoni* from other *Eliurus* species with similarly monochromatic dark-blackish brown tails (*E. myoxinus* and *E. webbi*) (Carleton, 1994; Carleton et al., 2001; Carleton, 2003; Goodman et al., 2009).

E. ellermani (Carleton, 1994) is another large member of the *Eliurus* genus (head-body length: 126 – 174 mm; Jansa et al., 2019) restricted to the north and north-eastern montane humid forests of Madagascar (Fig. S17). In particular, individuals of *E. ellermani*, whose species identity has been genetically confirmed, have been collected from the northernmost Montagne d'Ambre to the central mountains of the Masoala Peninsula. The habitat of *E. ellermani* consists of evergreen humid forest in montane formations, characterised by moist or sclerophyllous plant communities. It occurs in an elevation range between 440 m and 1570 m (Jansa et al., 2019). *E. ellermani* composes, together with *E. tanala* and *E. tsingimbato*, the so-called *E. tanala* species group. The three species are morphological difficult to distinguish, and also 'remarkably similar morphologically to members of the *E. antsingy* group (*E. antsingy* and *E. carletoni*) (Jansa et al., 2019). The tuft covers about

35% – 40% of tail length. The traits that most distinguish *E*. *ellermani* from the other species of the *E. tanala* group is the bicolored hairs and overall grayish appearance of the ventral pelage in *E*. *tanala*, in contrast to monochromatic ventral hairs and creamy white underparts in *E*. *ellermani.* The tail tuft with a brown tip distinguishes in most cases *E. tsingimbato* from *E*. *ellermani* (Jansa et al., 2019).

E. minor (Major, 1896) is the smallest member of the *Eliurus* genus (head-body length: 101 – 124 mm; Goodman and Carleton, 1996) (Fig. S17). The litter size of *E. minor* ranges from 2 to 4 (Carleton, 2003). The breeding season of this species occurs in the last quarter of the year (October to December; Goodman and Carleton, 1996; Carleton and Goodman, 1998; Goodman et al., 1999). However, this information is limited by the biased time period in which fieldwork data have been collected. Earlier sparse observations (Carleton, 1994) showed a male with scrotal testes recorded for May, a specimen in juvenile pelage for July, an embryo count of two for August and a lactating female for August. Trap capture data suggest that *E. minor* can be found on lianas, small trees, bamboo grass, on tree ferns as well as on dead wood, filigree structures or trees with a small trunk diameter. It has been observed to climb along filigree vegetation, lianas and in dense filigree undergrowth (Marquart, 2014), up to heights of several meters (Goodman and Carleton, 1996; Carleton and Goodman, 1998). Observations by Katrin Marquart (Marquart, 2014) suggests that in comparison to other *Eliurus* species (*E. tanala* and *E. webbi*), *E. minor* can be 'considered both a good large-tree-climbers as well as a good 'small-branch and bamboo- grass-climbers'. *E. minor* occurs in a wide elevation range (near sea level to 1875 m; Garbutt, 1999; Carleton, 2003) in correspondence of moist evergreen forests. Individuals whose species identity has been genetically confirmed show that *E. minor* has a wide distribution, extending from the south-east to the north-east of Madagascar (Everson et al., 2020). Morphologically, *E. minor* can be distinguished from the other *Eliurus* species by its smaller

cranial and body-size, the smallest of the genus. In addition, it is characterised by a typically dark-brown or blackish tuft-tail, a bright brown colouration of the dorsal pelage and the lack of a marked difference in dorsal-ventral coloration. The tuft covers between 60% - 66% of the tail (Carleton, 1994).

E. myoxinus (Milne Edwards, 1885) is a medium-size tuft-tailed rat (head-body length: 117 – 136 mm; Carleton et al., 2001) (Fig. S17). It occurs in the dry forests of the west and in the lowland humid forests of the far north/north-east Madagascar, which have a pronounced annual dry season (Soarimalala and Goodman, 2003). Litter size between 1- 3 (Carleton, 2003). It has been documented from localities in xerophytic scrub (spiny bush), dry deciduous forest, dry-humid transitional forest (Carleton, 1994; Goodman and Ganzhorn, 1994; Goodman and Rasoloarison, 1997; Goodman et al., 1999). It is strictly forest-dependent and occurs in dry and humid forests up to 900 m and 1240 m in elevation, respectively (Soarimalala and Goodman, 2011). *E. myoxinus* individuals have been collected also from the mesic (moist) canyons of Isalo and the upper portion of the Analavelona Massif, which have distinctly more humid vegetation than surrounding lowland areas (Goodman et al., 1999; Carleton et al., 2001; Soarimalala and Goodman, 2003; Goodman et al., 2009; Shi et al., 2013). *E. myoxinus* is characterised by a relatively shorter tail length compared to the other *Eliurus* species (105% - 110% of head-body length). *E. myoxinus*, although being a western-dry associated species, has a body size smaller than any of the eastern-humid *Eliurus* species. It is covered by a generally lighter brown dorsal pelage. *E. myoxinus* individuals present a decline in body size, going from the south to the north of its distribution, e.g., head-body-length is 127.7 mm $\pm$ 6.4 in the south and 123.6 mm $\pm$ 8.7 in the north (Table 3 in Carleton et al., 2001). The tuft is brown to brownish-black, long and bushy, covering between 60% - 70% of the tail (Carleton, 1994).

5.11.3 Sampling and laboratory procedures

We collected small ear biopsies from individuals captured with Sherman traps (H.B. Sherman Traps®) in northern Madagascar (Fig. 1 and Table 1), during the dry seasons of 2011, 2012, 2013, 2014, 2015, 2017 and 2018, according to Rakotondravony and Radespiel, (2009) field procedures. We ensured DNA preservation in field conditions by storing the samples in Queen's lysis buffer (Seutin et al., 1991) and at −20 °C once in the laboratory. We extracted total genomic DNA from 125 ear biopsies using a DNeasy® blood and tissue kit (Qiagen®). To increase the amount of extracted genomic DNA, tissues were incubated overnight at 56 °C in 300 µL of ATL digestion buffer (Qiagen®) including proteinase *K* (C_f = 1 mg/ mL), and 20 µL of dithiothreitol (1 M). The final DNA solution was then eluted in 60 µL of AE buffer. We quantified nucleic acids and checked for DNA purity using the Nanodrop spectrophotometer (ND-1000, Thermo Fisher) and quantified double stranded DNA (dsDNA) using a Qubit 2.0 Fluorometer (Invitrogen). For maximizing the amount of genomic data recoverable from high-throughput DNA sequencing, we selected samples with a dsDNA concentration superior to 4 ng/µL and a 260/280 absorbance ratio higher than 1.75.

We amplified and sequenced the mitochondrial (mtDNA) cytochrome b (*cytb*; 1140 bp) of 13 individuals, distributed in the four different clades identified with preliminary analyses on the nuclear RAD-seq data. The aim of these analyses was to clarify species identity, based on previous mtDNA phylogenetic studies, for individuals belonging to the genetic clades identified by nuclear data, since *Eliurus* species can be difficult to distinguish morphologically, but also due to the limited knowledge on the species diversity of this genus. mtDNA *cytb* was amplified using published primers: MVZ05 and UMMZ04 (Jansa et al., 1999). We carried out PCR amplifications in a 10-µL reaction with final concentrations of 0.8 µl of 1.0 µM for each primer,

5 µl of 1.5 mM GoTaqFlexi DNA polymerase Mix (Promega no. M8305), 1 µl of genomic DNA and 2.4 µl of H_2O. Thermocycling consisted of an initial denaturing temperature of 94 °C for 2 min, followed by 35 cycles of 1 min at 94 °C, 1 min at 51 °C (*cytb* annealing temperature) and 1 min at 72 °C, and a final extension of 10 min at 72 °C. We sequenced the PCR products in an ABI 3130 XL Genetic Analyzer (Applied Biosystems, Foster City, CA) and edited the sequences using GENEIOUS PRO v10.2.3 (http://www.geneious.com, Kearse et al., 2012).

5.11.4 RAD-sequencing

Restriction site-Associated DNA sequencing (RAD-seq) consists in sequencing regions neighbouring restriction sites, to obtain homologous sequences across many individuals, spread across the genome, with high coverage and at a reasonable cost (Baird et al., 2008; Andrews et al., 2016). We prepared RAD-seq libraries using 100-200 ng of genomic DNA from each sample and one negative control made up of water following the protocol described in Etter et al., (2011) and by the Genomic Resources Development Consortium (2015). Briefly, samples were digested with SbfI (New England Biolabs) before P1 adapter ligation. After adapter ligation, we pooled 10 µL of 18 samples in four sub-libraries, before shearing them in a Covaris M220 for 45 sec to obtain fragments with an average size of 500 bp. Following end repair, we conducted size selection using AMPure XP beads (Agencourt). After 3'adenylation, we ligated the P2 adapter. The libraries were finally enriched with 10 cycles of PCR. Libraries quality was verified using Fragment Analyzer® (Advanced Analytical) and quantified using qPCR (QuantStudio6®, Applied Biosystems). We made equimolar pooling with the sub-libraries and sequenced them on an Illumina HiSeq-3000 (150-pb paired-end reads; 72 individuals / lane) at the GetPlage Sequencing platform facility (Toulouse, France). Raw reads were first checked for quality using FastQC (Andrews, 2010) and MultiQC (Ewels et al., 2016), then demultiplexed,

adapter-removed and quality-filtered in STACKS v2.2 (Rochette et al., 2019) using default parameters. PCR duplicates, that is multiple copies of the same single-end and paired-end read pair DNA fragment, were removed using the *clone_filter* program in STACKS v2.2 under default settings. The last filtering step was carried out using the *kmer_filter* program in STACKS v2.2 with default settings. This step removes reads with putative sequencing errors (identifying kmers with a coverage lower than 15% of the median read coverage), decrease the number of over-abundant reads and normalize read-depth according to k-mer coverage.

5.11.5 RAD-seq data processing and *de novo* assembly

After filtering the raw short-read sequences, it is necessary to assemble the reads pairs into sets of reads (stacks) and merge overlapping stacks into putative loci, which will then be used for population genomic analyses. Since there is no reference-genome available for the *Eliurus* genus, we decided to proceed with a *de-novo* assembly of RAD-seq nuclear loci. *De-novo* assembly was performed in STACKS v2.2 using *ustacks*, *cstacks*, *sstacks*, *tsv2bam*, *gstacks* and *populations* programs, which we describe below with more details.

The *ustacks* program is run on each individual in the dataset, separately, to build loci from single-end short-reads. It can be distinguished in two steps which allows to i) create *stacks* (or putative alleles) of exactly matching reads; and ii) match putative alleles (*stacks*) into a locus. These two steps are controlled by two parameters: *m* (minimum depth of coverage), corresponding to the minimum number of raw reads required to form a *stack*; and *M* (maximum distance between stacks), equivalent to the maximum number of nucleotides differences allowed between stacks to merge them into a putative locus.

The *de-novo* analysis creates a 'synthetic' reference-genome (*catalog*) which is made of all loci and alleles assembled in the *ustacks* program across the individuals of the dataset. In other words, it creates a set of consensus loci by merging homologous alleles of different individuals together. The key parameter in this step is *n* (distance allowed between *catalog* loci), corresponding to the number of mismatches allowed between sample loci when building the *catalog* consensus loci. This parameter is fundamental since it determines the threshold for considering a given allele a homologous of other alleles in the whole dataset.

The program *sstacks* is used to match individuals' loci to the *catalog*, and loci matching more than one *catalog* locus are excluded because their true matching locus in the *catalog* is ambiguous, either due sequencing error or to true paralogy. Multiple loci, however, can still uniquely match the same *catalog* locus, however later analyses (in *gstacks* and *populations* program) can exclude these loci (Catchen et al., 2011). The *tsv2bam* program rearrange the format of the matched loci and retrieve the paired-end reads associated with the single-end reads that have been used for individuals' loci assembly. The *gstacks* program assembles the paired-end reads retrieved in the previous step into contigs which are then merged with the assembled single-end loci. After that, *gstacks* calls SNPs within the whole dataset using the information on the matched loci, and genotype each individual accordingly. For *gstacks*, we used a *--var-alpha* and *--gt-alpha* equal to 0.01. The *var-alpha* parameter defines the p-value threshold for calling a given nucleotide *polymorphic* rather than *invariant* across all individuals in the dataset. The *gt-alpha* parameter defines the p-value threshold of the likelihood-ratio test statistic between the significant heterozygous genotype (called depending on *var-alpha*) and the genotype homozygous for the more abundant nucleotide in the individual at a given site (Maruki and Lynch, 2017). If the likelihood of the heterozygous genotype is significantly greater than the likelihood of the homozygous genotype, then the heterozygous genotype is

considered, otherwise a homozygous genotype is called. In practice, a *gt-alpha* parameter of 0.01 would be equivalent to a Genotype Quality score ≥ 20, calculated as $-\log_{10}(\text{likelihood ratio test p} - \text{value})$.

The *gstacks* program generates the consensus *catalog* FASTA file which correspond to the 'synthetic' reference genome used for phylogenomic analyses. The *catalog* FASTA file was cleaned from possible contaminant sequences using *DeconSeq* software (Schmieder and Edwards, 2011). We decided to use *DeconSeq* on the *catalog* loci and not on the raw reads, because it requires long-read datasets (> 150 bp mean read length) and also because several steps in the STACKS v2.2 pipeline are made such to reduce the probability of calling nucleotides that are rarely represented in the dataset. Briefly, *DeconSeq* assesses sequence similarity between *catalog* loci and Bacteria, Virus and Human databases of potential contaminants. It uses the parameters *coverage* and *identity*, that is how many nucleotides of a *catalog* locus are covered by and identical to the reference contaminant sequence, respectively. We set *coverage* equal to 95 and *identity* equal to 94.

Finally, we used the *populations* program to analyse the genomic data from the whole dataset allowing to convert this information, for instance, into a dataset of single SNPs per locus or to FASTA sequences. In doing so, additional filtering parameters need to be specified, which allow to further increase confidence on the recovered genomic information by reducing biases due to sequencing errors, paralogy or repetitive regions. In the *populations* program, as conditions for processing a nucleotide at a locus, we set i) a minimum minor allele frequency (*min_maf*) equal to 5%; ii) a maximum observed heterozygosity (*max_obs_het*) equal to 70%; iii) restrict data analysis to one random SNP per locus, in order to reduce linkage disequilibrium among genotyped SNPs (*write-random-snp*); and iv) a minimum percentage of individuals in a population required to process a locus equal to 80% (*-r*).

As noted above, the assembly of nuclear loci from RAD-seq data relies on several parameters (*m*, *M*, *n*). The choice of the best combination of parameters is difficult, since several factors can influence the amount and quality of genomic information recoverable from the dataset, such as the biology and natural history of the studied population(s) or the deployed RAD-seq technique (Paris et al., 2017). Therefore, we followed the protocol proposed by (Paris et al., 2017) to optimize the three main parameters of the *de novo* assembly (*m*, *M*, *n*). The analysis for the optimization of *de novo* assembly parameters was performed on a subset of the whole dataset, following Rochette and Catchen, (2017). In particular, Rochette and Catchen, (2017) found that adding all samples does not increase significantly the number of widely shared *catalog* loci in the dataset and can add noise in the estimation of *m*, *M* and *n*. The individuals used for building the *catalog* were selected by considering i) those with highest coverage and ii) maximizing the geographic distribution of the selected individuals, as a proxy for maximizing the genetic diversity in the dataset (Rochette and Catchen, 2017). We thus performed the optimization analysis on twenty individuals from different regions. This included four *E. carletoni*, four *E. minor*, five *E. ellermani,* five *E. myoxinus* and one *Eliurus* sp. nova (see Fig. S2 and Tab. S1). These were also the individuals used for building the *catalog*. The optimization procedure consisted in iterating a range of values (2 to 8) for the *m*, *M* and *n* parameters, separately, while keeping all other parameters fixed at default values (*m* = 3, *M* = 2 and *n* = 1). For each parameter combination we extracted the genomic information shared by either 40%, 60% or 80% of the individuals in the dataset. In order to identify the best parameter combination, we extracted and assessed several assembly metrics: i) individual coverage i) the number of assembled loci; ii) the number of polymorphic loci; iii) the number of SNPs; iv) the variable and v) fixed sites identified across the whole population. According to Paris et al., (2017) the best parameter combination is identified considering: i) a stable set of loci that are highly replicated across 80% of the population, since it is unlikely to be derived from paralogous, repetitive

sequences or sequencing errors; and ii) the highest amount of polymorphism, or the parameters 'value at which the number of polymorphisms levels out. In other words, Paris et al., (2017) criteria provide a balance between obtaining true polymorphism and introducing sequencing error, or over-merging paralogous or repetitive loci. Once the informed choice of *m*, *M* and *n* parameters was done, we re-ran the STACKS v2.2 pipeline using the selected value and obtain the definitive *catalog* on which further genomic analyses are carried out.

5.11.6 Mitochondrial DNA analyses

We compared the mtDNA *cytb* of 13 sampled *Eliurus* individuals with published data on the other 12 *Eliurus* species. These data were retrieved from NCBI (Tab S2 for accession numbers): *E. antsingy* (N= 6), *E. carletoni* (N=67), *E. ellermani* (N = 38), *E. grandidieri* (N = 3), *E. majori* (N = 16), *E. minor* (N = 11), *E. myoxinus* (N = 60), *E. tanala* (N = 14), *E. tsingimbato* (N = 11), *E. webbi* (N = 16). As outgroup we used mtDNA *cytb* of *Rattus norvegicus* (EU349782.1). The complete dataset consisted of 263 mtDNA *cytb* sequences. We aligned and visually checked the sequences using the Clustal Omega method (Sievers et al., 2011). We performed format conversion of DNA alignments using ALTER (Glez-Peña et al., 2010). We used two statistical methods for phylogenetic analysis: a Bayesian MCMC approach implemented in MRBAYES v3.2.1 (Ronquist et al., 2012) and a maximum-likelihood (ML) approach implemented in RAXML v8.2.X (Stamatakis, 2014).

In MRBAYES v3.2.1, we used four MCMC chains, each at the default temperature and running for 15,000,000 generations. Trees were sampled every 15,000 generations and 25% of total sampled trees were discarded as the "burn-in." To assess MCMC convergence, we checked for stationarity of the log probability of the data over time, we verified that the potential scale reduction factor was close to 1 and we used Tracer v1.6 (Rambaut et al.,

2014) to confirm that all estimated sample size values were >200. In RAXML v8.2.X, we used the rapid Bootstrap option with automatic detection of the sufficient number of bootstrap replicates (option: “autoMR”). Retained posterior distributions of trees are summarized in MRBAYES to build a consensus tree, while in RAXML majority-rule consensus tree is estimated from the Bootstrap replicates. We estimated the best partitioning scheme and best-fit model of evolution for mtDNA *cytb* in PARTITIONFINDER v2.1.1 (Lanfear et al., 2017) using the greedy algorithm and selecting the best model according to AICc criteria. We performed a separate analysis for each phylogenetic method, given that MRBAYES and RAXML accept only some of the models of evolution that can be estimated in PARTITIONFINDER.

We measured pairwise mtDNA distance within and between species using the *dist.dna* function in the *adegenet* R package (Jombart, 2008; Jombart and Ahmed, 2011). We used the ‘raw’ evolutionary model to compute genetic distance, that is the proportion of sites that differ between each pair of mtDNA sequences.

5.11.7 RAD-seq datasets building

Genomic analyses were performed on several datasets, which varied in number and geographical distribution of the individuals, and list of loci to include in the analyses (*white list*). We decided to replicate several of the analyses on multiple datasets in order to gain confidence on the inferred pattern. Details on each of the dataset are presented in Table S3. Here we give a detailed description of the criteria of selection. In particular, we generated four types of datasets, differing on number and geographical distribution of the individuals: i) all 125 individuals (*N125*); ii) one individual per sampling site, per species (*N62*); iii) 10 individuals per species (*N48*); and iv) 3 individuals per species (*N15*). For each dataset, we created a *white list*, which included loci that were genotyped: *i*) if at least 80% of the individuals across taxa contained that locus (*pop1.r80*); or *ii*) if at least 80% of the

individuals within taxa contained that locus (*pop5.p0.r80*); or *iii*) if at least 80% of the individuals within taxa and 80% of the taxa (N = 4) contained that locus (*pop5.p4.r80*).

The *white lists* were then used for generating the genomic datasets on which to perform phylogenomic and genomic analyses. Most of the datasets used for phylogenomic analysis (maximum-likelihood tree inference) were composed of sites that are fixed within individuals but variables across individuals, since we are first interested in the topology of the species tree, and not to the branch lengths. For two datasets (*N125.pop5.p4.min100* and *N15.pop5.p4.min15*), we also generated a multialignment composed of all sites (variant + fixed). The *N125.pop5.p4.min100* dataset was used for computing average genetic difference between and within taxa, whereas the *N15.pop5.p4.min15* dataset was used for phylogenetic inference, species delimitation test and estimation of divergence times and species population size (see below). Finally, we generated a dataset, with all individuals across taxa, composed by one random SNPs per catalog locus, which needs to be present in at least 80% of the individuals within taxa and in at least four out of the five taxa (*N125.pop5.p4.r80*).

For comparison, we also generated one genomic dataset (*N125.pop5.p4.r80*) using a genotype-likelihood (GL) based approach implemented in ANGSD (Nielsen et al., 2012; Korneliussen et al., 2014). One of the differences that distinguish ANGSD from STACKS v2.2 is that the former does not call genotypes but rather output the posterior probability of all possible genotypes, given the data. Using the GLs based approach would allow thus to strength results by using a larger amount of information. The ANGSD pipeline was run on the full dataset (N=125) by first aligning cleaned sequences to the *de-novo* catalog generated in STACKS v2.2. The alignment was carried out with *bwa-mem* (Li, 2013). We estimated locus depth as the F1 read starting position (Sbf1 cutting site) using the “samtools-view” command (Li et al., 2009). We then estimated genotype likelihoods with the SAMtools approach (‘-GL 1’)

using only loci with *i*) a total depth of at least five (-setMinDepth), *ii*) a total depth of at most the sum of the 0.95 quantile of the depth distribution of considered individuals (-setMaxDepth), and *iii*) an individual depth of at most the 0.95 quantile of its depth distribution (-setMaxDepthInd). This strategy allows discarding loci present only in a very low number of individuals, or over-covered loci that are likely to be repeated or paralog regions (Heller et al., 2021). In addition, we considered only bases with a minimum quality of 20, reads that mapped to one unique *catalog* locus with a minimum quality of 30 and associated with their pair-end read. Finally, we only kept biallelic variants with a SNP probability lower than 1e-6, a minor allele frequency (MAF) of at least 0.05, and that were present in at least 100 out of 125 individuals.

5.11.8 Phylogenomic analysis

We employed a maximum-likelihood approach (RAXML v8.2.11; Stamatakis, 2014) to infer evolutionary relationships among the five *Eliurus* taxa identified in northern Madagascar.

We performed concatenated analyses on datasets composed by fixed sites. The concatenated analysis assumes that all the sites included in the multialignment share the same species tree topology (Felsenstein, 1981). We set the unpartitioned GTRCAT molecular model with joint branch length optimization, the '-f a' option to conduct 1000 rapid Bootstrap analysis and the 'autoMRE_IGN' option to carry out bootstrap convergence test, which determines whether sufficient bootstrap replicates have been run in order to get stable support values in the inferred phylogenetic tree. We removed invariant sites from the datasets and applied an ascertainment bias correction for binary SNP data in the RAXML analysis based on Paul O. Lewis formula (Lewis, 2001).

We also carried out a partitioned analysis on the *N15.pop5.p4.min15* dataset, including all sites (variant + fixed), setting each catalog locus as a

separate partition. The partitioned analysis estimates the model parameters of the most likely evolutionary model for each partition, while fixing the global set of branch lengths (genealogy). As for the concatenated analysis, we set GTRCAT molecular model for each partition. In this analysis we also used the argument '-M', which allows in addition to estimate a separate set of branch lengths for every partition.

For each maximum-likelihood analysis, the best likelihood tree with bootstrap supported value was visualized using *ggtree* and *treeio* R packages (Yu et al., 2017; Yu, 2020; Wang et al., 2020).

5.11.9 Clustering analyses (ADMIXTURE, PCA)

Clustering methods were used in order to identify genetic groups across and within *Eliurus* species. We used the widely-used maximum likelihood model-based approach implemented in ADMIXTURE v1.3 (Alexander et al., 2009). This program uses the same likelihood model as STRUCTURE software (Pritchard et al., 2000), but instead of sampling the posterior probability of a model given the data, it maximises the likelihood of the model. In the model, it is assumed that individuals are sampled from an admixed population with contribution from K *a priori* clusters. Then, individuals are assigned to K-clusters by maximizing the likelihood of the model. For each dataset, we ran the program for K values ranging from 1 to 8, and repeated the analysis for 30 independent runs. ADMIXTURE analysis was carried out using the default optimization method and the default termination criterion, which stops the analysis when the log-likelihood increases by less than 10^{-4} between iterations. K values were compared using the cross-validation method by setting $--cv = 20$ and by comparing likelihoods across runs. We used CLUMPAK (Kopelman et al., 2015) to permute the 30 independent runs so that clusters align across runs. We selected the best K number of clusters considering the K value with lower

average cross-validation error compared to other K values, but we also explored admixture models for K - 1 and K + 1. The final assignment of individuals to the best K-value was obtained by taking the mode containing most runs (*Major.Cluster*).

In addition, we used a model-free approach, the Principal Components Analysis (PCA), throught the function *glPca* implemented in the R package ADEGENET 2.0.1 (Jombart and Ahmed, 2011). PCA constructs low-dimensional projections of the data that explain the gross variation in marker genotypes, and thus it can be useful to assess whether most of the variation in the genomic dataset can be explained by species divergence. ADMIXTURE and PCA analyses were performed on the *N125.pop5.p4.r80* dataset, which includes all 125 individuals and loci that are present in at least 80% of the individuals within taxa and in at least four out of five taxa.

Admixture and PCA analyses were also performed on the GLs-based ANGSD dataset using the NGSadmix and PCAngsd programs (Skotte et al., 2013). NgsAdmix was run for 20 iterations per K value (1 to 12) and a stringent tolerance of 10^{-6} for convergence. We compared K values using the ΔK procedure (Evanno et al., 2005).

5.11.10 Species Validation

We used three approaches to test the species hypothesis for the newly identified taxa (*Eliurus sp. nova*): i) guided-tree species delimitation analyses; ii) estimation of the gdi-index; and iii) compared isolation-by-distance pattern within versus between taxa.

Species delimitation analysis with guided tree was performed with BPP (Yang and Rannala, 2010; Flouri et al., 2018, 2020) using full-length FASTA file for a subset of individuals, namely three individuals per species (*N15.pop5.p4.min15* dataset). We used six combinations of priors for divergence time (τ) and population size (θ). The θ and τ priors consisted of an inverse gamma distribution with α = 2 (θ) or α = 3 (τ) and β equal to either

0.002 or 0.02 or 0.2. Six combinations of priors are tested in order to ensure that model selection would not be influenced by prior information. Four MCMC runs were carried out for each combination of parameters, and convergence was checked by comparing the outputs from the four runs. Each run was conducted with a burnin of 5000, sample frequency of 100, and collected 200,000 samples. Each analysis was repeated using both rjMCMC algorithms in BPP, algorithm0 and algorithm1. Prior combinations, two distinct MCMC algorithms and four runs per analysis were conducted in order to ensure convergence of the results. The tree model with higher posterior probabilities across all runs was selected as the best tree topology.

The genealogical divergence index (*gdi*; Jackson et al., 2017; Leaché et al., 2019) was calculated using the posterior distribution of θ and τ from the analysis A00 carried out in BPP (Yang, 2015). Analysis A00 estimates divergence times and population sizes using a guided specie tree, under the MSC model. As for the species delimitation test, we ran the analysis for each of the six combinations of priors for θ and τ, and four runs for each combination. The details of the MCMC run were the same as for the species delimitation test. The posterior distribution of the gdi-index was calculated by computing the gdi-index for each of the 200,000 MCMC samples, across runs, using *equation 7* from Leaché et al., (2019): $gdi = 1 - e^{-2\tau/\theta}$. The gdi-index analysis was carried out to test the species hypothesis for *Eliurus sp. nova*, therefore, we simply used the estimated divergence time (τ) between *Eliurus sp. nova* and *E. myoxinus* (its sister-species), and alternatively used as population size (θ) either the one estimated for *Eliurus sp. nova* or for *E. myoxinus*.

Complementary to the previous species delimitation methods, which ignore the importance of geographically-limited dispersal in determining genetic differences between populations, we compared nuclear genetic and geographic (Euclidean) distances within taxon (*E. myoxinus* or *Eliurus* sp.

nova) and between individuals belonging to different taxa (*E. myoxinus* and *Eliurus* sp. nova). We performed this analysis i) on the STACKS generated datasets by estimating pairwise genetic distances using the *stamppNeisD* function in R (R Core Team, 2021) which computes the Nei's genetic distance (Nei, 1972); and on the ANGSD generated datasets by estimating the covariance of nuclear RADseq GLs in PCAngsd (Meisner and Albrechtsen, 2018).

5.12 Results

5.12.1 Species identification from mtDNA *cytb* reference phylogeny

We sequenced the mtDNA *cytb* for 13 individuals (Tab. S4 or Fig. S3) which were distributed in distinct genetic clusters in the neighbor-joining tree estimated using the RAD-seq dataset (Fig. S3). They were analyzed with published *Eliurus sp.* mtDNA *cytb* sequences. Using both maximum-likelihood and Bayesian based phylogenetic analyses, we identified within the 13 individuals four previously described *Eliurus* species: *E. carletoni*, *E. ellermani*, *E. minor* and *E. myoxinus* (Fig. 2a; Fig. S3-S4). We observed major topological differences between ML and Bayesian trees at the deeper nodes. The evolutionary relationships reconstructed using Bayesian-based phylogenetic method (MRBAYES) were consistent with previous *Eliurus* phylogenies (e.g., Jansa et al., 2019).

The phylogenetic analyses have also revealed a novel mtDNA *Eliurus* lineage, sister-clade of *E. myoxinus*, which hereafter we will refer to as *Eliurus* sp. nova (Fig. 2a; Fig. S3-S4). The level of average genetic differences ($p-distance$) across all *Eliurus* species pairs ranged between 6.12% (*E. antsingy* – *E. carletoni*) and 17.6% (*E. grandidieri* – *Eliurus* sp. nova), with a median of 13.4% (Fig. 2b, c). Among all *Eliurus* sister species pairwise comparisons, the $p-distance$ value of *E. antsingy* – *E. carletoni* was the lowest (6.12%), whereas the value of *E. antsingy* – *E. majori* was the highest (12.9%) (Fig. 2d). The two sister taxa *Eliurus* sp. nova and *E. myoxinus* showed a high level of average genetic differences among sister-species ($p-distance$ = 11.3%), similar to that between *E. minor* - *E. myoxinus* ($p-distance$ = 11.6%).

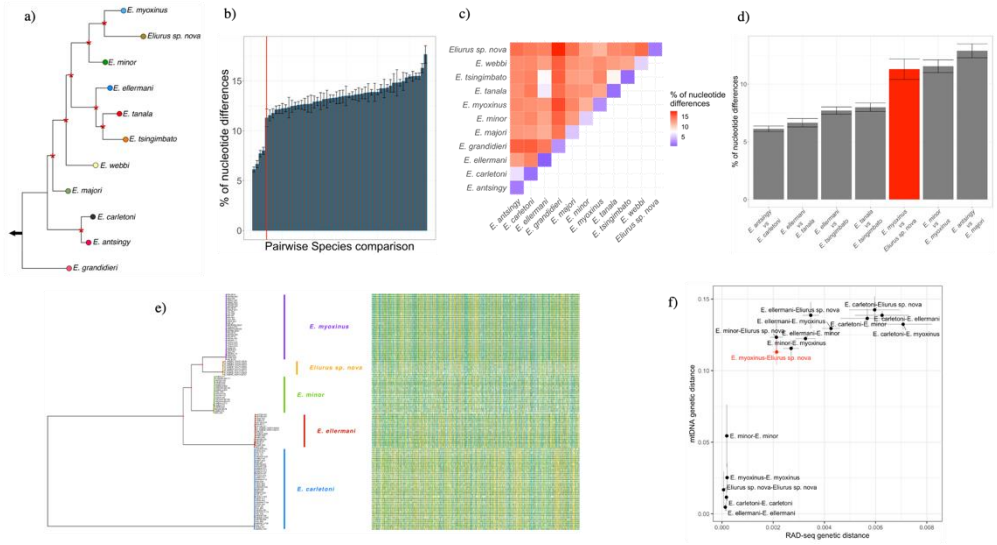


Figure 2. Inference of the phylogenetic relationships within the *Eliurus* genus. a) Bayesian-based phylogeny obtained from MRBAYES. The outgroup EU349782.1 corresponds to mtDNA *cytb* of *Rattus norvegicus*. b) Mean proportion of mtDNA *cytb* nucleotide differences (mtDNA *cytb* p-distance) between *Eliurus* species pairs. Error bars represent the standard deviation around the mean. The vertical redline marks the *Eliurus sp.* nova – *E. myoxinus* comparison. c) Matrix of the mtDNA *cytb* p-distances between *Eliurus* species pairs. d) mtDNA *cytb* p-distances between *Eliurus* sister-species pairs. The red bar represents the *Eliurus sp.* nova – *E. myoxinus* mtDNA *cytb* p-distance, which is larger than the other four comparisons of closely-related *Eliurus* species pairs. e) Maximum-likelihood based phylogeny of *Eliurus* samples based on nuclear RAD-seq data (*N125.pop5.p4.min100*; 4.351 SNPs). On the right side is showed the multialignment from which the phylogeny has been inferred. f) Correlation of pairwise *Eliurus* sp. genetic differences between mtDNA *cytb* and nuclear RAD-seq data. The results suggest a strong correlation between mtDNA *cytb* and nuclear RAD-seq phylogenies (Spearman rho: 0.896; p-value<2.2e-16), and that *Eliurus sp.* nova could be considered a newly discovered *Eliurus* species, holding among the highest level of *Eliurus* sister-species genetic divergence.

5.12.2 RAD-data processing: assembly of nuclear loci *de novo*

After the three filtering steps described in the *Material and Methods* section (*process_radtags*, *clon_filter*, *kmer_filter*) and data processing (*ustacks*, *cstacks*, *sstacks*, *tsv2bam*, *gstacks*) of each RAD-seq dataset, we obtained in total about 500 million short-reads (half paired-end and half single-end), with an average of ~4 M reads per individual. The minimum number of forward reads recovered per individual was 201 K and the maximum was 5.2 M reads (Tab. S3). The optimization analysis of *de novo* assembly parameters shows that the highest amount of polymorphism for $r = 80\%$ was at $m = 2$, $M = 5$, $n = 8$ (Fig. S5). The cleaned *catalog* of *Eliurus spp.* was composed of 327,870 loci with average length of ~780 bp. Across the whole RAD-seq dataset, the weighted mean individual coverage was between ~5 and ~40X (Tab. S3; Fig. S6).

For phylogenomic analyses, we used STACKS to construct eleven RAD-seq datasets (Tab. S3), which differ in sample size, sampling scheme and the selecting criteria for genotyping a given locus (see Material and Methods). In addition, we generated a dataset using the genotype-likelihood based program ANGSD, which unlike STACKS, takes into account genotype uncertainty rather than definite genotype calls. Across the eleven STACKS datasets, we recovered between 8,471 and 108,662 variant sites, whereas the ANGSD dataset included 392,405 sites.

5.12.3 Phylogenomic-data analysis

The maximum-likelihood based phylogeny inferred using the full RAD-seq dataset (*N125.pop5.p4.min100*) is consistent with the mtDNA *cytb* phylogeny (Spearman rho: 0.896, p-val < 2.2e-16; Fig. 2a, e, f), although not all *Eliurus* species are present in our RAD-seq dataset. The maximum-likelihood based phylogenies were identical across all RAD-seq datasets, regardless of the sampling scheme, sample size or list of included *catalog* loci

(Fig. S7). For the dataset including the lowest sample size (*N15.pop5.p4.min15*), we also performed partitioned analysis on RAXML, in which, for each partition (equivalent to gene), the method estimates the model parameters of the most likely evolutionary model, while fixing the global set of branch lengths (genealogy) (Fig. S8a). In addition, we also inferred the maximum likelihood tree for each partition independently (Fig. S8b). A qualitative comparison of Fig. S7 and Fig. S8 show consistencies between concatenated, partitioned and gene phylogenies.

5.12.4 Genomic structure

Analysis of hidden genetic structure with both ADMIXTURE and NGSadmix identify the best number of genetic clusters K = 6 (Fig. S9). The coefficients of individual ancestry estimated from the two approaches show that at K = 4, *E. myoxinus*, *E. carletoni*, *E. ellermani* and *E. minor* are primarily separated, whereas *Eliurus sp. nova* show mixed ancestry between *E. minor* and *E. myoxinus* (Fig. 3). At K = 5, ADMIXTURE results identify population structure within *E. carletoni* as an additional genetic cluster, and at K= 6, *Eliurus sp. nova* individuals can be genetically separated from the two closely-related species *E. minor* and *E. myoxinus*. NGSadmix results show clear genetic differences between *Eliurus sp. nova* and the other *Eliurus* species at K = 5, while reveling population structure within *E. carletoni* at K= 6 (Fig. S10, right panel).

The PCA analysis shows five distinct clusters of individuals (Fig. 3), corresponding to the genetic clades identified in the phylogenetic analyses. The first principal component (PC1) explains ~50% of the variance, contributing mostly to the delimitation of *E. carletoni* against the other *Eliurus* species. The PC2 explains ~15% of the variance, clearly separating the remaining four *Eliurus* taxa. A similar pattern of separation can be seen with PC3 (~5%) while at PC4 (~1%) major differences are observed between

Eliurus sp. nova and the other four *Eliurus* species. The PCA analysis performed on the ANGSD dataset further confirms these results (Fig. S10).

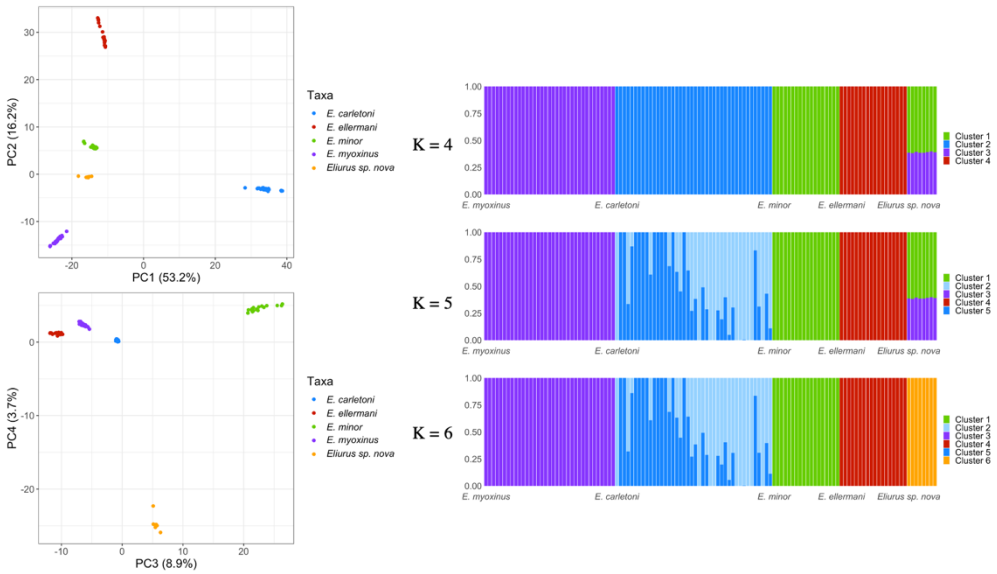


Figure 3. Genetic structure of sampled *Eliurus* individuals in northern Madagascar. PCA and ADMIXTURE analyses separate individuals according to the hypothesized species identification (*E. carletoni*; *E. myoxinus*; *E. ellermani*; *E. minor*) while identifying the hypothetically new *Eliurus* species (*Eliurus sp. nova*) as being genetically distinct from the others, and, in particular, from the sister-species *E. myoxinus*. Both analyses have been performed using the *N125.pop5.p4.min100* dataset which includes 346,893 sites, out of which 108,662 are variants.

5.12.5 *Eliurus* species delimitation

The phylogenomic and genomic structure analyses above appear thus to support the hypothesis of a new distinct *Eliurus* taxon, to which we have referred as *Eliurus sp. nova*. In this section, we present further analyses aimed at quantitatively testing the species delimitation hypothesis, taking into account the effect of population structure as a possible confounding factor.

We carried out species delimitation analyses with BPP, using all sites (variant and invariant) of the *N15.pop5.p4.min15* dataset, for testing the

species delimitation hypothesis of the three closely related species: *E. myoxinus*, *E. minor* and *Eliurus sp. nova*. The distributions of divergence times parameters were concordant across the four MCMC runs, the six combinations of priors for divergence times and population size parameters, and for both MCMC algorithms (algorithm0 and algorithm1) (Fig. S11). Similarly, we did not observe major differences in the distribution of log-likelihood values across runs, suggesting MCMC convergence of the analysis. The results of the BPP species delimitation analysis strongly support the delimitation of *E. myoxinus*, *E. minor* and *Eliurus sp. nova* as distinct species, since across all runs, algorithms and priors, this topology was always the best species tree with a posterior probability of 1.0 (Fig. 4a).

The multialignmnet including sequences of *E. myoxinus*, *E. minor* and *Eliurus sp. nova* (*N15.pop5.p4.min15* dataset) was also used in a separate analysis to estimate divergence time and population size parameters. The analysis was performed in BPP under the multispecies-coalescent (MSC) model, and the estimated values were used to calculate the *gdi*-index. The distribution of MCMC samples for most parameters were consistent across runs, prior's combinations and algorithms, suggesting convergence of the MCMC analysis (Fig. S12). The *gdi*-index was calculated as $gdi = 1 - e^{-2\tau/\theta}$ (Jackson et al., 2017; Leaché et al., 2019), using as τ the divergence time between the two sister-taxa *E. myoxinus* and *Eliurus sp. nova* ($\tau_{E.myoxinus-Eliurus\ sp.nova}$) and as θ, either *E. myoxinus* ($\theta_{E.myoxinus}$) or *Eliurus sp. nova* ($\theta_{Eliurus\ sp.nova}$) population size. We computed *gdi*-index for each run and combination of priors, separately, and compared the results. Fig. 4b shows little variations in *gdi*-index across runs, with values ranging between 0.99994 and 1.

We performed an *ad hoc* test that compares isolation-by-distance (IBD) patterns for individuals belonging to the same taxa (*E. myoxinus* - *E. myoxinus* or *Eliurus* sp. nova- *Eliurus* sp. nova) with IBD patterns for

individuals assigned to different taxa (*E. myoxinus* - *Eliurus* sp. nova) (Fig. 4c). We used this approach to assess whether *Eliurus* sp. nova would correspond to a new *Eliurus* species or to intraspecific spatial genetic structure within *E. myoxinus*. The results show positive and significant IBD for *E. myoxinus* (Mantel'R= 0.484; p-value=0.001) and *Eliurus* sp. nova (Mantel'R= 0.7386; p-value=0.001), whereas negative significant IBD was detected for *Eliurus* sp. nova-*E. myoxinus* (Spearman *rho*= -0.187; p-value= 0.0017). The results show also that, for similar geographical distances, *Eliurus* sp. nova- *E. myoxinus* genetic differences are significantly higher than within taxa genetic differences (Fig. 4d). Similar patterns can be observed for two well-recognized *Eliurus* species, *E. ellermani* and *E. minor*, for which the average proportion of genetic differences is slightly higher than *Eliurus* sp. nova- *E. myoxinus* species pairs ($p - distance_{cytb}$ = 12.3%; Fig. S14). Patterns of covariance estimates within and between taxa obtained in ANGSD were consistent with patterns of pairwise genetic differences computed from SNPs called in STACKS (Fig. S13).

We also measured the percentage of GC content within each species and compared these values across the *Eliurus* species included in our dataset. We used a multialignment which included loci present in at least 100 out of the 125 individuals in the dataset. All sites (variant and invariant) were extracted and used to compute %GC (Fig. 4e). The average %GC content ranged between 46.06% (*E. carletoni*) and 46.18% (*E. myoxinus*). We did not observe significant differences among *E. ellermani*, *E. minor* and *Eliurus sp. nova*, however we found significant differences between *Eliurus sp. nova* and its sister-taxa *E. myoxinus*.

5.12.6 *Eliurus* species distribution in northern Madagascar

We estimated range areas for the five *Eliurus* species by computing the minimum convex hull among all sampling sites for which species identity

has been genetically determined in the present and previous studies (Fig. S16). Among the four previously known *Eliurus* species, *E. myoxinus* and *E. minor* present the widest distribution in Madagascar, ~378,000 km^2 and ~261,423 km^2, respectively, followed by *E. ellermani* and *E. carletoni* with, ~45,500 km^2 and ~8,700 km^2. *Eliurus* sp. nova was found in Montagne d'Ambre (~ 249 km^2), and never identified in the other forests we visited. The altitudinal range varied across *Eliurus* species (Fig. S15): between 37 m and 1,002 m for *E. carletoni*; between 40 m and 1,284 m for *E. myoxinus*; between 289 m and 1,558 m for *E. ellermani*; between 289 m and 2,016 m for *E. minor*; between 812 m and 1,057 m for *Eliurus* sp. nova. We found significant differences in elevational range among all *Eliurus* sp. pairwise comparisons, except for *E. minor* - *Eliurus sp. nova* (p-value=0.14) and *E. ellermani* - *Eliurus* sp. nova (p-value=0.77).

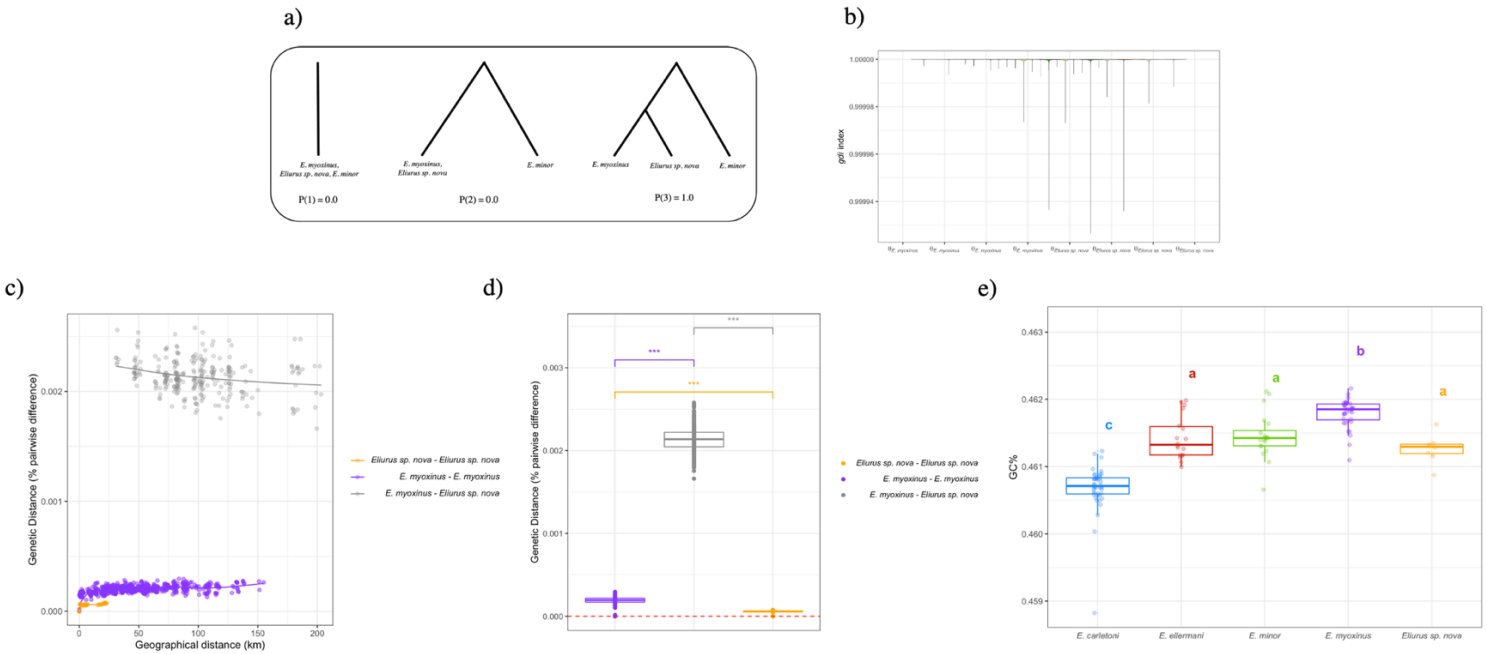


Figure 4. Species delimitation analyses for *E. myoxinus*, *Eliurus* sp. nova and *E. minor*. a) Alternative species delimitation models generated from the hypothetical fixed guide tree ((*E. myoxinus*, *Eliurus sp. nova*), *E. minor*) in BPP. P(1), P(2), P(3) refers to the posterior probabilities for the best species tree model across four runs for each of the two tested rjMCMC algorithms (finietune0 and finetune1) and six combinations of population size and divergence time priors; b) *gdi*-index calculated using the estimated of parameters ($\theta_{E.\ myoxinus}$; $\theta_{Eliurus\ sp.nova}$, $\tau_{E.myoxinus-Eliurus\ sp.nova}$) under the multispecies-coalescent model implemented in BPP; c) Isolation by distance within taxa (*E. myoxinus*, *Eliurus sp. nova*) and between taxa; d) Comparison of genetic distance within and between taxa; e) percentage of GC content (%GC) across *Eliurus* species, estimated from all sites (variant and invariant) extracted from the *N125.pop5.p4.min100* dataset (6,162,356 bp).

5.13 Discussion

In the present study, we analysed mitochondrial and nuclear genomic data of 125 Malagasy endemic rodents (*Eliurus* genus) collected in northern Madagascar. We assessed the extent to which species delimitation based on mitochondrial data is supported by nuclear genomic data. Our analyses have identified an undescribed *Eliurus* clade (*Eliurus sp. nova*), recovered from individuals collected in the humid forests of Montagne d'Ambre, in the far north of Madagascar. We hypothesized that the newly identified *Eliurus* genetic clade would correspond to an evolutionary unit distinct from the other known *Eliurus* species. We tested our hypothesis by performing a series of phylogenetic inferences using ML and MSC based approaches, and by evaluating pattern of genomic diversity. We assessed the support for the *Eliurus sp. nova* species hypothesis by using three approaches: i) the species delimitation tests based on MSC model, ii) *gdi*-index and iii) the *ad-hoc* test of isolation-by-distance within *versus* between sister-taxa. Our results support the delimitation of the newly discovered *Eliurus* clade as a distinct species. In addition, we have also shown that, in northern Madagascar, the dry-habitat associated *E. myoxinus* can actually be mostly found in humid habitats. We discuss below the implications of such findings in the context of *Eliurus* species evolution and diversification. Lastly, we confirm that *Eliurus* species co-occurrence is common in northern Madagascar.

5.13.1 *Eliurus* mitochondrial genetic diversity

Among the 125 individuals sampled in dry, humid and transition forests across a study region of ~12,000 km^2, in northern Madagascar, we identified four previously known *Eliurus* species (*E. carletoni*, *E. myoxinus*, *E. minor*, *E. ellermani*) and one new *Eliurus* taxon (*Eliurus* sp. nova). The *Eliurus* evolutionary relationships inferred from mtDNA *cytb* were in agreement with previous phylogenies (Jansa et al., 1999, 2019), and placed *Eliurus* sp. nova as sister-taxa of *E. myoxinus* (Fig. 2a). The percentage of genetic differences

at the mtDNA *cytb* between *Eliurus* species pairs suggest strong level of genetic differentiation within the *Eliurus* genus (6.12% - 17.6%; Fig. 2b). Previous analyses at the mtDNA *cytb* of six out of thirteen *Eliurus* species had shown similarly high values of pairwise species genetic differences (4.7% - 13.2%; Jansa et al., 2019). For a comparison, for instance, the average species genetic differences measured at the same mitochondrial locus in a small primate genus endemic of Madagascar (*Microcebus* spp.) ranged between 0.9% and 15.3% (e.g., Sgarlata et al., 2019). The *Eliurus* sp. nova - *E. myoxinus* sister-taxa showed the highest level of genetic differences compared to the other *Eliurus* sister-species comparisons, and the second highest when also the third closely-related species were considered (Fig. 2c), strongly supporting, on the basis of the mtDNA *cytb*, the species hypothesis for *Eliurus sp. nova*. While no new *Eliurus* species has been described in Montagne d'Ambre, the identification of a new *Eliurus* taxon sampled in the northwest, sister of *E. myoxinus*, was reported in an unpublished study using two mtDNA and three nuclear loci (poster in Tysdal and Jansa, 2014). Future genomic studies including the *Eliurus* individuals sampled in the northwest by Tysdal and Jansa, (2014) will confirm whether the *Eliurus* sp. nova identified in the present study and the new taxon in Tysdal and Jansa, (2014) belong to the same lineage.

5.13.2 Nuclear genomic diversity of *Eliurus* species in northern Madagascar

Up to now, genetic data on *Eliurus* species were limited to few mtDNA (*cytb*, *ND2* and *D-loop*) and nuclear loci (intron FGB, exon vWF, exon IRBP, exon BRCA1, exon GHR, exon RAG1, and one intron and two exons of Acp5) (Jansa et al., 1999; Goodman et al., 2009; Rakotoarisoa et al., 2013a; Tysdal and Jansa, 2014; Steppan and Schenk, 2017; Jansa et al., 2019). Phylogenetic inference using mtDNA or mtDNA+nuclear loci were generally concordant on the evolutionary relationships within the *Eliurus* genus (Jansa

and Weksler, 2004; Tysdal and Jansa, 2014; Steppan and Schenk, 2017; Jansa et al., 2019). This is why we used mtDNA phylogeny as a 'reference' for the species assignment of the 125 *Eliurus* individuals for which we have generated genome-wide data (RAD-seq). The comparison of individuals genotyped at both mtDNA and nuclear DNA have confirmed the presence of five *Eliurus* taxa in our dataset (Fig. S3). The identification of the five *Eliurus* taxa using neighbour-joining tree analysis was followed by extensive maximum-likelihood phylogenetic inferences on several genomic datasets, which were generated from the same initial 125 individual's dataset, differing in number of individuals per species, geographic distribution and list of genotyped loci (Fig. S7). The results on genome-wide data have consistently confirmed the tree topology recovered from mtDNA *cytb*, and in particular, the identification of the *Eliurus sp. nova*. This was true also for the partitioned phylogenetic analysis and single-gene phylogenies, which were performed for taking into account gene tree variation across the genome (substitution model, branch lengths and topology) (Fig. S8). Genome-wide PCA and clustering analysis have further confirmed the genomic differentiation across the five *Eliurus* taxa (Fig. 3 and Fig. S10). We did not find any signature of admixed co-ancestry among *Eliurus* species, which lead to exclude events of recent admixture or hybridization among them. The congruence obtained among our different analyses can be considered a good indicator of true evolutionary signal in the data (Hillis, 1995; Rubin et al., 2012), and therefore we hypothesise that, according to the phylogenetic species concept (Rosen, 1979; Baum and Shaw, 1995), *Eliurus sp. nova* may correspond to an undescribed *Eliurus* species, possibly the same as that mentioned by Tysdal and Jansa, (2014).

5.13.3 *Eliurus* species delimitation

It is well understood now that concatenation models of gene sequences, which assumes topologically congruent genealogies across the genome, oversimplify the genome-wide variability of the biological processes

(e.g., incomplete lineage sorting, recombination, ancient coalescence) (Rannala and Yang, 2003; Edwards et al., 2016). The multispecies-coalescent (MSC) model has been proposed as an alternative model able to capture such variability and estimate gene trees, together with species trees, by taking into account the diversity of genealogical histories in the genome (Rannala and Yang, 2020, 2003). The MSC model has been shown to perform better for species identification than genetic distance approaches that rely on empirical distance thresholds (Yang and Rannala, 2017). Moreover, compared to concatenation models for species tree inference, recent studies have shown that MSC-based approaches outperform concatenation ones (Jiang et al., 2020). For instance, Jiang et al., (2020) have re-analysed 47 phylogenomic datasets collected across several taxa and compared the accuracy of concatenation and MSC models in inferring species trees. They found that i) all 47 datasets were rejecting the hypothesis of congruent gene trees across the genome, ii) 38% of gene trees across all datasets were rejecting the concatenation tree, and that iii) 11% of loci rejected the MSC model, thus, showing that a significantly lower percentage of loci reject MSC model compared to concatenation models. Thus, we decided to perform additional analysis under the MSC model implemented in BPP, to test the species delimitation hypothesis of the three closely-related species: *E. minor*, *E. myoxinus* and *Eliurus sp. nova*. Our results strongly support the separation of the three taxa as separate species, since ((*E. myoxinus*, *Eliurus sp. nova*), *E. minor*) was the species tree with highest support ($P_{(3)} = 1$; Fig. 4), across all runs and combinations of priors.

Two recent studies (Jackson et al., 2017; Leaché et al., 2019) have shown that even in cases where two populations (or putative species) exchange genes at a relatively high rate (Nm = 10), such that they should be considered nearly one population, MSC-based species delimitation approaches (e.g., implemented PHRAPL and BPP) would select a two-species model over a one-species model. In other words, these results have shown that Bayesian model selection in MSC-based approaches may favour

species over-split in case of ongoing gene-flow between populations, and thus mis-interpret population structure as evidence for species delimitation. To overcome this limitation in MSC-based species delimitation models, Jackson et al., (2017) have proposed the *gdi* index, which calculates the degree of non-monophyly in a set of gene-trees. This index provides a continuous measure, that uses information on the estimated divergence times and population sizes for a fixed species tree under the MSC, to assess where, a given taxon, lies on the path of speciation. Although the *gdi* index does not really solve the dichotomy of population structure *versus* species delimitation in MSC models, it allows to go beyond the discrete model selection (e.g., one-species *versus* two-species). *A priori*, we may exclude the possibility of ongoing gene-flow among the three closely-related taxa (*E. minor*, *E. myoxinus* and *Eliurus sp. nova*) for which we have performed species delimitation test. In fact, clustering and PCA analyses have clearly identified five genetic clades corresponding to the five *Eliurus* taxa, with no evidence of admixture. Moreover, the study region from which we have sampled *Eliurus* individuals is fragmented with open habitat (Fig. 1), and thus it seems unlikely that ongoing gene-flow may occur between forest fragments. However, we note that some *Eliurus* species show tolerance to forest disturbance (Stephenson, 1995; Lehtonen et al., 2001; Goodman et al., 2013), which may suggest that forest fragmentation may not necessarily represent a strict barrier to gene-flow. These observations leave open the possibility of an over-split due to population structure with ongoing gene-flow within *E. myoxinus*. This is why, in addition to species delimitation analysis under MSC, we also computed the *gdi* index using the parameters estimated under MSC with ((*E. myoxinus*, *Eliurus sp. nova*), *E. minor*) as species tree. The *gdi*-index we calculated across runs and combinations of priors was consistently higher than 0.999. Based on Pinho and Hey, (2010) meta-analysis, Jackson et al., (2017) proposed that $gdi < 0.2$, should be interpreted as evidence for one single species, a $gdi > 0.7$ as evidence for two-species model, and $0.2 \leq gdi \leq 0.7$ as ambiguous delimitation. According to this rule of thumb, our gdi

results strongly support the delimitation of *Eliurus sp. nova* a separate species.

In addition to MSC-based species delimitation and *gdi*-index, we also performed an *ad hoc* test, aimed at evaluating the extent on which the genetic clade assigned to *Eliurus sp. nova* is actually the result of insufficient taxon sampling and spatial genetic structure within the sister-taxa *E. myoxinus*. The reasoning underlying such test is that species can be characterized by geographically limited dispersal, such that individuals geographically close tend to be genetically more related than individuals far apart. The signal of spatial genetic structure due to limited dispersal (isolation-by-distance, IBD) can create cline in genetic diversity that, if not taken into account and in case of insufficient taxon sampling, can lead to misinterpret intraspecies population structure as evidence for species delimitation (Barley et al., 2018; Hausdorf and Hennig, 2020; Mason et al., 2020). We have shown that mtDNA identified *E. myoxinus* individuals exhibit signatures of IBD at the nuclear genome, suggesting that *E. myoxinus* belong to those species characterised by geographically-limited dispersal. Therefore, we compared IBD pattern within *E. myoxinus* and *Eliurus sp. nova*, with IBD computed only among *E. myoxinus* - *Eliurus sp. nova* individual pairs. If *Eliurus sp. nova* individuals are not a population of *E. myoxinus*, we would expect i) no IBD for *E. myoxinus* - *Eliurus sp. nova* comparisons; and ii) significantly higher genetic distances between taxa than within taxa, keeping in mind that genetic differentiation is expected to reach plateau (stationary value of differentiation) for large geographical distances (Rousset, 1997, 2000). In this regard, our results have shown that it seems unlikely that *Eliurus sp. nova* is the result of insufficient taxon sampling and spatial genetic structure within the sister-taxa *E. myoxinus* (Fig. 4c, d).

Finally, we quantified the percentage of GC content across the nuclear genome of each *Eliurus* species (Fig. 4e). We found significant differences in GC content for all *Eliurus* species pairwise comparisons, expect *E. ellermani*-*E. minor*-*Eliurus* sp. nova. Most importantly, we showed significant

differences between the two sister-taxa *Eliurus* sp. nova and *E. myoxinus*. Quantify GC content across species can be informative on differences in genomic architecture, since it has been shown that GC content correlates with several genomic features, such as gene density, methylation rate, recombination rate, and expression levels (Eyre-Walker and Hurst, 2001; Kudla et al., 2006). The small differences recovered among *Eliurus* species are consistent with GC content pattern across placental mammals (Romiguier et al., 2010). Romiguier et al., (2010) compared GC content at the third codon of over 1,000 orthologous genes across 33 species and showed similar small differences, for instance, between *Homo sapiens* (GC3 = 46.1%) and *Pan troglodytes* (GC3 = 46.09%), or between *Rattus norvegicus* (GC3 = 51.46%) and *Mus musculus* (GC3 = 51.24%).

5.13.4 *Eliurus* species distribution in northern Madagascar

Among the five *Eliurus* taxa herein studied, *E. myoxinus* and *E. minor* are the species with the widest distribution across Madagascar, respectively ~378,000 km^2 and ~261,423 km^2 (Fig. S16). Prior to this study, *E. myoxinus* was mostly known from the dry deciduous forests and xerophytic scrub bush of western Madagascar, with few reported occurrences in the humid forests of northern-east Madagascar (Puy and Moat, 1996; Soarimalala and Goodman, 2003; Shi et al., 2013). The present study reveals frequent occurrences of *E. myoxinus* also in northern Madagascar, in both dry and lowland humid forests at altitudes ranging from 40 – 1284 m (Fig. S15). In particular, we found that among the sites where *E. myoxinus* individuals have been sampled, 77% (14/18) are characterised by humid habitat and 22% (4/18) by dry habitat (Fig. 1 and Tab. 1).

In northern Madagascar, the first reporting of *E. myoxinus* individuals in humid habitats back to the work of Goodman and Soarimalala, (2002) and Soarimalala and Goodman, (2003), documenting the presence of this species in the Special Reserve of Manongarivo (in the northwest), in the humid forests

of Loky-Manambato protected area (in the north) and National Park Marojejy (in the northeast). Morphological comparisons of *E. myoxinus* individuals collected from the south to the north of Madagascar showed a cline in body size, from the largest size in the south to the smallest in the north (Carleton and Goodman, 2007). Such decline in body size in the direction of the most humid northern populations, coincides with the increase in morphological similarity of *E. myoxinus* with another *Eliurus* species known to be associated with humid habitats of north and east Madagascar, *E. minor*. The increasing similarity with the humid-associated *E. minor*, and the occurrence of *E. myoxinus* in both dry and mesic habitats, may point out to a mechanism of local adaptation or phenotypic plasticity. However, the fact that the other two closely-related *Eliurus* lineages (*E. minor* and *Eliurus sp. nova*) are humid-associated species may suggest that *E. myoxinus* was originally adapted to humid habitat. Assuming a mechanism of local adaptation to dry habitats from an ancestral humid-associated *E. myoxinus*, the increasing aridification of forested habitat during the late Pleistocene – mid Holocene may have promoted local adaptation and expansion towards the south. A similar hypothesis has been proposed in the unpublished work of Tysdal and Jansa, (2014), where the authors compared *Eliurus* phylogeny with the habitat type to which each *Eliurus* species is associated. They found that i) seven out of eleven *Eliurus* species occurs in humid habitats, and that ii) dry-associated *Eliurus* species are not necessarily phylogenetically related. Therefore, they proposed that most speciation in the *Eliurus* genus occurred in humid and evergreen forests, and that at least three independent events of adaptation to dry habitats occurred within the genus, specifically for: i) *E. myoxinus*; ii) *E. tsingimbato*; and iii) *E. carletoni* + *E. antsingy*.

With this study, we also report, for the first-time, evidence for the presence of *E. carletoni* in humid habitats of northern Madagascar, corresponding to 24% (5/21) of *E. carletoni* sampling sites (e.g., ANDRA, ANDR, SAL, BEZ and ANALV; Fig. 1). Similar findings have been reported

by Goodman et al., (2009), which sampled *E. carletoni* individuals in mesic vegetations within the deep canyons of the Ankarana Massif. The presence of *E. carletoni* in both dry and humid habitats could be explained by i) ecological tolerance, allowing this species to occupy areas of moister habitats; or ii) the presence in mostly humid localities, of dry microclimatic conditions that favours *E. carletoni* persistence. Together with the observations on *E. myoxinus*, we can then state that among the dry-associated *Eliurus* species, *E. myoxinus* and *E. carletoni* are those for which there are also populations living in humid habitats, which make them two interesting study cases for possibly studying adaptation to dry and humid habitats. At this stage, this scenario is largely speculative and it will require genetic and detailed phenotypic data across the whole species range to test such hypothesis and assess to which extent dry-adaptation is due to genetic adaptation or phenotypic plasticity.

In the dry southern Madagascar, previous studies have reported the presence of *E. myoxinus* in the small moisty areas of the canyons of Isalo and Analavelona Massif (Carleton et al., 2001). These two localities have been suggested to be Quaternary relics of wetter climatic conditions across this region (Burney, 1993; Goodman and Rakotozafy, 1997; Raxworthy and Nussbaum, 1997; Carleton et al., 2001), further supported by evidence of ancient connectivity between the present-day humid eastern and dry western Madagascar (Goodman and Rakotozafy, 1997; Jansa et al., 1999; Muldoon et al., 2009; Steven M. Goodman et al., 2013; Yoder et al., 2016). In fact, Jansa et al., (1999) found in the canyons of Isalo (in the southwest), together with *E. myoxinus*, also an *Eliurus* form related to the eastern *E. majori*. Muldoon et al., (2009) found in the Ankilitelo Cave (in the south-west) subfossils of shrew tenrecs, carnivorans and rodents which have modern distributions in areas more humid than the present-day dry southwestern Madagascar. Yoder et al., (2016) found stronger genetic proximity between two mouse lemurs (*Microcebus*) occurring in different ecoregions, the humid

forests of the east (*M*. *rufus*) and the dry deciduous forests of the west (*M*. *berthae*), than mouse lemurs occurring in the same ecoregion, suggesting that forest-dwelling species such as *Microcebus* could disperse across the Central Highlands in the ancient past.

Assuming that *E. myoxinus* was originally restricted to humid habitats, we propose that also the humid forest fragments where *E. myoxinus* has been found in northern Madagascar represents relics of an ancient wetter period. In fact, these humid forest fragments coincide with sites that have been suggested to represent humid refugia (Sgarlata et al., 2019), when dry conditions increased in northern Madagascar during the Holocene (Jungers et al., 1995; Simons et al., 1995; Godfrey et al., 1999). For instance, Sgarlata et al., (2019) have reported the presence in these fragmented humid forests of another small mammal, the Arnhold's mouse lemur (*Microcebus arnholdi*), which until 2019 was known only from Montagne d'Ambre (in the far north) and from one site in north-west Madagascar (Weisrock et al., 2010). The discovery of new *M. arnholdi* populations in fragmented humid forests surrounded by dry forests has led the authors, supported by fossil evidences (Jungers et al., 1995; Simons et al., 1995; Godfrey et al., 1999), to hypothesise that in the recent past northern Madagascar was most likely covered by humid habitat, and that following the aridification of the Holocene only few mountain forests could maintain humid conditions suitable to the persistence of humid-adapted species. According to this scenario, only some populations of *E. myoxinus* and *M. arnholdi* have persisted during the dry conditions of the Holocene. Similar scenarios have been proposed to explain the demographic history, and distribution of two *Eliurus* species in northern Madagascar (*Eliurus carletoni* and *Eliurus tanala*; Rakotoarisoa et al., 2013a, b), sifaka (*Propithecus tattersalli*; Quéméré et al., 2012; Salmona et al., 2017) and the leaf chameleons (e.g., *Brookesia ebenaui* and *Brookesia minima*; Raxworthy and Nussbaum, 1995). Future population genomic analyses may shed light on the time scale of forest cover change in northern Madagascar

and shed light on the environmental changes that have shaped present-day biodiversity in this region.

Table 1. Localities and summary statistics of nuclear genomic data of the *Eliurus* individuals sampled across northern Madagascar. Forest: complete forest name; Site: forest abbreviation; Nloci: number of genotyped loci; NFWReads: number of used forward reads; Coverage: mean coverage; WCoverage: weighted mean coverage calculated by giving more weight to loci that are more represented across the dataset; N: number of samples used for genomic analyses; Tot: number of samples in the full dataset (347 individuals); Sympatry: co-occurring *Eliurus* species.

Species	Forest	Site	Latitude	Longitude	Nloci	NFWReads	Coverage	WCoverage	N	Tot	Sympatry
E. carletoni	Ambohitsitondroina	AMBO	-13.13039	49.4511	131783	1802023	13.674	14.158	3	8	/
E. carletoni	Analabe	ANALABE	-12.76921	49.04428	/	/	/	/	0	1	/
E. carletoni	Analamerana	ANALM	-12.81209	49.38934	122720	1091989	9.316	9.615	3	22	/
E. carletoni	Ampasimaty	ANALM_2	-12.76715	49.47568	121271	1507823	11.6145	12.1625	2	9	/
E. carletoni	Analalava	ANALV	-13.61952	50.00191	96117	486790	5.065	5.102	1	1	*E. myoxinus*
E. carletoni	Andrafiamena	ANDF	-12.90593	49.33555	137048	2351686	16.8825	17.767	4	7	*E. myoxinus*
E. carletoni	Andoharano	ANDR	-13.77516	49.52557	152396	4155229	27.266	29.417	1	1	*E. myoxinus , E. minor*
E. carletoni	Andraharikely	ANDRA	-13.6246	49.24428	131829	2182490	16.555	17.204	1	1	*E. myoxinus*
E. carletoni	Andrafiabe_tenan'ankarana	ANDRAF	-12.95576	49.04309	124751	1667971.5	13.0695	13.708	4	12	/
E. carletoni	Ankaramy	ANKAMY	-13.27271	49.65465	150178	3329347	22.169	23.636	1	2	/
E. carletoni	Antsahabe	ANTSB	-13.20946	49.56042	131833.5	1547645	11.6695	12.154	2	6	*E. myoxinus , E. ellermani*
E. carletoni	Bekaraoka	BEK	-13.16378	49.70763	127615.5	1661317	12.6925	13.233	2	16	/
E. carletoni	Bezavona - Ankirendrina	BEZ	-13.55742	49.89869	127616	1757407	13.771	14.242	1	1	*E. myoxinus , E. ellermani*
E. carletoni	Binara	BIN	-13.25624	49.61579	121874	1456062	11.947	12.345	1	3	*E. myoxinus , E. ellermani*
E. carletoni	Bobankora	BOB	-13.24654	49.76853	129350	2260666.5	17.297	18.157	2	11	/
E. carletoni	Mahamasina/Ankarana	MAHAR	-12.94873	49.11851	149625	2928045.5	19.5705	20.7885	2	11	/
E. carletoni	Marovato/Ankarana	MAROV	-12.84803	49.13856	152650	3896453	25.324	27.709	3	15	/
E. carletoni	Salafaina	SAL	-13.44336	49.72103	138273.5	3280349.5	24.0055	25.279	2	6	*E. myoxinus*
E. carletoni	Antserasera/Masiamposa	SERA	-12.95584	49.20876	132829	2809299	21.15	22.189	3	8	/
E. carletoni	Solaniampilana	SOL	-13.0925	49.58031	145733.5	2398163	16.4195	17.2745	4	8	*E. myoxinus*
E. carletoni	Andavakoera	TSAKA	-13.1176	49.21146	119764	1298170	10.839	11.209	1	1	/
E. carletoni		TOT							43	150	43% sites

Species	Forest	Site	Latitude	Longitude	Nloci	NFWReads	Coverage	WCoverage	N	Tot	Sympatry
E. ellermani	Ambaliha	AMBA	-13.65282	49.55326	113191	1653024	14.604	15.288	1	1	*E. minor*
E. ellermani	Ambalasoa	AMBALA	-13.89701	49.44081	134933	1756733	13.019	13.741	1	1	*E. minor*
E. ellermani		ANDOHA	-14.16233	49.2418	/	/	/	/	0	1	/
E. ellermani	Ankinjanala	ANKI	-13.34506	49.44872	132482.5	1719220.5	12.6295	13.29	2	2	*E. myoxinus*
E. ellermani	Antsahabe	ANTSB	-13.21235	49.55932	127564	1415416	10.831	11.328	5	9	*E. myoxinus , E. carletoni*
E. ellermani	Bezavona - Ankirendrina	BEZ	-13.55604	49.87817	118304	2110304	17.838	18.423	1	2	*E. myoxinus , E. carletoni*
E. ellermani	Binara	BIN	-13.28065	49.60196	135273	2272588.5	16.7775	17.6855	4	6	*E. myoxinus , E. carletoni*
E. ellermani	Anketrakabe	KABE	-13.45813	49.56894	138659.5	2715244	19.3405	20.564	2	3	*E. myoxinus*
E. ellermani	Montagne d'Ambre	M_AMBRE_NORTH	-12.52451	49.18241	149407	2745101.5	18.399	19.641	2	2	*Eliurus sp. nova*
E. ellermani	Montagne d'Ambre	M_AMBRE_SOUTH	-12.69369	49.16734	/	/	/	/	0	4	*Eliurus sp. nova*
E. ellermani		TOT							18	31	90% sites

Species	Forest	Site	Latitude	Longitude	Nloci	NFWReads	Coverage	WCoverage	N	Tot	Sympatry
E. minor	Ambaliha	AMBA	-13.65099	49.55507	98912	986247.5	9.9235	10.3345	2	5	*E. ellermani*
E. minor	Ambalasoa	AMBALA	-13.91712	49.40907	133997	1820287	12.484	13.4	3	6	*E. ellermani*
E. minor	Andohanibemarivo	ANDOHA	-14.1517	49.24149	117323	1750474	15.043	15.989	3	7	/
E. minor	Andoharano	ANDR	-13.75711	49.56868	158304	2623176	16.466	17.734	5	9	*E. myoxinus, E. carletoni*
E. minor	Antsahavary	ANTSA	-13.55831	49.53822	107856	1250540	11.595	12.152	3	7	/
E. minor	Beamalonkely	BEAMA	-14.22791	49.64715	114319	1327160	11.609	12.002	1	1	/
E. minor	Lohananjialava	LOHANA	-14.18806	49.39257	157708	3158713	20.029	21.914	1	3	/
E. minor	Maladialina	MAL	-13.75644	49.40949	119975	2000167.5	14.2995	15.2495	2	4	/
E. minor		TOT							20	42	38% sites

Species	Forest	Site	Latitude	Longitude	Nloci	NFWReads	Coverage	WCoverage	N	Tot	Sympatry
E. myoxinus	Ambatomainty	AMBATO	-14.25017	49.61373	146089	3047060	20.858	21.823	1	2	/
E. myoxinus	Analafiana	ANALF	-13.4542	49.82948	149027	3950060.5	26.5015	28.0945	2	3	/
E. myoxinus	Analalava	ANALV	-13.61609	49.991	125739	1354342	10.771	10.96	1	2	*E. carletoni*
E. myoxinus	Andrafiamena	ANDF	-12.9226	49.32506	147004	3474441	23.635	24.863	1	3	*E. carletoni*
E. myoxinus	Andoharano	ANDR	-13.74913	49.59581	145232	2645416	18.215	19.18	1	1	*E. carletoni*
E. myoxinus	Andraharikely	ANDRA	-13.62457	49.24422	141238	2658730	18.4	19.3365	2	8	*E. carletoni*
E. myoxinus	Andranomenabe	ANDRANO	-14.24828	49.74902	137275	3005142	21.891	22.711	1	2	/
E. myoxinus	Ankinjanala	ANKI	-13.35139	49.44964	127697	1352955	10.595	10.842	1	12	/
E. myoxinus	Antsahavary	ANTSA	-13.5416	49.54489	149623	2992206	19.998	21.226	1	7	/
E. myoxinus	Antsahabe	ANTSB	-13.21014	49.56031	150325	3274397	21.802	23.038	4	8	*E. carletoni, E. ellermani*
E. myoxinus	Beamalonkely	BEAMA	-14.24036	49.65551	123195	759389	6.164	6.258	1	3	/
E. myoxinus	Bezavona - Ankirendrina	BEZ	-13.55385	49.89297	139890	2373081.5	16.6315	17.3745	2	12	*E. carletoni, E. ellermani*
E. myoxinus	Binara	BIN	-13.27597	49.61046	146916	2795339	18.947	19.932	7	18	*E. carletoni, E. ellermani*
E. myoxinus	Anketrakabe	KABE	-13.46114	49.5738	142213.5	3082868.5	21.091	22.2835	2	11	/
E. myoxinus	Lohananjialava	LOHANA	-14.19134	49.38949	/	/	/	/	0	1	/
E. myoxinus	Salafaina	SAL	-13.44309	49.71178	132691	1682583	12.68	12.995	3	11	*E. carletoni*
E. myoxinus	Solaniampilana	SOL	-13.09044	49.57916	131809	1826501	13.857	14.188	1	1	*E. carletoni*
E. myoxinus	Andavakoera	TSAKA	-13.11661	49.22099	149323.5	4330517	28.8305	30.7735	4	13	*E. carletoni*
		TOT							35	118	56% sites

Species	Forest	Site	Latitude	Longitude	Nloci	NFWReads	Coverage	WCoverage	N	Tot	Sympatry
Eliurus sp. nova	Montagne d'Ambre	M_AMBRE_NORTH	-12.51457	49.18181	150944.5	2693287.5	17.64	18.446	2	4	*E. ellermani*
Eliurus sp. nova	Montagne d'Ambre	M_AMBRE_SOUTH	-12.67684	49.16095	153288.5	2830566	18.289	19.2255	6	9	*E. ellermani*
Eliurus sp. nova		TOT							8	13	100% sites

5.13.5 *Eliurus* sympatry: more common than rare

In the present study, we also document frequent cases of *Eliurus* species co-occurrence (sympatry and sintopy) (Tab. S5; Fig. 1). In particular, we found a co-occurrence frequency of 43%, 90%, 38% and 56%, respectively for *E. carletoni*, *E. ellermani*, *E. minor* and *E. myoxinus*. *Eliurus sp. nova* is currently known only from Montagne d'Ambre, and thus in co-occurrence with *E. ellermani*. Future studies with samples from the north-west may reveal further information on this new *Eliurus* taxon (Tysdal and Jansa, 2014) and with which *Eliurus* species it may co-occur.

Eliurus records collected in previous studies have shown that sympatry between *Eliurus* species is common (Carleton, 1994; Jansa et al., 2019; Tab. S5). As pointed out by Goodman et al., (2013), however, *Eliurus* sympatry rarely involves closely-related species. The genetic analyses we performed, such as ADMIXTURE or PCA, did not reveal signature of admixture between sympatric *Eliurus* species, suggesting that even if in co-occurrence, these species are most likely reproductively isolated. Not many studies have looked at how several *Eliurus* species can co-exist in the same sites and what are the mechanism behind reproductive isolation. A PhD thesis by Katrin Marquart (Marquart, 2014), has however shown that niche

repartition may be one of the mechanisms that make *Eliurus* species coexistence possible. In fact, Marquart, (2014) observed that, among the four *Eliurus* taxa found in Maromizaha, north-eastern Madagascar (*E. minor*, *E. tanala*, *E. webbi*, *E. grandidieri*), some species could be preferentially captured at ground level and medium high vegetation (e.g., *E. tanala*), while others exclusively in high vegetation (*E. webbi*). The author also identified different behaviours across *Eliurus* taxa, some considered great climbers (*E. minor*) whereas others seen using ground level or swimming along riverside (*E. tanala*). Some of these differences in behaviour and microhabitat use were qualitatively correlated with differences in the morphology of the posterior hind (*E. webbi* versus *E. tanala*-*E. grandidieri*-*E. minor*), which led the author to conclude that certain morphological structures of *Eliurus* feet give evidence of a specific way of life and that foot pad morphology in particular, mirrors special adaptations to a species' habitat.

5.13.6 Conclusions and Future Directions

In the present work, we have assessed the species diversity of the *Eliurus* genus in northern Madagascar. We document the discovery of a new species, *Eliurus sp. nova*, in the mountain humid forest of Montagne d'Ambre, supporting the idea that this mountain forest contains extraordinary levels of biodiversity across several taxa and represent an 'island' of humid habitat surrounded by drier habitat (Raxworthy and Nussbaum, 1994; Louis et al., 2008; Rasolonjatovo et al., 2020). This work represents the first genomic study of the *Eliurus* genus. In addition to the discovery of a new taxa, we confirm the species status of the four previously known species through the use of thousands of nuclear genomic loci, therefore going beyond the use of few mtDNA and nuclear loci (two and three, respectively). We also showed that, in northern Madagascar, *E. myoxinus* occurs mostly in humid habitats, in contrast to what was previously known for this species. We also document the presence of the dry-associated *E. carletoni* in humid forests of northern

Madagascar. These findings might have important implications for our understanding of *Eliurus* evolution and diversification. In fact, the *E. myoxinus* populations identified in the humid forests of northern Madagascar may contain genetic information that could contribute to our understanding of the genomic basis of adaptation to dry and humid habitats. Ultimately, we confirmed that *Eliurus* species often co-exist in the same forest fragments, suggesting niche separation of *Eliurus* species within the same environment. Additional work is required to uncover the eco-evolutionary mechanisms that make such co-existence possible.

5.14 References

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19, 1655–1664.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics 17, 81–92.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLOS ONE 3, e3376.

Barley, A.J., Brown, J.M., Thomson, R.C., 2018. Impact of Model Violations on the Inference of Species Boundaries Under the Multispecies Coalescent. Systematic Biology 67, 269–284.

Baum, D.A., Shaw, K.L., 1995. Genealogical perspectives on the species problem., in: Experimental and Molecular Approaches to Plant Biosystematics. Missouri Botanical Garden, St. Louis, pp. 289–303.

Bésairie, H., 1964. Carte géologique de Madagascar, au 1:1,000,000e, trois feuilles en couleur.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology & Evolution 22, 148–155.

Buřivalová, Z., 2011. Remote sensing of vegetation in conservation: A case study from the dry and transitional forests of Andrafiamena, northern Madagascar. Université de Genève and Conservatoire et Jardin botaniques de la Ville de Genève, Genève.

Burney, D.A., 1993. Late Holocene Environmental Changes in Arid Southwestern Madagascar. Quaternary Research 40, 98–106.

Carleton, M., Goodman, S., 1998. New taxa of nesomyine rodents (Muroidea: Muridae) from Madagascar's northern Highlands, with taxonomic comments on previously described forms. Fieldiana Zoology 1998, 163–200.

Carleton, M.D., 2003. Eliurus, tufted-tailed rats., in: The Natural History of Madagascar. University of Chicago Press, Chicago, pp. 1373–1380.

Carleton, M.D., 1994. Systematic studies of Madagascar's endemic rodents (Muroidea: Nesomyinae) : revision of the genus Eliurus. American Museum novitates ; no. 3087.

Carleton, M.D., Goodman, S.M., 2007. A new species of the Eliurus majori complex (Rodentia, Muroidea, Nesomyidae) from south-central Madagascar, with remarks on emergent species groupings in the genus Eliurus ; American Museum novitates, no. 3547. Eliurus majori complex.

Carleton, M.D., Goodman, S.M., Rakotondravony, D., 2001. A new species of tufted-tailed rat, genus Eliurus (Muridae : Nesomyinae), from western Madagascar, with notes on the distribution of E-myoxinus.

Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J.H., 2011. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. G3 (Bethesda) 1, 171–182.

Chenuil, A., Cahill, A.E., Délémontey, N., Du Salliant du Luc, E., Fanton, H., 2019. Problems and Questions Posed by Cryptic Species. A Framework to Guide Future Studies, in: Casetta, E., Marques da Silva, J., Vecchi, D. (Eds.), From Assessing to Conserving Biodiversity: Conceptual and Practical Challenges, History, Philosophy and Theory of the Life Sciences. Springer International Publishing, Cham, pp. 77–106.

Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., Leaché, A.D., Liu, L., Davis, C.C., 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Molecular Phylogenetics and Evolution 94, 447–462.

Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., Cresko, W.A., 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol Biol 772, 157–178.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology 14, 2611–2620.

Everson, K.M., Goodman, S.M., Olson, L.E., 2020. Speciation and gene flow in two sympatric small mammals from Madagascar, Microgale fotsifotsy and M. soricoides (Mammalia: Tenrecidae). Molecular Ecology 29, 1717–1729.

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354
Eyre-Walker, A., Hurst, L.D., 2001. The evolution of isochores. Nat Rev Genet 2, 549–555.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 17, 368–376.

Flouri, T., Jiao, X., Rannala, B., Yang, Z., 2020. A Bayesian Implementation of the Multispecies Coalescent Model with Introgression for Phylogenomic Analysis. Molecular Biology and Evolution 37, 1211–1223.

Flouri, T., Jiao, X., Rannala, B., Yang, Z., 2018. Species Tree Inference with BPP Using Genomic Sequences and the Multispecies Coalescent. Molecular Biology and Evolution 35, 2585–2593.

Fowler, S.V., Chapman, P., Checkley, D., Hurd, S., McHale, M., Ramangason, G.-S., Randriamasy, J.-E., Stewart, P., Walters, R., Wilson, J.M., 1989. Survey and management proposals for a tropical deciduous forest reserve at Ankarana in Northern Madagascar. Biological Conservation 47, 297–313.

Garbutt, N., 1999. Mammals of Madagascar. Yale University Press.

Gautier, L., Ranirison, P., Nusbaumer, L., Wohlhauser, S., 2006. Aperçu des massifs forestiers de la région Loky-Manambato, in: Inventaires de La Faune

et de La Flore Du Nord de Madagascar Dans La Région Loky-Manambato, Analamerana et Andavakoera. Rech. Développement. Sér Sci Biol.

Glez-Peï¿½a, D., Gï¿½mez-Blanco, D., Reboiro-Jato, M., Fdez-Riverola, F., Posada, D., 2010. ALTER: program-oriented conversion of DNA and protein alignments. Nucleic Acids Res 38, W14–W18.

Godfrey, L.R., Jungers, W.L., Simons, E.L., Chatrath, P.S., Rakotosamimanana, B., 1999. Past and Present Distributions of Lemurs in Madagascar, in: Rakotosamimanana, B., Rasamimanana, H., Ganzhorn, J.U., Goodman, S.M. (Eds.), New Directions in Lemur Studies. Springer US, Boston, MA, pp. 19–53.

Goodman, S.M., Carleton, M.D., 1996. The rodents of the Réserve Naturelle Intégrale d'Andringitra, Madagascar. Fieldiana: Zoology (new series) 257–283.

Goodman, S.M., Carleton, M.D., Pidgeon, M., 1999. The rodents of the Réserve Naturelle Intégrale d'Andohahela, Madagascar. Fieldiana: Zoology (new series) 217–249.

Goodman, S.M., Ganzhorn, J.U., 1994. Les petits mammiferes, in: Inventaire Biologique Foret de Zombitse,Recherches Pour Le Developpement, Serie Sciences Biologiques. Centre d'Information et de Documentation Scientifique et Technique, Antananarivo, Madagascar, pp. 58–63.

Goodman, S.M., Raheriarisena, M., Jansa, S.A., 2009. A new species of Eliurus Milne Edwards, 1885 (Rodentia: Nesomyinae) from the Réserve Spéciale d'Ankarana, northern Madagascar | Request PDF. Bonner Zoologische Beiträge 56, 133–149.

Goodman, Steven M., Raherilalao, M.J., Muldoon, K., 2013. Bird fossils from Ankilitelo Cave: inference about Holocene environmental changes in southwestern Madagascar. Zootaxa 3750, 534–548.

Goodman, S.M., Raherilalao, M.J., Wohlhauser, S. (Eds.), 2021. The Terrestrial Protected Areas of Madagascar: Their History, Description, and Biota, Volume 3: Western and southwestern Madagascar - Synthesis. Association Vahatra in Antananarivo.

Goodman, S.M., Rakotozafy, L.M.A., 1997. Subfossil birds from coastal sites in western and southwestern Madagascar: a paleoenvironmental reconstruction, in: Natural Change and Human Impact in Madagascar. Smithsonian Institution Press, Washington, D. C., pp. 257–279.

Goodman, S.M., Raselimanana, A.P., Wilmé, L., 2007. Inventaires de la faune et de la flore du couloir forestier d'Anjozorobe-Angavo. Centre d'Information et de Documentation Scientifique et Technique.

Goodman, S.M., Rasoloarison, R., 1997. Les petits mammiferes, in: Inventaire Biologique Foret de Vohibasia et d'Isoky-Vohimena. Recherches Pour Le Developpement, Serie Sciences Biologiques. Centre d'Information et de Documentation Scientifique et Technique, Antananarivo, Madagascar, pp. 144–155.

Goodman, S.M., Rasolonandrasana, B.P.N., 1999. Inventaire biologique de la Réserve spéciale du Pic d'Ivohibe et du couloir forestier qui la relie au Parc national d'Andringitra. Centre d'information et de documentation scientifique et technique, Antananarivo, Madagascar.

Goodman, S.M., Soarimalala, V., 2002. Les petits mammifères de la Réserve Spéciale de Manongarivo, Madagascar, in: Inventaire Floristique et

Faunistique de La Réserve Spéciale de Manongarivo (NW Madagascar). pp. 383–401.

Goodman, S. M., Soarimalala, V., Raheriarisena, M., Rakotondravony, D., 2013. Small mammals or tenrecs (Tenrecidae) and rodents (Nesomyinae), in: Atlas of Selected Land Vertebrates of Madagascar. Association Vahatra, Antananarivo, pp. 211–269.

Goodman, S.M., Wilmé, L., 2006. Inventaires de la faune et de la flore du nord de Madagascar dans la région Loky-Manambato, Analamerana et Andavakoera.

Hausdorf, B., Hennig, C., 2020. Species delimitation and geography. Molecular Ecology Resources n/a.

Hawkins, A., Chapman, P., Ganzhorn, J., Bloxam, Q., Barlow, S., Tonge, S., 1990. Vertebrate conservation in Ankarana special reserve, Northern Madagascar.

Heller, R., Nursyifa, C., Garcia-Erill, G., Salmona, J., Chikhi, L., Meisner, J., Korneliussen, T.S., Albrechtsen, A., 2021. A reference-free approach to analyse RADseq data using standard next generation sequencing toolkits. Molecular Ecology Resources 21, 1085–1097.

Hillis, D.M., 1995. Approaches for Assessing Phylogenetic Accuracy. Systematic Biology 44, 3–16.

Jackson, N.D., Carstens, B.C., Morales, A.E., O'Meara, B.C., 2017. Species Delimitation with Gene Flow. Syst Biol 66, 799–812.

Jansa, S.A., Carleton, M.D., 2003. Systematics and phylogenetics of Madagascar's native rodents., in: The Natural History of Madagascar. University of Chicago Press, Chicago.

Jansa, S.A., Carleton, M.D., Soarimalala, V., Rakotomalala, Z., Goodman, S.M., 2019. A review of the Eliurus tanala complex (Rodentia, Muroidea, Nesomyidae), with description of a new species from dry forests of western Madagascar. (Bulletin of the American Museum of Natural History, no. 430).

Jansa, S.A., Goodman, S.M., Tucker, P.K., 1999. Molecular Phylogeny and Biogeography of the Native Rodents of Madagascar (Muridae: Nesomyinae): A Test of the Single-Origin Hypothesis. Cladistics 15, 253–270.

Jansa, S.A., Weksler, M., 2004. Phylogeny of muroid rodents: relationships within and among major lineages as determined by IRBP gene sequences. Molecular Phylogenetics and Evolution 31, 256–276.

Jiang, X., Edwards, S.V., Liu, L., 2020. The Multispecies Coalescent Model Outperforms Concatenation Across Diverse Phylogenomic Data Sets. Systematic Biology 69, 795–812.

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.

Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070–3071.

Jungers, W.L., Godfrey, L.R., Simons, E.L., Chatrath, P.S., 1995. Subfossil Indri indri from the Ankarana Massif of northern Madagascar. Am J Phys Anthropol 97, 357–366.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15, 1179–1191.

Korneliussen, T.S., Albrechtsen, A., Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15, 356.

Kudla, G., Lipinski, L., Caffin, F., Helwak, A., Zylicz, M., 2006. High Guanine and Cytosine Content Increases mRNA Levels in Mammalian Cells. PLOS Biology 4, e180.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol Biol Evol 34, 772–773.

Leaché, A.D., Zhu, T., Rannala, B., Yang, Z., 2019. The Spectre of Too Many Species. Syst. Biol. 68, 168–181.

Lehtonen, J.T., Mustonen, O., Ramiarinjanahary, H., Niemelä, J., Rita, H., 2001. Habitat use by endemic and introduced rodents along a gradient of forest disturbance in Madagascar. Biodiversity and Conservation 10, 1185–1202.

Lewis, P.O., 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst Biol 50, 913–925.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio].

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Louis, E.E.L., Engberg, S.E., McGuire, S.M., McCormick, M.J., Randriamampionona, R., Ranaivoarisoa, J.F., Bailey, C.A., Mittermeier, R.A., Lei, R., 2008. Revision of the Mouse Lemurs, Microcebus (Primates, Lemuriformes), of Northern and Northwestern Madagascar with Descriptions of Two New Species at Montagne d'Ambre National Park and Antafondro Classified Forest. prco 23, 19–38.

Major, C.I.F., 1896. Descriptions of four additional new mammals from Madagascar. Annals and Magazine of Natural History 461–463.

Marquart, K., 2014. Habitat use and morphological adaptations of endemic rodents (Muroidea: Nesomyinae) of East Madagascar. Universität Hohenheim, Munich.

Maruki, T., Lynch, M., 2017. Genotype Calling from Population-Genomic Sequencing Data. G3: Genes, Genomes, Genetics 7, 1393–1404.

Mason, N.A., Fletcher, N.K., Gill, B.A., Funk, W.C., Zamudio, K.R., 2020. Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. Systematics and Biodiversity 18, 269–280.

Meisner, J., Albrechtsen, A., 2018. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. Genetics 210, 719–731.

Michaux, J., Reyes, A., Catzeflis, F., 2001. Evolutionary History of the Most Speciose Mammals: Molecular Phylogeny of Muroid Rodents. Mol Biol Evol 18, 2017–2031.

Milne Edwards, A., 1885. Description d'une nouvelle espèce de rongeur provenant de Madagascar. Annales des Sciences naturelles, Zoologie et Paléontologie.

Muldoon, K.M., de Blieux, D.D., Simons, E.L., Chatrath, P.S., 2009. The Subfossil Occurrence and Paleoecological Significance of Small Mammals at Ankilitelo Cave, Southwestern Madagascar. Journal of Mammalogy 90, 1111–1131.

Musser, G.G., Carleton, M.D., 2005. Superfamily Muroidea., in: Mammal Species of the World: A Taxonomic and Geo- Graphical Reference. Johns Hopkins University Press, Baltimore, pp. 894–1531.

Nei, M., 1972. Genetic Distance between Populations. The American Naturalist 106, 283–292.

Nicoll, M.E., Legrand, 1989. Madagascar : revue de la conservation et des aires protégées.

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y., Wang, J., 2012. SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data. PLOS ONE 7, e37558.

Paris, J.R., Stevens, J.R., Catchen, J.M., 2017. Lost in parameter space: a road map for stacks. Methods in Ecology and Evolution 8, 1360–1373.

Pinho, C., Hey, J., 2010. Divergence with Gene Flow: Models and Data. Annual Review of Ecology, Evolution, and Systematics 41, 215–230.

Poux, C., Madsen, O., Marquard, E., Vieites, D.R., de Jong, W.W., Vences, M., 2005. Asynchronous colonization of Madagascar by the four endemic clades of primates, tenrecs, carnivores, and rodents as inferred from nuclear genes. Syst. Biol. 54, 719–730.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Puy, D.D., Moat, J., 1996. A refined classification of the primary vegetation of Madagascar based on the underlying geology: Using GIS to map its distribution and to assess its conservation status.

Quéméré, E., Amelot, X., Pierson, J., Crouau-Roy, B., Chikhi, L., 2012. Genetic data suggest a natural prehuman origin of open habitats in northern Madagascar and question the deforestation narrative in this region. PNAS 109, 13028–13033.

R Core Team, 2021. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [WWW Document]. URL https://www.R-project.org/

Rakotoarisoa, J.E., Raheriarisena, M., Goodman, S.M., 2013. Late Quaternary climatic vegetational shifts in an ecological transition zone of northern Madagascar: insights from genetic analyses of two endemic rodent species. Journal of Evolutionary Biology 26, 1019–1034.

Rakotoarisoa, J. E., Raheriarisena, M., Goodman, S.M., 2013. A phylogeographic study of the endemic rodent Eliurus carletoni (Rodentia: Nesomyinae) in an ecological transition zone of Northern Madagascar. J. Hered. 104, 23–35.

Rakotoarisoa, J.-E., Raheriarisena, M., Goodman, S.M., 2010. Phylogeny and species boundaries of the endemic species complex, Eliurus antsingy and E. carletoni (Rodentia: Muroidea: Nesomyidae), in Madagascar using mitochondrial and nuclear DNA sequence data. Molecular Phylogenetics and Evolution 57, 11–22.

Rakotondravony, R., Radespiel, U., 2009. Varying patterns of coexistence of two mouse lemur species (Microcebus ravelobensis and M. murinus) in a heterogeneous landscape. American Journal of Primatology 71, 928–938.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2014. Tracer v1.6.

Rannala, B., Yang, Z., 2020. Species delimitation, in: Phylogenetics in the Genomic Era. Self published.

Rannala, B., Yang, Z., 2003. Bayes Estimation of Species Divergence Times and Ancestral Population Sizes Using DNA Sequences From Multiple Loci. Genetics 164, 1645–1656.

Rasolonjatovo, S.M., Scherz, M.D., Hutter, C.R., Glaw, F., Rakotoarison, A., Razafindraibe, J.H., Goodman, S.M., Raselimanana, A.P., Vences, M., 2020. Sympatric lineages in the Mantidactylus ambreensis complex of Malagasy frogs originated allopatrically rather than by in-situ speciation. Molecular Phylogenetics and Evolution 144, 106700.

Raxworthy, C.J., Nussbaum, R.A., 1997. Biogeographic patterns of reptiles in eastern Madagascar, in: Natural Change and Human Impact in Madagascar. Smithsonian Institution Press, Washington, D. C., pp. 124–141.

Raxworthy, C.J., Nussbaum, R.A., 1995. Systematics, speciation and biogeography of the dwarf chameleons (Brookesia; Reptilia, Squamata, Chamaeleontidae) of northern Madagascar. Journal of Zoology 235, 525–558.

Raxworthy, C.J., Nussbaum, R.A., 1994. A rainforest survey of amphibians, reptiles and small mammals at Montagne d'Ambre, Madagascar. Biological Conservation 69, 65–73.

Rochette, N.C., Catchen, J.M., 2017. Deriving genotypes from RAD-seq short-read data using Stacks. Nat Protoc 12, 2640–2659.

Rochette, N.C., Rivera-Colón, A.G., Catchen, J.M., 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Molecular Ecology 28, 4737–4754.

Romiguier, J., Ranwez, V., Douzery, E.J.P., Galtier, N., 2010. Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. Genome Res 20, 1001–1009.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Rosen, D.E., 1979. Fishes from the uplands and intermontane basins of Guatemala : revisionary studies and comparative geography. Bulletin of the AMNH ; v. 162, article 5.

Rossi, G., 1974. Morphologie et évolution d'un karst en milieu tropical : l'Ankarana (Madagascar)., in: Mém. et Doc. Du C.N.R.S. Vol. 15. Phénomènes Karstiques II. Paris, pp. 279–298.

Rousset, 2000. Genetic differentiation between individuals. Journal of Evolutionary Biology 13, 58–62.

Rousset, F., 1997. Genetic Differentiation and Estimation of Gene Flow from F-Statistics Under Isolation by Distance. Genetics 145, 1219–1228.

Rubin, B.E.R., Ree, R.H., Moreau, C.S., 2012. Inferring Phylogenies from RAD Sequence Data. PLOS ONE 7, e33394.

Salmona, J., Heller, R., Quéméré, E., Chikhi, L., 2017. Climate change and human colonization triggered habitat loss and fragmentation in Madagascar. Mol. Ecol. 26, 5203–5222.

Schmieder, R., Edwards, R., 2011. Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets. PLOS ONE 6, e17288.

Seutin, G., White, B.N., Boag, P.T., 1991. Preservation of avian blood and tissue samples for DNA analyses. Can. J. Zool. 69, 82–90.

Sgarlata, G.M., Salmona, J., Pors, B.L., Rasolondraibe, E., Jan, F., Ralantoharijaona, T., Rakotonanahary, A., Randriamaroson, J., Marques, A.J., Aleixo-Pais, I., Zoeten, T. de, Ousseni, D.S.A., Knoop, S.B., Teixeira,

H., Gabillaud, V., Miller, A., Ibouroi, M.T., Rasoloharijaona, S., Zaonarivelo, J.R., Andriaholinirina, N.V., Chikhi, L., 2019. Genetic and morphological diversity of mouse lemurs (Microcebus spp.) in northern Madagascar: The discovery of a putative new species? American Journal of Primatology 81, e23070.

Shi, J.J., Chan, L.M., Rakotomalala, Z., Heilman, A.M., Goodman, S.M., Yoder, A.D., 2013. Latitude drives diversification in Madagascar's endemic dry forest rodent Eliurus myoxinus (subfamily Nesomyinae). Biological Journal of the Linnean Society 110, 500–517.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7, 539.

Simons, E.L., Burney, D.A., Chatrath, P.S., Godfrey, L.R., Jungers, W.L., Rakotosamimanana, B., 1995. AMS 14C Dates for Extinct Lemurs from Caves in the Ankarana Massif, Northern Madagascar. Quaternary Research 43, 249–254.

Skotte, L., Korneliussen, T.S., Albrechtsen, A., 2013. Estimating individual admixture proportions from next generation sequencing data. Genetics 195, 693–702.

Soarimalala, V., Goodman, S.M., 2011. Les Petits Mammifères de Madagascar. Association Vahatra, Antananarivo, Madagascar.

Soarimalala, V., Goodman, S.M., 2003. Diversité biologique des micromammifères non volants (Liptotyphla et Rodentia) dans le complexe

Marojejy-Anjanaharibe-Sud. Recherches pour le Développement, Série Sciences Biologiques 231–278.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Stephenson, P.J., 1995. Small mammal microhabitat use in lowland rain forest of north-east Madagascar. Acta Theriologica 40, 425–438.

Steppan, S.J., Schenk, J.J., 2017. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. PLOS ONE 12, e0183070.

Struck, T.H., Feder, J.L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V.I., Kistenich, S., Larsson, K.-H., Liow, L.H., Nowak, M.D., Stedje, B., Bachmann, L., Dimitrov, D., 2018. Finding Evolutionary Processes Hidden in Cryptic Species. Trends in Ecology & Evolution 33, 153–163.

Tattersall, I., 2007. Madagascar's Lemurs: Cryptic diversity or taxonomic inflation? Evolutionary Anthropology: Issues, News, and Reviews 16, 12–23.

Tattersall, I., Cuozzo, F.P., 2019. Systematics of the extant Malagasy lemurs (order Primates), in: The Terrestrial Protected Areas of Madagascar: Their History, Description, and Biota. Association Vahatra, Antananarivo, Madagascar, pp. 403–424.

Tysdal, J.A., Jansa, S.A., 2014. Phylogeny and biogeography of an endemic Madagascar rodent.

Veress, M., Lóczy, D., Zentai, Z., Tóth, G., Schläffer, R., 2008. The origin of the Bemaraha tsingy (Madagascar). International Journal of Speleology 37.

Wang, L.-G., Lam, T.T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C.W., Jones, B.R., Bradley, T., Zhu, H., Guan, Y., Jiang, Y., Yu, G., 2020. Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. Molecular Biology and Evolution 37, 599–603.

Webb, C.S., 1954. The odyssey of an animal collector, [1st American ed.]. ed. Longmans, Green, New York.

Weisrock, D.W., Rasoloarison, R.M., Fiorentino, I., Ralison, J.M., Goodman, S.M., Kappeler, P.M., Yoder, A.D., 2010. Delimiting Species without Nuclear Monophyly in Madagascar's Mouse Lemurs. PLOS ONE 5, e9883.

Yang, Z., 2015. The BPP program for species tree estimation and species delimitation. Current Zoology 61, 854–865.

Yang, Z., Rannala, B., 2017. Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses. Molecular Ecology 26, 3028–3036.

Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. PNAS 107, 9264–9269.

Yoder, A.D., Campbell, C.R., Blanco, M.B., Reis, M. dos, Ganzhorn, J.U., Goodman, S.M., Hunnicutt, K.E., Larsen, P.A., Kappeler, P.M., Rasoloarison, R.M., Ralison, J.M., Swofford, D.L., Weisrock, D.W., 2016. Geogenetic patterns in mouse lemurs (genus Microcebus) reveal the ghosts of Madagascar's forests past. PNAS 113, 8049–8056.

Yu, G., 2020. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinformatics 69, e96.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.-Y., 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8, 28–36.

5.15 Supplementary Material

Table S1. List of *Eliurus* individuals used for the parameter's optimization analysis and for the construction of the *de-novo* catalog. Forest: complete forest name; Site: forest abbreviation; ID: identifier of the individual; Nloci: number of genotyped loci; Altitude: altitude of the capture site; NFWReads: number of used forward reads; Coverage: mean coverage; WCoverage: weighted mean coverage calculated by giving more weight to loci that are more represented across the dataset.

Forest	Site	ID	Latitude	Longitude	Altitude (m)	NFWReads	Nloci	Coverage	WCoverage	Species
Mahamasina/Ankarana	MAHAR	MAHAR.F19	-12.95819	49.11944	110	2987760	149227	20.022	21.255	*E. carletoni*
Salafaina	SAL	SAL.B64	-13.44856	49.72165	336	4922720	147346	33.409	35.418	*E. carletoni*
Solaniampilana	SOL	SOL.K35	-13.09551	49.57943	138	3453474	147705	23.381	24.735	*E. carletoni*
Andrafiamena	ANDF	ANDF.A11	-12.90433	49.33725	408	5024452	149154	33.686	36.042	*E. carletoni*
Ambohitsitondroina	AMBO	AMBO.F57	-13.13172	49.43522	132	2948307	148000	19.921	20.968	*E. carletoni*
Binara	BIN	BIN.K71	-13.28993	49.60661	767	3723416	143267	25.989	27.328	*E. ellermani*
Antsahabe	ANTSB	ANTSB.H21	-13.21329	49.5569	737	3561710	146529	24.307	25.794	*E. ellermani*
Anketrakabe	KABE	KABE.G09	-13.45726	49.56654	818	1886359	141833	13.3	13.858	*E. ellermani*
Ankinjanala	ANKI	ANKI.G52	-13.34159	49.45042	1084	2306297	143695	16.05	16.993	*E. ellermani*
Ambalasoa	AMBALA	AMBALA.B25	-13.89701	49.44081	918	1755775	131620	13.34	13.934	*E. ellermani*
Andoharano	ANDR	ANDR.E32	-13.74079	49.58126	1025	2695557	149061	18.084	19.135	*E. minor*
Ambalasoa	AMBALA	AMBALA.B72	-13.91378	49.41881	1075	1811511	143090	12.66	13.37	*E. minor*
Andoharano	ANDR	ANDR.E30	-13.763	49.6118	930	2499976	136600	18.301	19.342	*E. minor*
Maladialina	MAL	MAL.A69	-13.75956	49.41195	1014	3433107	153047	22.432	23.951	*E. minor*
Binara	BIN	BIN.BE78	-13.262	49.60656	822	4531347	147315	30.76	32.811	*E. myoxinus*
Andavakoera	TSAKA	TSAKA.D75	-13.11674	49.23788	474	3634867	145332	25.011	26.531	*E. myoxinus*
Andrafiamena	ANDF	ANDF.B51	-12.91797	49.32937	475	3457600	146125	23.662	24.916	*E. myoxinus*
Anketrakabe	KABE	KABE.G6	-13.46135	49.56617	818	4504087	145442	30.968	32.548	*E. myoxinus*
Salafaina	SAL	SAL.B81	-13.44763	49.70726	669	3654993	146804	24.897	26.261	*E. myoxinus*
Montagne d'Ambre	AMBRE_SOU	MBRE_SOUTH	-12.67702	49.17084	1002	1802583	144602	12.466	12.989	*Eliurus sp. Nova*

Table S2. Nuclear RAD-seq datasets generated for phylogenomic and genomic analysis. *White list* indicates the lists of loci used for genotyping within each dataset. *pop5.p4.r80*: includes loci present in at least 80% of individuals within species (indv rule), and in at least four out of five species (Pop rule). *pop5.p0.r80*: includes loci present in at least 80% of individuals within species, with no constraints on the number of species in which a locus is present in order to be genotyped. Sampling indicates the criteria used to compose each dataset. *All*: all individuals; *1 indv/site/species*: one individual per sampling site per species; *10 indv/species* or *3 indv/species*: ten or three individuals per species. STACKS: defined genotype calls. ANGSD: genotype uncertainty is associated to each site. N: number of individuals included in the dataset. #loci: number of genotyped loci; #length (bp): mean length of the assembled loci; #geno/sites (bp): mean number of sites genotyped per locus; #sites: total number of sites; #variants: total number of SNPs. #fixed: total number of fixed sites within but variable across individuals.

Dataset Name	White list	Sampling	Genotyping	N	#loci	#length (bp)	#geno/sites (bp)	#sites	#variants	#fixed	pop rule	indv rule
N125.pop1.p0	pop1.r80	All	STACKS	125	15258	747.5	609.26	497398	8471		NA	NA
N125.pop5.p4.min100	pop5.p4.r80	All	STACKS	125	9044	748.36	681.22	346893	108662		80%	80%
N62.pop5.p0	pop5.p0.r80	1 indv/site/species	STACKS	62	71879	739.6	625.7	956653	71847		NA	NA
N62.pop5.p4	pop5.p4.r80	1 indv/site/species	STACKS	62	24997	749.07	683.59	422000	24994		80%	NA
N62.pop5.p0.min50	pop5.p0.r80	1 indv/site/species	STACKS	62	22289	745.92	628.1	633923	22223		NA	80%
N62.pop5.p4.min50	pop5.p4.r80	1 indv/site/species	STACKS	62	22052	748.33	682.32	646587	22031		80%	80%
N48.pop5.p0	pop5.p0.r80	10 indv/species	STACKS	48	89210	741.83	636.4	931991	89184		NA	NA
N48.pop5.p4	pop5.p4.r80	10 indv/species	STACKS	48	89210	741.83	636.4	931991	89184		80%	NA
N48.pop5.p0.min38	pop5.p0.r80	10 indv/species	STACKS	48	25884	745.22	635.46	651165	25833		NA	80%
N48.pop5.p4.min38	pop5.p4.r80	10 indv/species	STACKS	48	28253	747.68	679.91	734510	28231		80%	80%
N15.pop5.p4.min15	pop5.p4.r80	3 indv/species	STACKS	15	2493	748.91	682.68	1665164	34524		80%	100%
N125.min100	NA	All	ANGSD	125	58124	NA	NA	392405	NA	NA	NA	80%

Table S3. List of *Eliurus* individuals for which the mtDNA *cytb* has been sequenced. Forest: complete forest name; Site: forest abbreviation; ID: identifier of the individual; CYTB(bp): sequence length.

Forest	Site	ID	Latitude	Longitude	CYTB (bp)	Species
Maladialina	MAL	MAL.A80	-13.75381	49.41059	944	*E. minor*
Ambalasoa	AMBALA	AMBAL.B78	-13.91646	49.40948	928	*E. minor*
Binara	BIN	BIN.K71	-13.28993	49.60661	968	*E. ellermani*
Anketrakabe	KABE	ANKET.G09	-13.45726	49.56654	1048	*E. ellermani*
Ankinjanala	ANKI	ANKI.G52	-13.34159	49.45042	991	*E. ellermani*
Salafaina	SAL	SAL.B64	-13.44856	49.72165	1064	*E. carletoni*
Analalava	ANALV	ANALV.A39	-13.61952	50.00191	1064	*E. carletoni*
Montagne d'Ambre	M_AMBRE_NORTH	M_AMBRE_NORTH.BC51	-12.51994	49.18104	1067	*Eliurus sp. nova*
Montagne d'Ambre	M_AMBRE_NORTH	M_AMBRE_NORTH.BD29	-12.5082	49.18153	966	*Eliurus sp. nova*
Montagne d'Ambre	M_AMBRE_SOUTH	M_AMBRE_SOUTH.BG77	-12.70927	49.13926	956	*Eliurus sp. nova*
Montagne d'Ambre	M_AMBRE_SOUTH	M_AMBRE_SOUTH.BG72	-12.69357	49.14442	926	*Eliurus sp. nova*
Montagne d'Ambre	M_AMBRE_SOUTH	M_AMBRE_SOUTH.BH10	-12.70601	49.13793	952	*Eliurus sp. nova*
Montagne d'Ambre	M_AMBRE_SOUTH	M_AMBRE_SOUTH.BE35	-12.67702	49.17084	921	*Eliurus sp. nova*
Montagne d'Ambre	M_AMBRE_SOUTH	M_AMBRE_SOUTH.BG61	-12.70442	49.14583	919	*Eliurus sp. nova*
Salafaina	SAL	SAL.B82	-13.44596	49.71053	1090	*E. myoxinus*
Analalava	ANALV	ANALV.A28	-13.619	49.99098	993	*E. myoxinus*
Bezavona - Ankirendrina	BEZ	BEZAV.C81	-13.56176	49.88984	1064	*E. myoxinus*
Bezavona - Ankirendrina	BEZ	BEZAV.C58	-13.56158	49.90219	1064	*E. myoxinus*
Bezavona - Ankirendrina	BEZ	BEZAV.C71	-13.54106	49.89734	1064	*E. myoxinus*
Analafiana	ANALF	ANALF.B20	-13.46023	49.82661	1060	*E. myoxinus*

Table S4. Cases of sympatric occurrence in the *Eliurus* genus. Species pairs: species in co-occurrence; Region: region of Madagascar in which co-occurrence has been documented; Site: locality of observation (forest or protected area); Reference: study where the co-occurrence has been documented or mentioned.

Species pairs	Region	Site	Reference
E. antsingy - *E. tanala*			Carleton et al., 2001
E. tsingimbato - *E. antsingy*	North-western, West	Foret d'Ambovonomby, Mahabo, Bekopaka	Jansa et al., 2019
E. tsingimbato - *E. myoxinus*	West	/	Jansa et al., 2019
E. ellermani - *E. carletoni*	North	Binara, Ambohibe	Jansa et al., 2019
E. ellermani - *E. majori*	North	/	Jansa et al., 2019
E. tanala - *E. majori*	Central Highlands	/	Jansa et al., 2019
E. tanala - *E. minor*	Central Highlands	Perinet, Ranomafana	Carleton et al., 1994
E. minor - *E. myoxinus*	North	Marojejy	Soarimalala and Goodman, 2003
E. tanala - *E. webbi*	Central Highlands	Andrambovato, Ranomafana	Carleton et al., 1994
E. minor - *E. webbi*		Maintimbato, Vondrozo	Carleton et al., 1994
Eliurus sp. nova - *E. ellermani*	North	M. d'Ambre	Present study
E. myoxinus - *E. carletoni*	North	ANDF, SOL, SAL, ANALV, ANDRA	Present study
E. myoxinus - *E. carletoni* - *E. ellermani*	North	ANTSB, BIN, BEZ	Present study
E. myoxinus - *E. ellermani*	North	ANKI, KABE	Present study
E. minor - *E. ellermani*	North	AMBA, AMBALA	Present study
E. myoxinus - *E. carletoni* - *E. minor*	North	ANDR	Present study

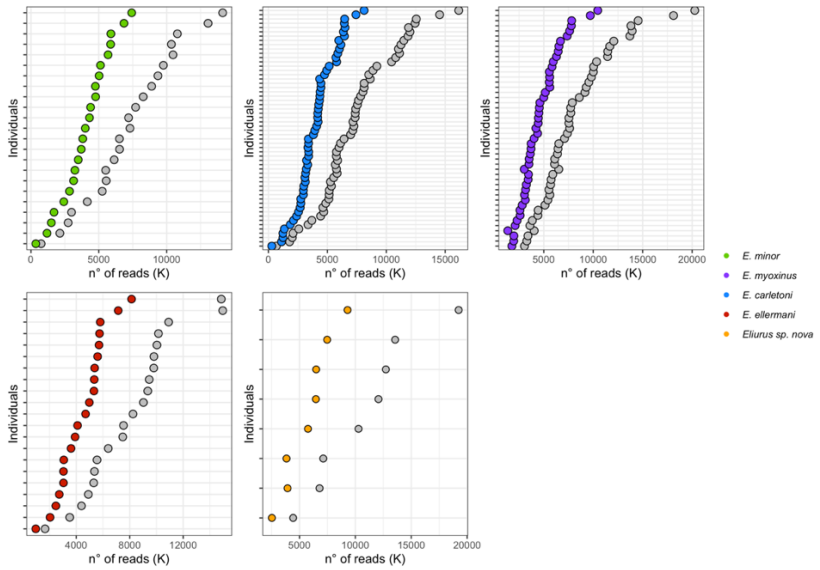


Figure S1. Number of reads before and after filtering. In the x-axis are indicated the number of reads and in the y-axis, individuals are ordered by increasing number of reads. Grey dots indicate number of reads before filtering.

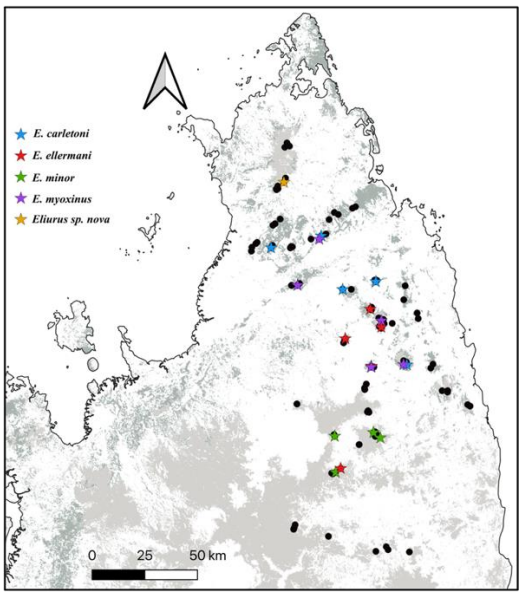


Figure S2. *Eliurus* samples selected for building the *Eliurus* spp. catalog. Black dots indicate the complete *Eliurus* dataset (N = 125), whereas coloured stars mark the individuals that have been selected for *catalog* assembly (Tab. S1). Individuals have been selected such to increase the geographic representation of each species, and use individuals with higher genomic coverage.

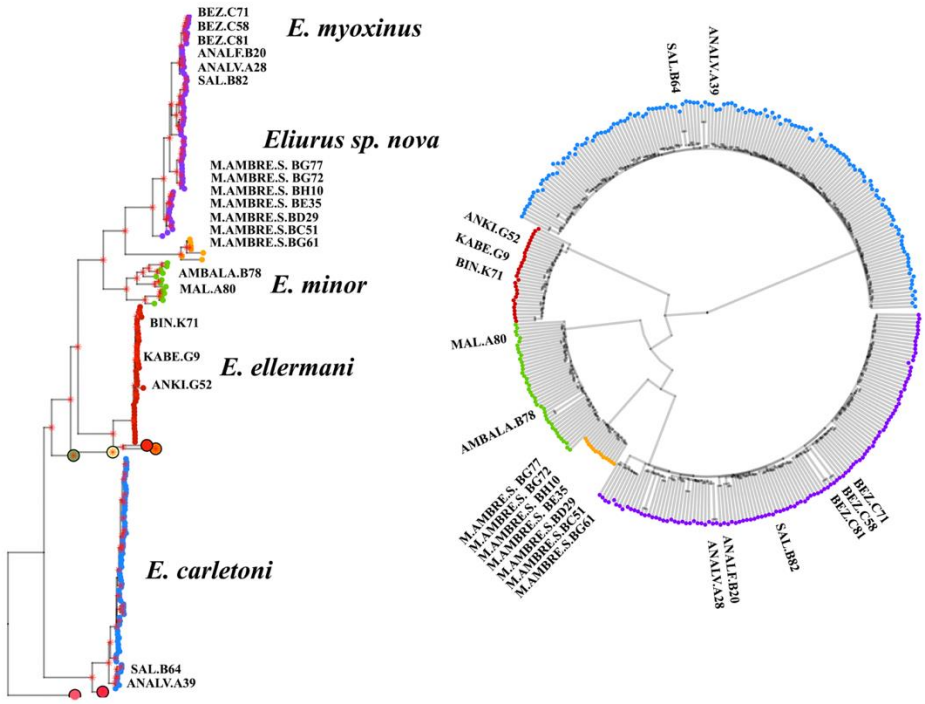


Figure S3. Consistent inference of phylogenetic relationships from mtDNA *cytb* and nuclear genomic data. Left: Bayesian mtDNA *cytb* phylogenetic tree of the *Eliurus* genus. The labelled individuals are the samples genotyped for *cytb* in the present study. Right: Neighbouring-joining tree of the RAD-seq nuclear data. This RAD-seq dataset included 347 individuals sampled across northern Madagascar, and loci that are present in at least 80% of the individuals (#SNPs: 8471). The concordant topological relationships recovered by the two phylogenetic trees suggest that our dataset is composed of five *Eliurus* taxa: *E. carletoni*; *E. myoxinus*; *E. minor*; *E. ellermani*; *Eliurus* sp. nova.



Figure S4. Maximum-likelihood and Bayesian-based phylogenetic tree estimated from mtDNA *cytb*. Left: Maximum-likelihood tree inferred in RAXML; Right: Bayesian-based tree inferred in MRBAYES. The two phylogenetic trees differ only on the topological relationships reconstructed at the deeper nodes, while confirming the phylogenetic units (taxa) identified in previous studies. MRBAYES phylogenetic tree is consistent with previous and well-established *Eliurus* phylogenies.

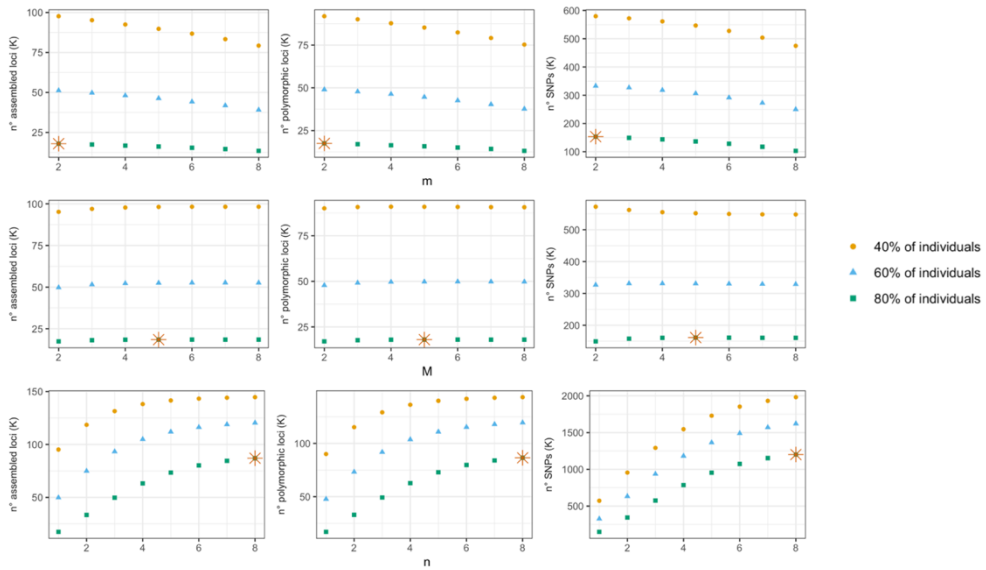


Figure S5. Optimization of *Eliurus spp. de novo* assembly of RAD-seq loci using STACKS. Iterating values of i) the minimum number of raw reads required to form a stack (*m*;

First row), ii) the distance allowed between two stacks (*M*; Second row), and iii) the number of mismatches allowed between stacks during construction of the *catalog* (*n*; Third row). a) number of assembled loci; b) number of polymorphic loci; c) number of SNPs. The arrows indicate the best parameter combination, that is *m* = 2; *M* = 5 and *n* = 8.

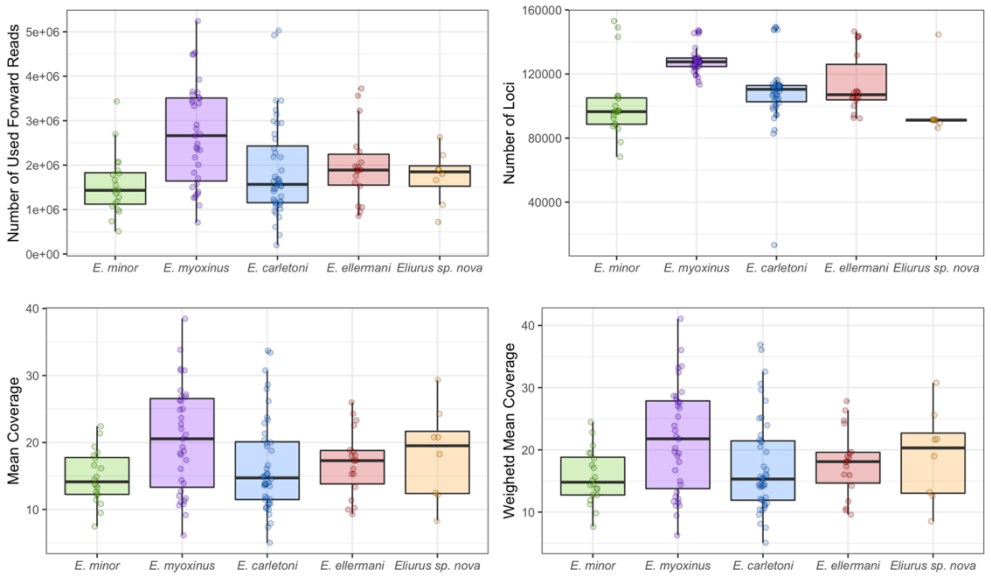


Figure S6. Summary statistics of the assembled RAD-seq loci across *Eliurus* species. Each panel shows individual variation within each *Eliurus* species in *i*) number of used forward reads; *ii*) number of assembled loci; *iii*) mean coverage; and *iv*) weighted mean coverage calculated by giving more weight to loci that are more represented across the dataset.

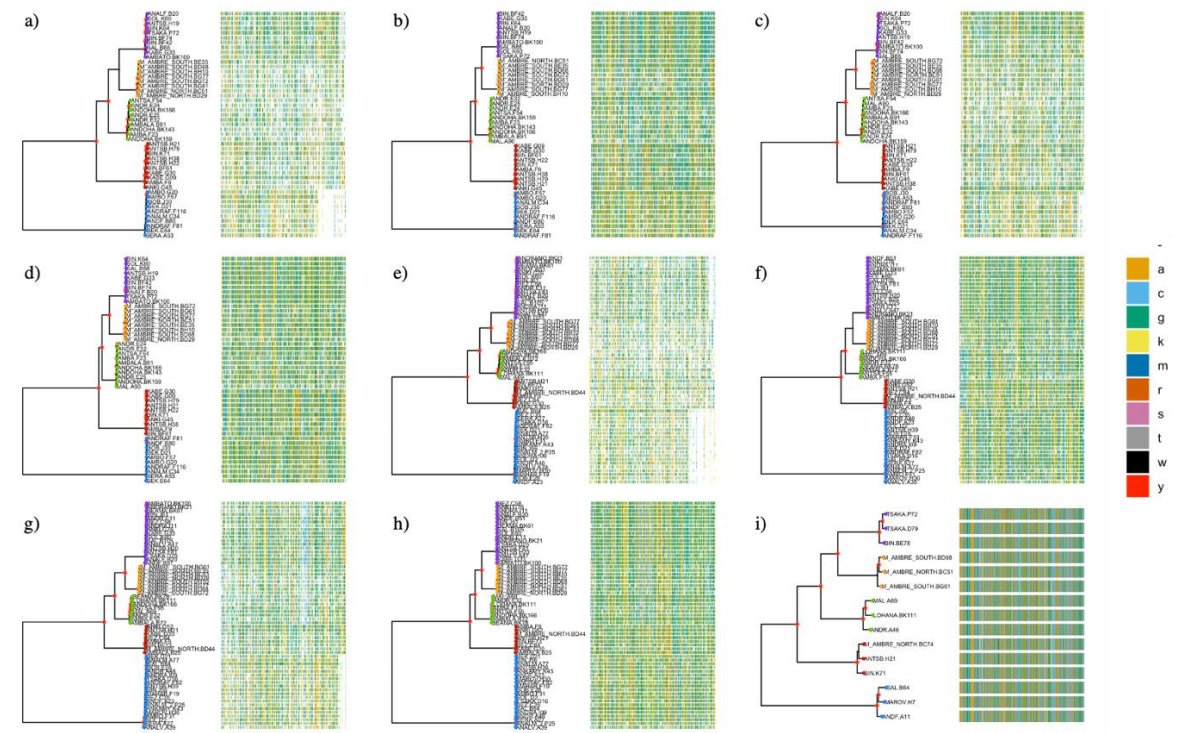


Figure S7. Maximum-likelihood phylogenies inferred from concatenated sites in each RAD-seq datasets. Phylogenies have been inferred using multialigments (on the right of each phylogeny) constructed by concatenating only sites that are fixed within individual and variants across individuals, with the exception of panel i) which includes all sites (variants and fixed). The phylogenies have been recovered using the following datasets (Tab. S2): a) N48.pop5.p0; b) N48.pop5.p0.min38; c) N48.pop5.p4; d) N48.pop5.p4.min38; e) N62.pop5.p0; f) N62.pop5.p0.min50; g) N62.pop5.p4; h) N62.pop5.p4.min50; i) N15.pop5.p4.min15. The legend indicates the color codes for each possible state in the alignment according to IUPAC nomenclature.

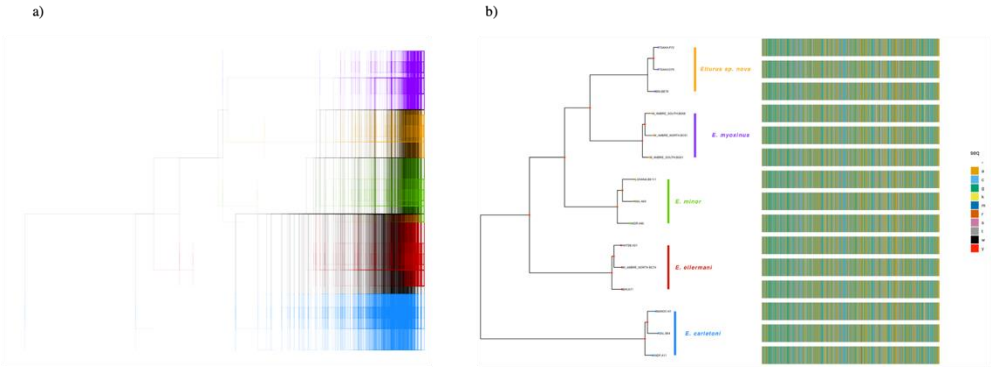


Figure S8. Partitioned phylogenomic inference. Maximum-likelihood phylogenies inferred for a) each partition (gene) independently, and b) using partitioned analysis, which estimates model parameters of the evolutionary model for each partition independently, while fixing the global set of branch lengths.

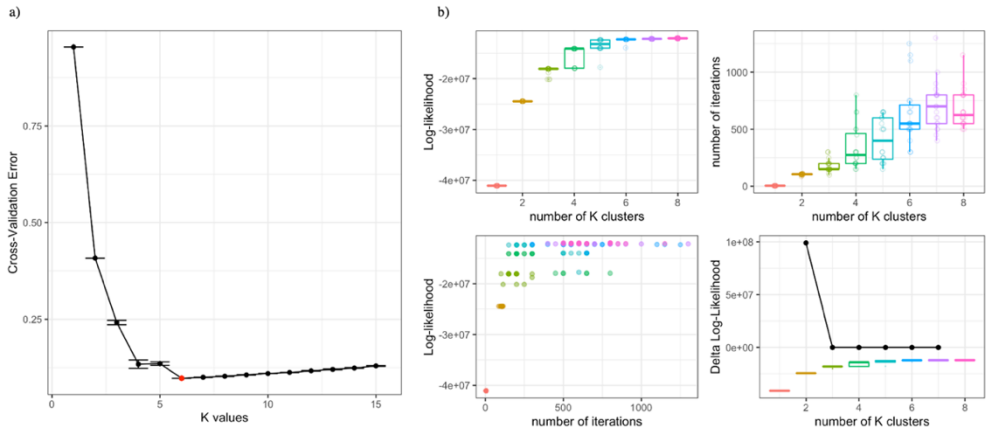


Figure S9. ADMIXTURE cross validation error and NGSadmix log-likelihood across several K-cluster values. a) ADMIXTURE model evaluation using 20-fold cross-validation (CV) error. The minimum CV error is identified for K = 6. b) Comparison of the log-likelihood of several K-values in NGS admix. The log-likelihood curve reaches plateau at K = 6. These results suggests that K = 6 is the minimum number of genetic cluster present in the dataset.

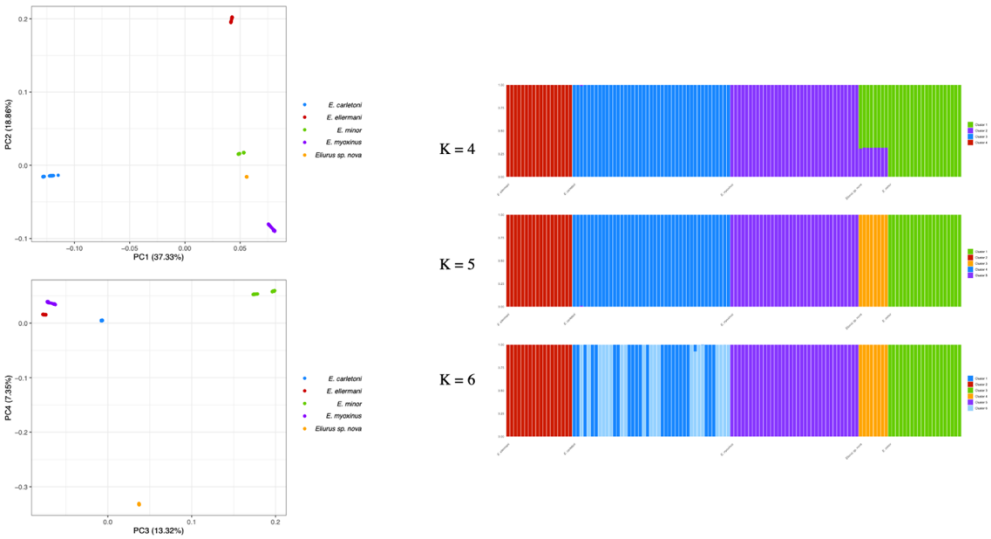


Figure S10. *Eliurus* genomic structure using PCAangsd and NGSadmix. PCA and ADMIXTURE analysis performed on genotype probabilities of the 125 *Eliurus* samples show five distinct clusters corresponding to the four previously identified species (*E. carletoni*; *E. myoxinus*; *E. ellermani*; *E. minor*) and the new taxon described in the present study (*Eliurus* sp. nova).

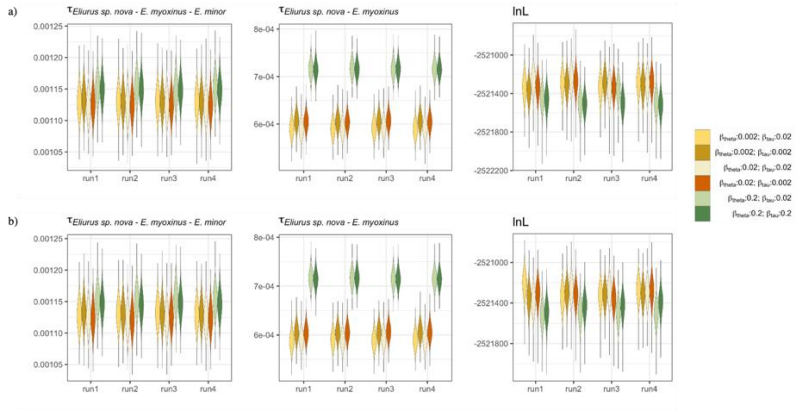


Figure S11. MCMC samples for divergence times parameters across four MCMC runs of the species delimitation analysis. Results obtained using algorithm0 (a) and algorithm1 (b) are concordant. The distribution of estimated $\tau_{E.myoxinus-Eliurus\ sp.nova-E.minor}$ values were similar across the six combinations of priors for divergence times and population size, whereas $\tau_{E.myoxinus-Eliurus\ sp.nova}$ showed significant differences for $\beta_{theta} = 0.2 - \beta_{tau} = 0.02$ and $\beta_{theta} = 0.2 - \beta_{tau} = 0.2$. Considering the large agreement across runs and algorithms, the results support the convergence of BPP analysis.

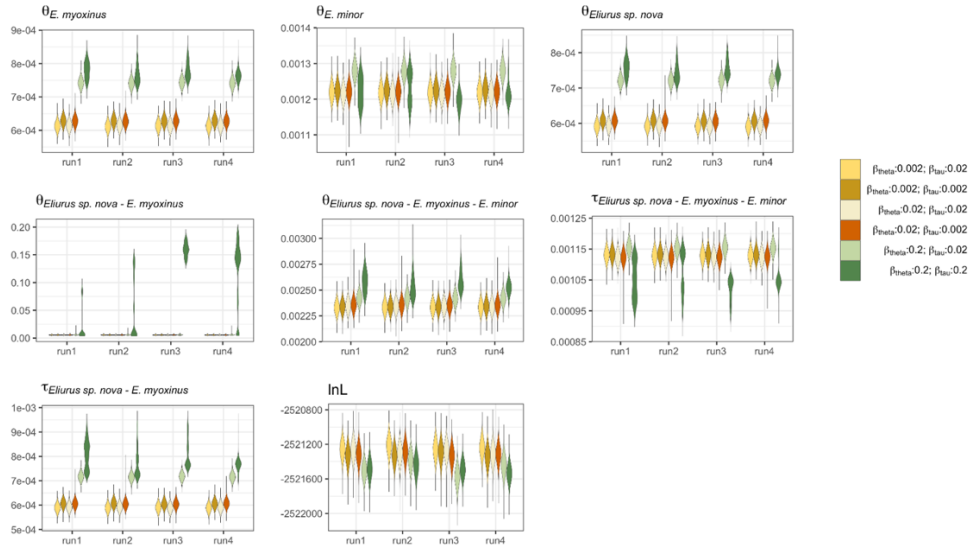


Figure S12. Estimation of the divergence times and population size parameters under the multispecies-coalescent model for the hypothesized species phylogeny ((*E. myoxinus*, *Eliurus sp. nova*), *E. minor*). The distribution of estimated values for $\theta_{E.minor}$, $\tau_{Eliurus\ sp.nova-E.myoxinus-E.minor}$, $\theta_{Eliurus\ sp.nova-E.myoxinus-E.minor}$, and log-likelihood were similar across the six combinations of priors and the four MCMC runs. $\theta_{E.myoxinus}$, $\tau_{E.myoxinus-Eliurus\ sp.nova}$, $\theta_{Eliurus\ sp.nova}$, $\theta_{Eliurus\ sp.nova-E.myoxinus}$ showed significant differences for $\beta_{theta} = 0.2 - \beta_{tau} = 0.02$ and $\beta_{theta} = 0.2 - \beta_{tau} = 0.2$. Considering the large agreement across runs and algorithms, the results support the convergence of BPP analysis.

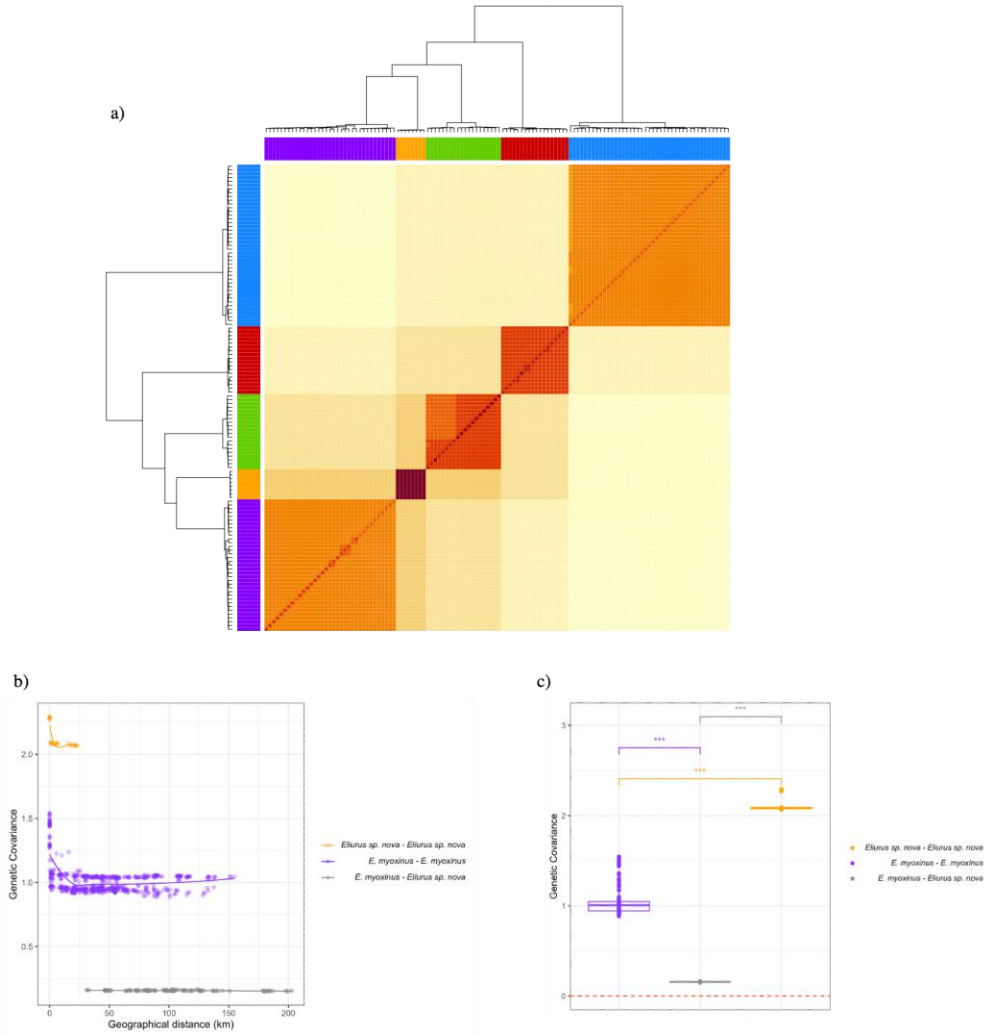


Figure S13. Genomic covariance across and within *Eliurus* species. a) Heatmap of pairwise genomic covariance across all individuals in the dataset (*N125.pop5.p4.min100*; 4.351 SNPs); b) genomic covariance within and between taxa (*E. myoxinus*, *Eliurus sp. nova*); c) Differences in the distribution of genetic covariance, within and between taxa. The covariance values were estimated from the ANGSD pipeline, and the results are consistent with the those obtained from the STACKS pipeline (Fig. 4c, d).

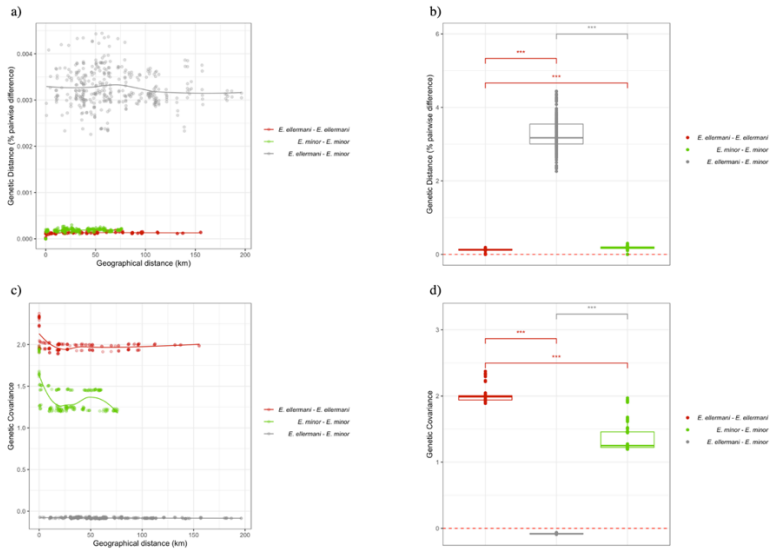


Figure S14. Pairwise genomic differences and genomic covariance within and between *E. ellermani* and *E. minor*. a) Isolation by distance within and between taxa. Genetic distance was measured as 'percentage of genetic differences on variant sites', calculated from the *N125.pop5.p4.min100* dataset (4.351 SNPs); b) Comparison of genetic distance within and between taxa; c) genomic covariance within and between taxa; d) Differences in the distribution of genetic covariance, within and between taxa.

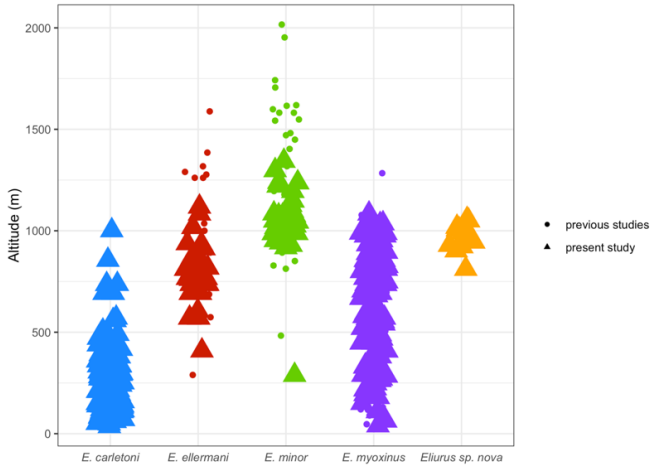


Figure S15. Altitudinal range of *Eliurus* species sampled in northern Madagascar, across their entire distribution. Significant differences in elevational range were identified for *E. carletoni*, *E. myoxinus*, *E. ellermani* and *E. minor*, whereas no significant differences were detected between *E. ellermani* – *Eliurus sp. nova* and *E. minor* - *Eliurus sp. nova* (Tukey multiple comparisons test).

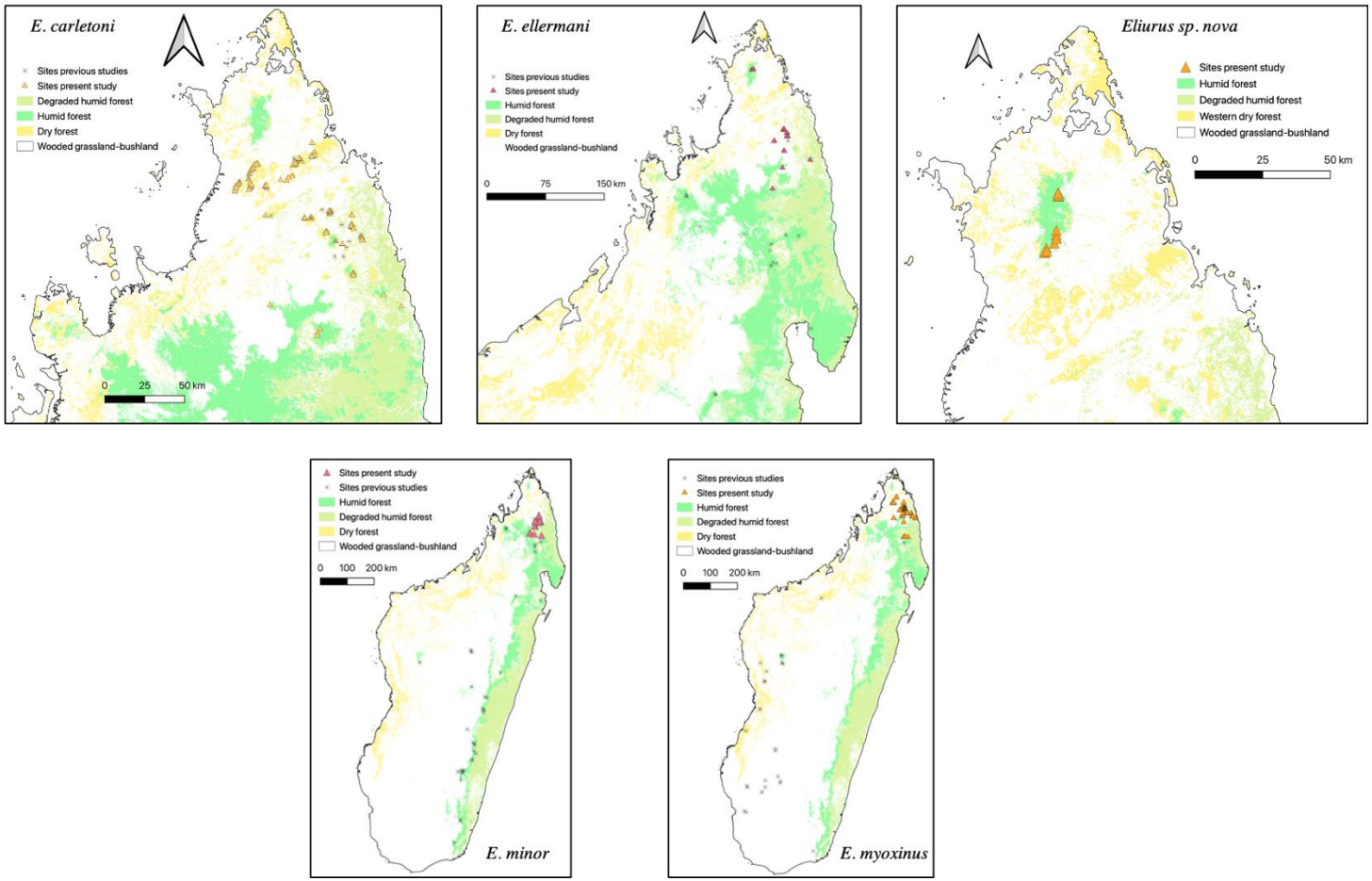


Figure S16. Range distribution of the five *Eliurus* species identified in the present study.



Figure S17. Pictures of genetically identified *Eliurus* species. *E. ellermani*: individual BEZ.C84 sampled in Bezavona – Ankirendrina (BEZ); *E. carletoni*: individual SAL.B62 sampled in Salafaina (SAL); *E. myoxinus*: individual ANDRA.I11 sampled in Andraharikely (ANDRA); *Eliurus* sp. nova: individual M_AMBRE_NORTH.BC51 sampled in St. Roussettes, Montagne d'Ambre (M_AMBRE_NORTH); *E. minor*: individual ANDR.E32 sampled in Andoharano (ANDR).

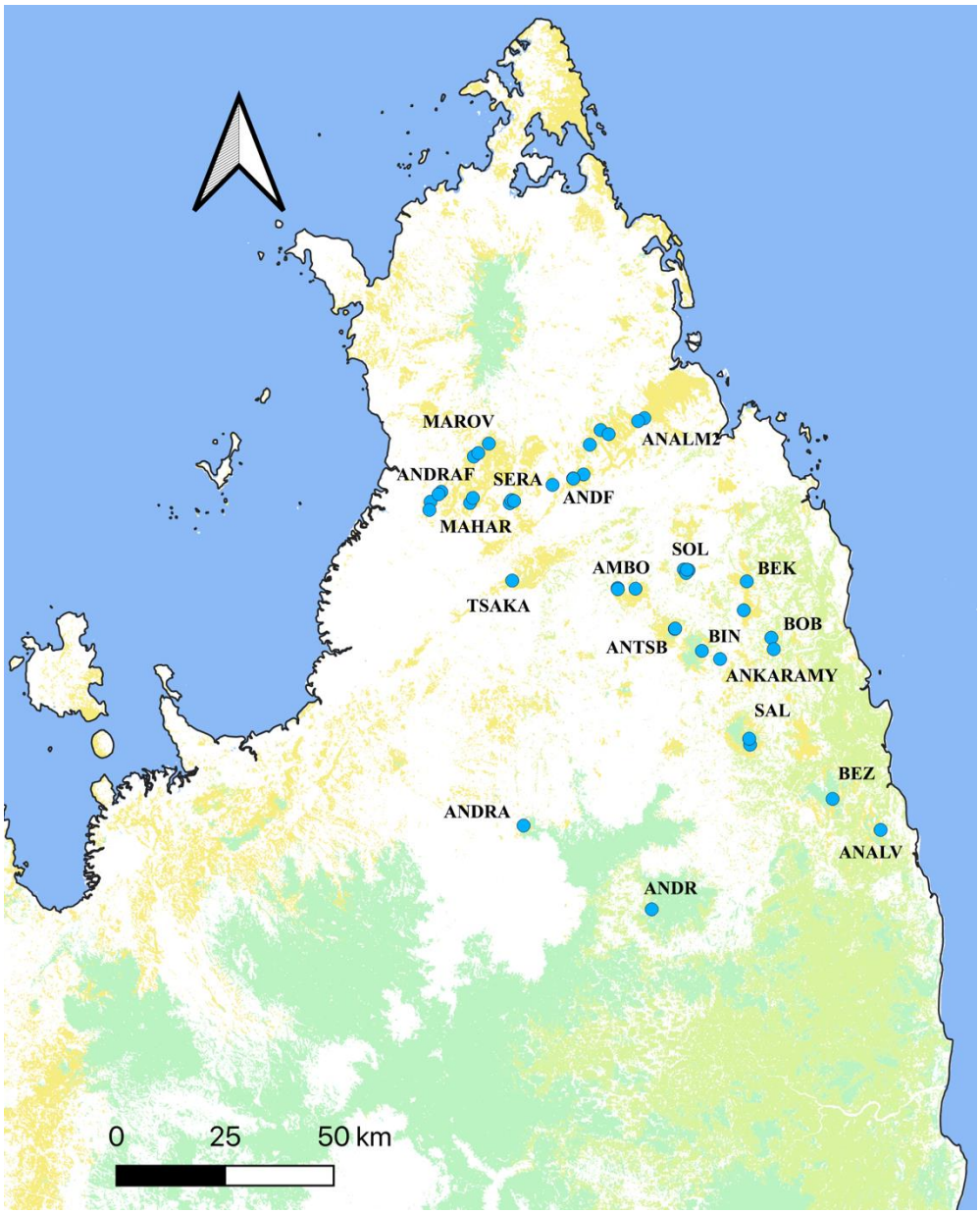


Figure S18. Sampling distribution of *Eliurus carletoni*.

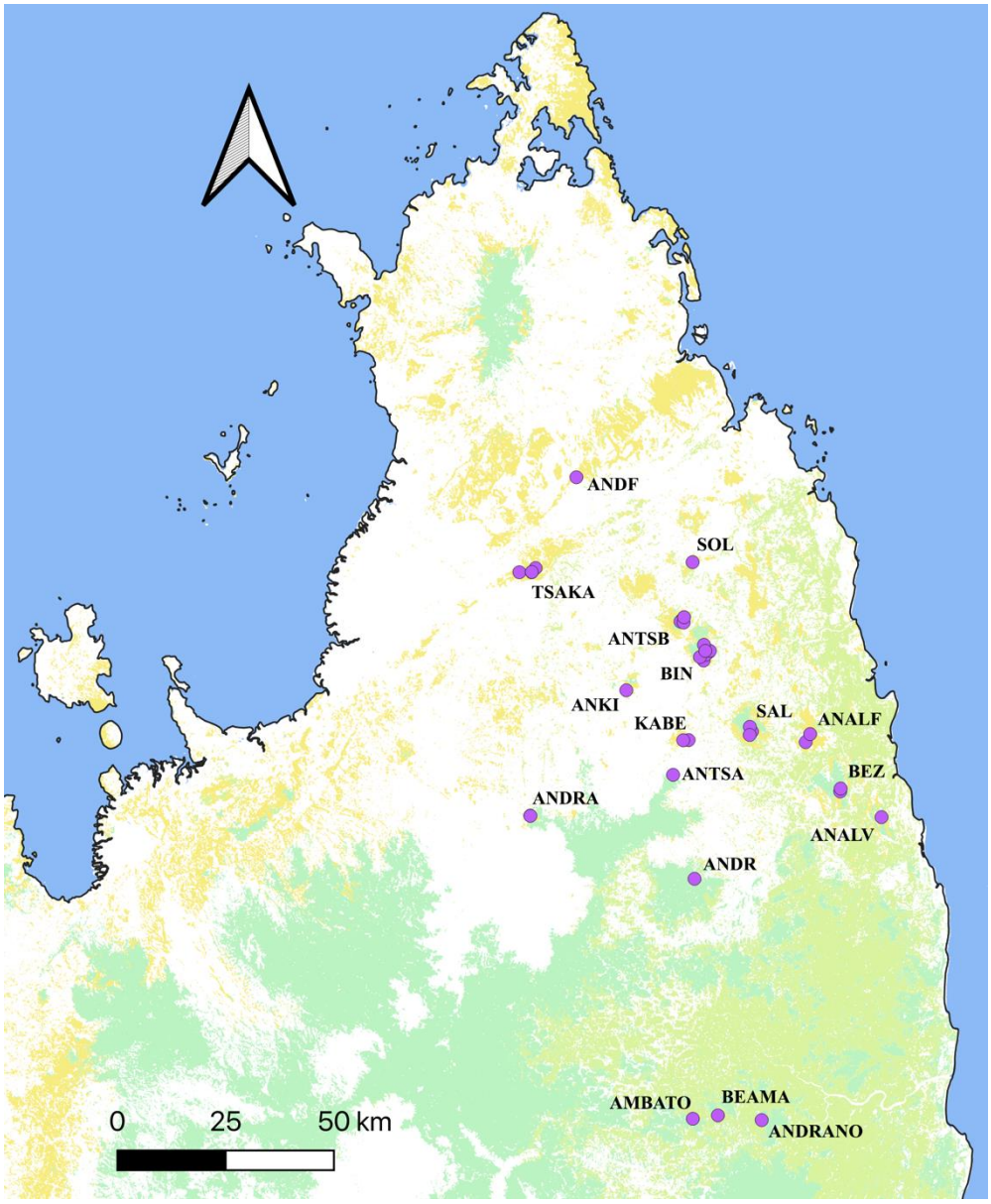


Figure S19. Sampling distribution of *Eliurus myoxinus*.

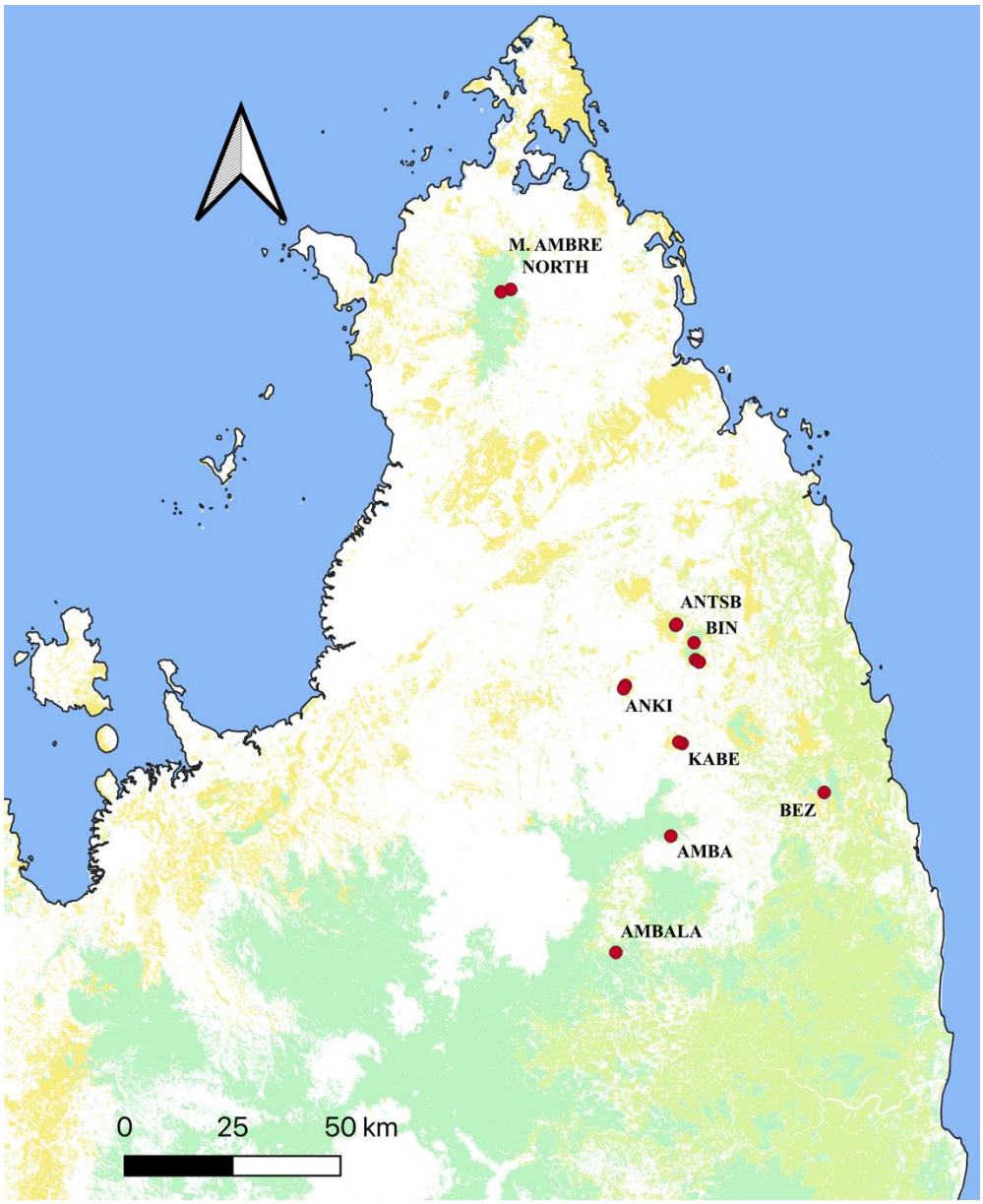


Figure S20. Sampling distribution of *Eliurus ellermani*.

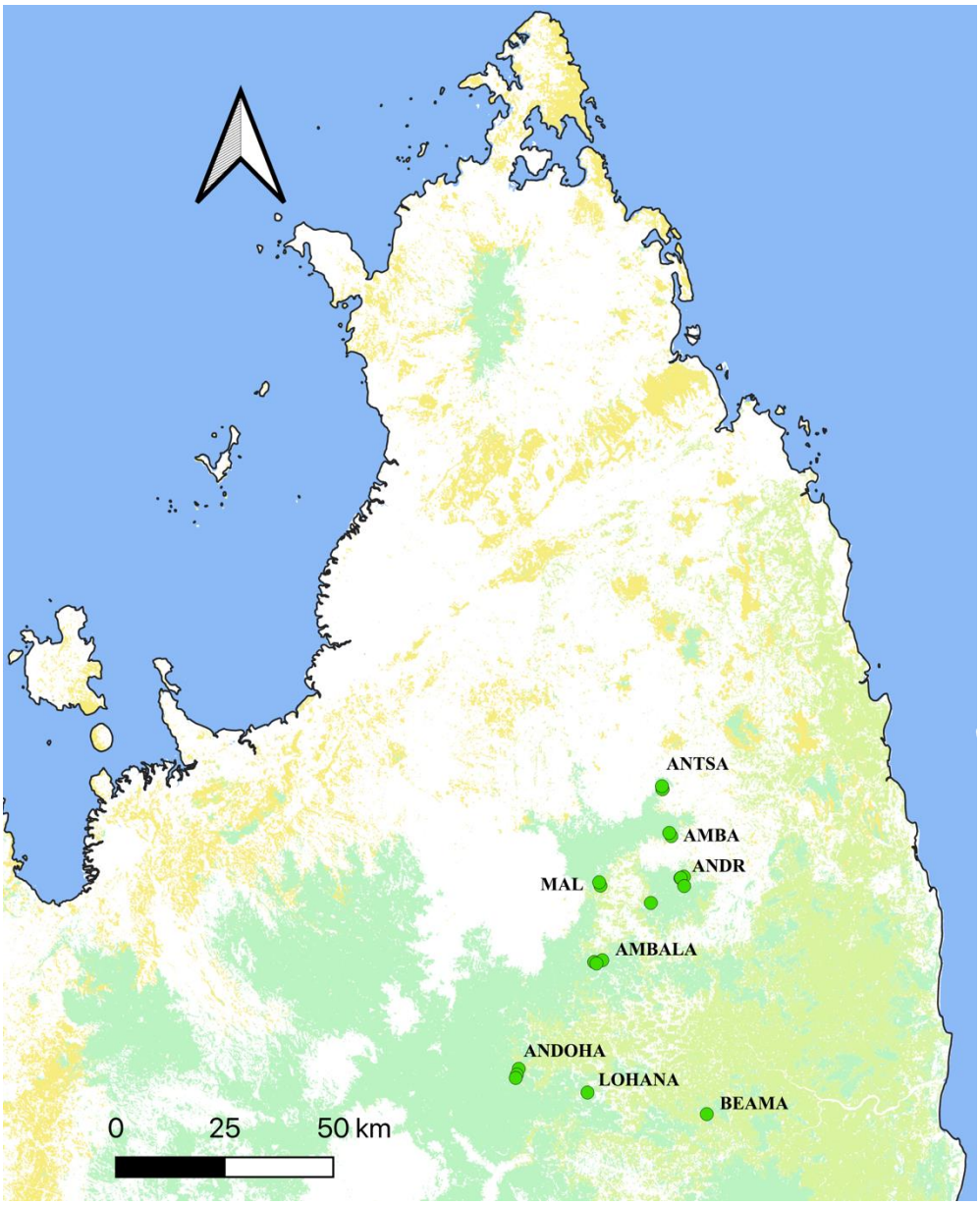


Figure S21. Sampling distribution of *Eliurus minor*.

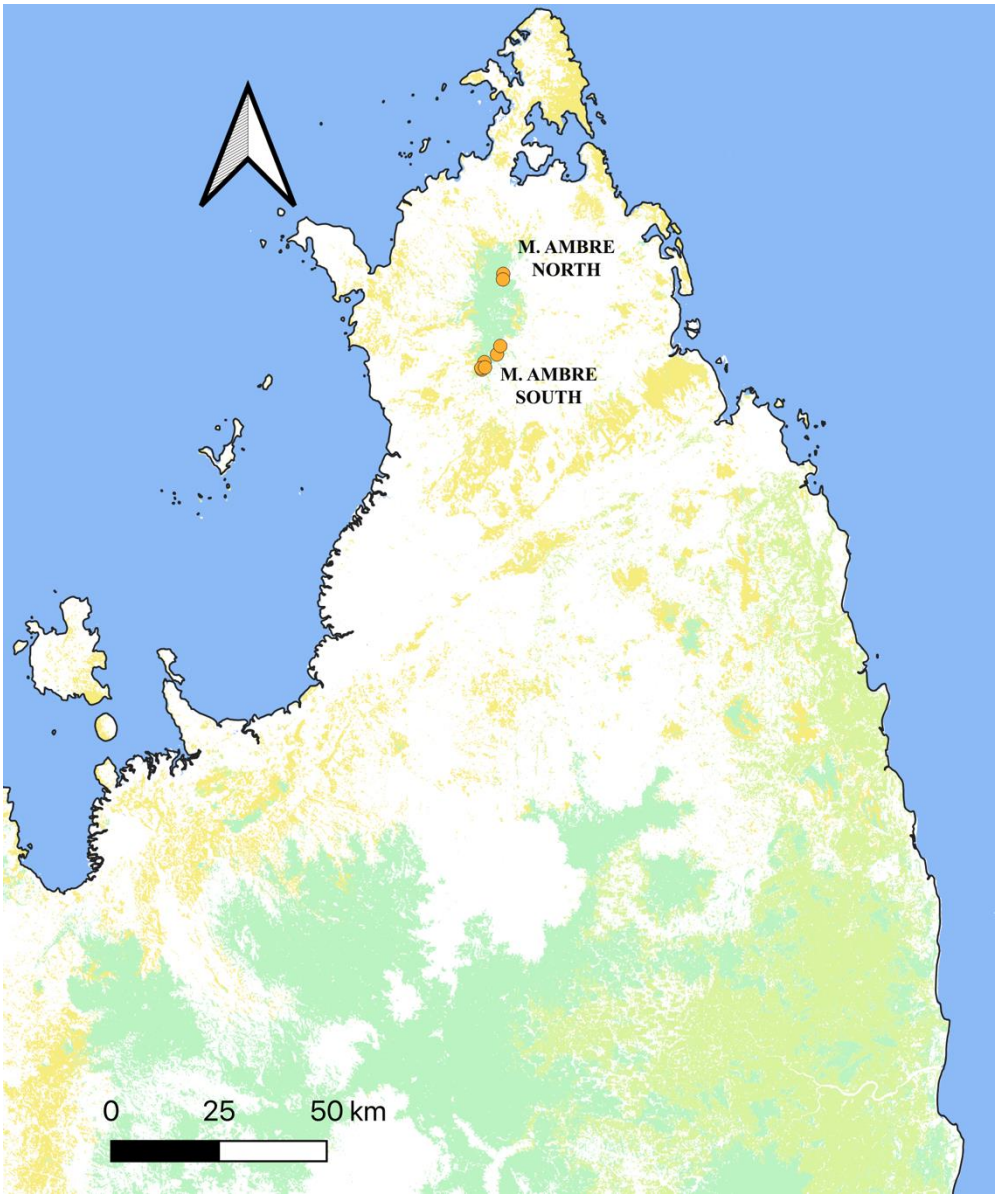


Figure S22. Sampling distribution of *Eliurus sp. nova*.

Chapter 6

6 General Discussion

6.1 Overview on scientific path: initial projects, fieldwork, conservation activities and trainings

The aim of this last chapter is to provide an overview of i) what it has been my scientific path across these years; ii) discuss the main results of my research, and iii) propose future directions for better understanding and predict the impact of habitat loss and fragmentation (HL&F) on species genetic diversity.

I started being interested on the genetic consequences of HL&F at the end of my Master degree project. At that time, the aim of my work was to identify mitochondrial DNA (mtDNA) pattern of social structure within a population of the endangered lemur species *Propithecus coronatus*, in north-west Madagascar. Later, this project evolved by involving a collaboration with my current PhD supervisor, Dr. Lounès Chikhi.

From a local-scale sociogenetic perspective, my Master project extended in spatial and temporal scales. The aim of the work was now to investigate mtDNA variability and differentiation, as well as the demographic history, of *P. coronatus* in relation to habitat loss and fragmentation. Indeed, the question of how HL&F influence the genetic diversity of species is one of the main research topics of Dr. Lounès Chikhi, and, along the path of my PhD, it became a major focus for my research.

When I started, I did not know much about population genetics. I was trained in Molecular Biology and only at the last year of my Master in Paris (France) I had the opportunity to attend few courses on population genetics.

However, at that time, I was already interested in population genetics. Inspired by a class on population genetics given by Prof. Guido Barbujani at the University of Padova, at the end of my Bachelor degree I had decided that population genetics was what I was looking for. I was fascinated by how diversity is expressed in all its forms in natural populations (including humans).

The collaboration with Dr. Lounès Chikhi and his previous PhD student, now Dr. Jordi Salmona, supported me in the publication of my Master project as a peer-reviewed paper (Sgarlata et al., 2016). This was my first published contribution to the scientific community.

Meanwhile, in 2015, I was accepted in the Integrative Biology and Biomedicine PhD program of the Instituto Gulbenkian de Ciencia, which would have started only in early January 2016. While waiting the beginning of this exciting experience, I had the chance to go to the field in the forests of northern Madagascar, together with the research team of Dr. Lounès Chikhi. There, I had my first glimpse on the exceptional biodiversity of Malagasy forests. In that period, I was also continuing my research on the genetic consequence of HL&F. This time, we were investigating the mtDNA diversity of another lemur species, the northern rufous mouse lemur (*Microcebus tavaratra*), occurring in the Loky-Manambato region, northern Madagascar. In that study (Sgarlata et al., 2018; see Annex), we showed that i) pattern of mtDNA genetic variability were consistent with evidence of recent connectivity between the two only humid forest fragments of the study area, ii) that rivers may not be strict barriers to gene flow in this species, iii) that *M. tavaratra* demographic history was rather stable over the last millennia, and iv) that in the more ancient past two waves of range expansion may have occurred across its distribution (Fig. 1).

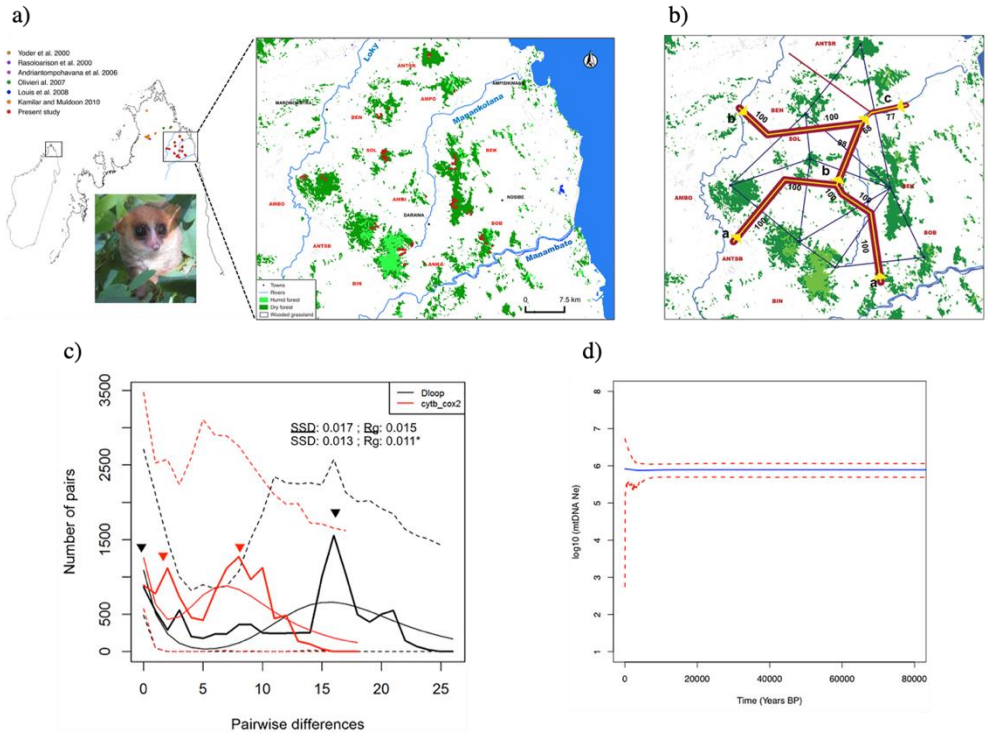


Figure 1. Genetic differentiation and demographic history of the northern rufous mouse lemur (*Microcebus tavaratra*) across a fragmented landscape in northern Madagascar. **a)** Sampling distribution in the Loky-Manambato region, Madagascar; Red dots indicate sample's locations; Dark-green: dry-deciduous forests; Light-green: humid forests. **b)** Identification of barriers to gene-flow among subpopulations of *M. tavaratra*; Red segments constitute barriers computed from 100 bootstrapped matrices (numbers); Blue lines represent the Delaunay triangulation; Yellow lines represent the genetic barriers estimated from the Φ_{st} matrix; a-first, b-second and c-third barriers. **c)** Mismatch distribution under the assumption of spatial expansion; SSD and SSD represent the sum of the squared deviation for the D-loop and cytb_cox2 dataset, respectively; Rg and Rg indicate the raggedness index of the D-loop and cytb_cox2 dataset, respectively; Continuous lines show the expected distribution; thick continuous lines show the observed distribution; dashed lines show the 95% confidence interval; triangles show the two modes identified in the distributions. **d)** Demographic history inferred using EBSP analysis and including all samples together. The solid line represents the median of the mtDNA effective population size over time. Dashed lines above and below the median values bound the 95%-HPD.

A follow up of Sgarlata et al., (2018) study aimed at assessing the extent on which mtDNA and nuclear genetic (microsatellite markers) patterns were consistent (Aleixo-Pais et al., 2019; see Annex). Our results showed that geographic distance along with open habitat, vegetation type and, to some extent, the Manankolana River, appeared to be the main factors responsible for *M. tavaratra* population structure in the Loky-Manambato region (Fig. 2).

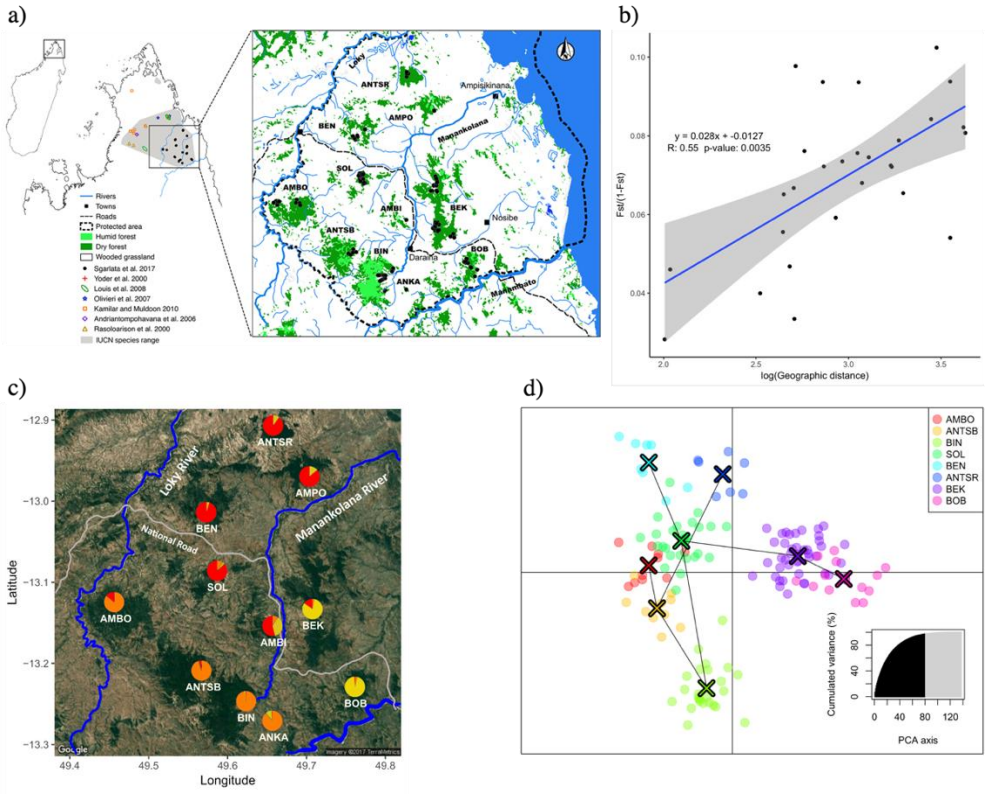


Figure 2. Genetic structure of the northern rufous mouse lemur (*Microcebus tavaratra*) in the fragmented forests of northern Madagascar. a) Geographic range of *M. tavaratra* across northern Madagascar and sampling distribution in the Loky-Manambato region (LM region). b) Isolation-by-distance pattern among *M. tavaratra* habitat fragments in LM region. c) Bayesian clustering of *M. tavaratra* individuals across the LM region for K=3 a priori ancestral populations. Color size within each pie chart is proportional to the percentage of ancestry from each of the K ancestral populations. d) Discriminant analysis of principal component analysis of sampled individuals. The colored dots represent the position of each sample on the two first

discriminant axes. The lines connecting each cross represent the minimum spanning tree based on the squared distances between populations, showing a close correspondence between genetic structure and the geographical location of the sampled forests.

In a way that was similar to a previous study on another mouse lemur species occurring in northwest Madagascar (Schneider et al., 2010), Aleixo-Pais et al., (2019) found that *M. travaratra* genetic data exhibited a significant isolation-by-distance (IBD) pattern (correlation between genetic and geographic distance), despite the high level of forest fragmentation in the study area. The detection of IBD signatures in a fragmented landscape was somewhat surprising, since one of the main assumptions behind the IBD model is that populations are connected by gene-flow (Wright, 1943; Malécot, 1948). This apparent inconsistency between spatial genetic pattern and theoretical expectations, motivated our initial question: What is the influence of HL&F on isolation-by-distance patterns? This question led us to investigate the spatial and temporal processes that influence species genetic diversity in the context of HL&F.

A related direction of investigation was also to understand the extent on which present patterns of habitat fragmentation are attributable to ancient natural environmental changes or anthropogenic activities. Genetic studies investigating the origin of HL&F in Madagascar and elsewhere often apply inferential approaches that ignore the effect of geographically-limited dispersal on species genetic diversity. This was the case of several studies in our research group (e.g., Quéméré et al., 2012; Salmona et al., 2017; Sgarlata et al., 2018; Teixeira et al., 2021). This is a general trend in population genetics, due to the complexity of spatial genetic models compared to other models of population structure (e.g., Bradburd and Ralph, 2019). This knowledge gap motivated part of my PhD research.

In the meantime, my PhD had started and fieldwork seasons in northern Madagascar as well. In the following years, I have been surveying several natural forests of northern Madagascar, for a total of 10 months. Our

work did not only consist in collecting small tissue samples from wild lemurs and endemic rodents for DNA sequencing, but also in assessing species richness and abundance within those forests. We were performing population surveys with the aim of estimating the population density of lemur species, as well as documenting opportunistic observations of present-day biodiversity: birds, reptiles, amphibians, primates and rodents. These surveys were performed together with our Malagasy collaborators, mainly Emmanuel Rasolondraibe, Ando Rakotonanahary, Jacquis Randriamaroson, Tantely Ralantoharijaona and the great support of Prof. Solofonirina Rasoloharijaona and Prof. Nicole Volasoa Andriaholinirina. As part of our collaboration, we were also training Malagasy Master students to capture wild lemurs and rodents, collect tissue and faecal samples, as well as estimate lemur population density across the surveyed forests.

The biodiversity surveys carried out over these years, and way before the start of my PhD by others, gave rise to publications aiming at i) documenting the wide and still unknown biodiversity existing in the remaining (and threatened) forests of Madagascar, and ii) understand the link between species distributions and ecological pattern. On this regard, during my PhD, we published three papers (Salmona et al., 2018; Hending et al., 2020; Le Pors et al., 2020), which I will briefly describe below.

In Salmona et al., (2018), we documented the presence of a poorly-studied lemur genus (*Phaner* spp.) in the only dry forest fragment present at the northern side of the island' humid ecoregion, the lowland dry forest of Analafiana (~ 29 km^2). The work represented one of the few studies including genetic data (mtDNA) from this genus. Genetic and morphological analyses confirmed that the individual collected in this isolated forest was closer to *P. electromontis* than to the other three *Phaner* species (*P. furcifer*, *P. pallescens*, *P. parienti*). The presence of a probable *P. electromontis* population in such unique and isolated dry forest suggested that *P. electromontis* was able to survive in likely complete isolation for relatively long time (Fig. 3).

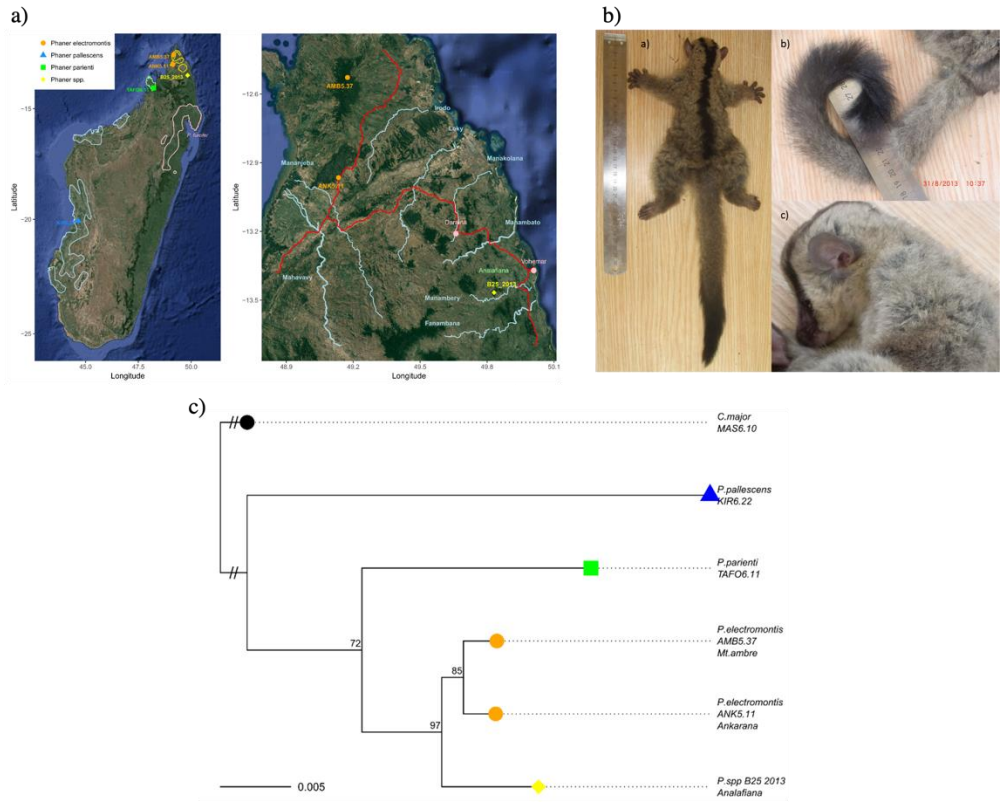


Figure 3. Re-discovering the forgotten *Phaner* population of the small and isolated Analafiana forest (Vohémar, SAVA). a) *Phaner* species distributions and locations of the samples used for phylogenetic analysis. The samples B25_2013 correspond to the only juvenile *Phaner* individual captured in Analafiana in the present study. **b)** Photographic illustration of the juvenile *Phaner* specimen B25_2013. **c)** Maximum-likelihood phylogeny of the *Phaner* genus obtained from cytb and cox2 mitochondrial loci. Bootstraps values are indicated as percentage derived from 500 replications.

Hending et al., (2020) extended Salmona et al., (2018) work by including observations and acoustic data of *P. electromontis* collected across northern Madagascar between 2011 and 2018. We estimated relative population density across the surveyed sites, compared them in relation to vegetation types and investigated the relationship between relative population density and climate variables. Although we found *P. electromontis* most

frequently in dry-transitional forests, our results suggested that geography (spatial autocorrelation) rather than environmental variables explained the species' distribution (Fig. 4).

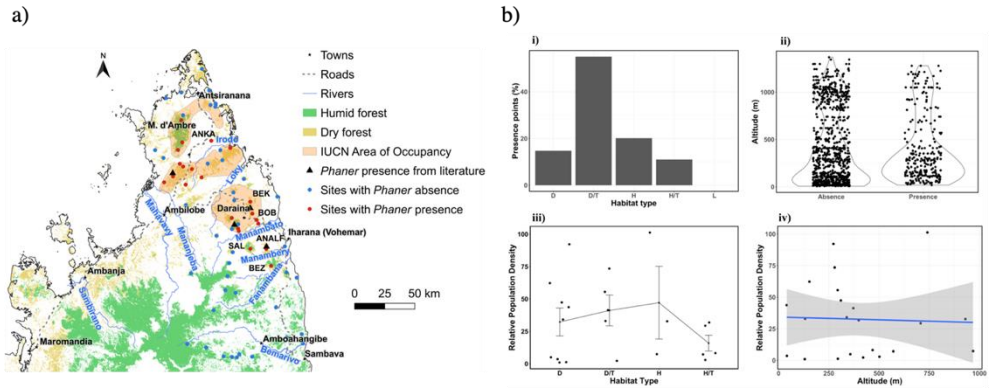


Figure 4. Distribution and conservation status of the endangered Montagne d'Ambre fork-marked lemur (*Phaner electromontis*). a) Presence and absence over the 66 study sites and four additional sites from the literature. Shaded polygons show the IUCN species distribution. The seven red dots located outside the polygons thus represent new localities where a likely *P. electromontis* species is present. b-i) Percentages of sites with *P. electromontis* presence with respect to habitat type; b-ii) elevation distribution of acoustic points with *P. electromontis* presence or absence; b-iii) relative *P. electromontis* population density with respect to habitat type; c-iv) relationship between relative population density and elevation. Forest types: D-dry; D/T-dry/transitional; H-humid; H/T-humid/transitional; L-littoral.

In Le Pors et al., (2020), we documented the observation of five lemur species in the isolated lowland Analabe dry forest (~ 10.64 km^2), in northern Madagascar. We also reported the observation of an albino lemur individual, likely belonging to *Eulemur coronatus*, which represents one of the few observations of albinism in wild primates, together with the albino gorilla identified in West Africa by Sabater Pi, (1967). Overall, our naturalistic observations suggested that small forests, such as Analabe, are still able to harbour diversity in Madagascar (Fig. 5), and as such they require to be investigated and preserved before their unique diversity disappears in the next future.



Figure 5. Lemur inventory of the Analabe forest, SAVA region, and observation of an albino *Eulemur.* Left: Surveyed forest is delimited by a red square. **Right:** A) Northern rufuous mouse lemur (*Microcebus tavaratra*); B) Aye-aye (*Daubentonia madagascariensis*) tree traces; C) northern sportive lemur (*Lepilemur septentrionalis*); D) Crowned lemur (*Eulemur coronatus*) group; E) Albino *Eulemur* spp.

In my last fieldwork season (2018), I had the chance to participate at an 'exceptional' International Union for Conservation of Nature (IUCN) red-listing workshop, organized in Antananarivo, Madagascar. 'Exceptional' because usually IUCN red-listing workshops are organized every 20 years, whereas the last IUCN red-listing workshop on lemurs occurred only five years before, in 2013. The main reason for this 'exceptional' workshop was the increasing discovery of new taxa and populations of lemurs across the island in the last years (~112 species today), as well as the high rate of HL&F that is threatening these micro-endemic species. In occasion of this workshop, Barbara Le Pors and I gathered all observations acquired across northern Madagascar by the Population and Conservation Genetics (PCG) research group and our Malagasy collaborators since 2011. In particular, we provided the first extensive assessment of population density of two mouse lemurs species, *Microcebus tavaratra* and *M. arnholdi*, and updated their geographical distribution (see Annex). In addition, we updated the distribution

of the Ankarana dwarf lemur *Cheirogaleus shethi*, Claire's mouse lemur *Microcebus mamiratra* and the Amber Mountain fork-marked lemur *P. electromontis*. At the end of this workshop, we contributed to the publication of 13 reports updating the conservation status of several lemur species: *Microcebus tavaratra* (Vulnerable, Sgarlata et al., 2020a), *M. arnholdi* (Vulnerable, Sgarlata et al., 2020c), *M. ravelobensis* (Vulnerable, Blanco et al., 2020d), *M. danfossi* (Vulnerable, Blanco et al., 2020e), *M. mittermeieri* (Endangered, Blanco et al., 2020c), *M. macarthurii* (Endangered, Blanco et al., 2020f), *M. griseorufus* (Least Concern, Ganzhorn et al., 2020), *M. simmonsi* (Endangered, Blanco et al., 2020b), *M. jollyae* (Endangered, Borgerson et al., 2020), *M. mamiratra* (Endangered, Blanco et al., 2020g), *Cheirogaleus shethi* (Endangered, Sgarlata et al., 2020b), *C. major* (Vulnerable, Blanco et al., 2020a), *P. electromontis* (Endangered, Sgarlata et al., 2020d) and *Propithecus coronatus* (Critically Endangered, Razafindramanana et al., 2020).

The fieldwork, the training of Master's students, the IUCN workshop as well as meetings with Malagasy institutions and stakeholders, were a key part of my scientific path. These achievements contributed to build my line of thoughts on what are the natural forests of Madagascar, how they are structured and to which extent my research could contribute to the conservation of the exceptional biodiversity of Malagasy forests. In particular, fieldwork was a fundamental aspect of my research, beyond the collection of tissues for genetic analysis. It helped me to look at the theoretical problems tackled during my PhD with other lenses. That is, develop hypotheses on how much some species would rely on forest cover for their persistence (habitat *versus* no-habitat), how much riparian forests along river streams could allow connectivity between apparently fragmented forests, or on how the persistence of forests would depend on climate and on the geological structure of soil (e.g., Sgarlata et al., 2019 in Chapter 5). Naturalistic observations reported in Sgarlata et al., (2019) were somehow in line with earlier hypotheses on the role played by rivers and geomorphology on the

persistence of forests during the dry periods of the late Quaternary (Wilmé et al., 2006, 2012; Mercier and Wilmé, 2013). For instance, as discussed in Chapter 5, we found that the humid-associated mouse lemur *Microcebus arnholdi* could survive in a lowland forest mainly composed of dry vegetation (Antsoroby, ~ 4 km^2) probably because of humid microclimatic conditions that permitted the persistence of a small patch of humid vegetation. Using the knowledge of local people and our observations of the soil structure in Antsoroby forest, we proposed that local geomorphology probably played a major role in generating such humid microclimate, even at a small spatial scale such as the one found in the humid patch of the dry Antsoroby forest. Although in the literature, naturalistic observations do not seem to have as much importance as other type of studies, my experience in the field made me realise that they are actually fundamental, and complementary to the theoretical work I will discuss in the following sections.

In conclusion, the research I have developed over these years on the natural history of Madagascar and on the genetic diversity of Malagasy mammals led to a final work of synthesis on the "genetics of Malagasy mammals" (Chikhi and Sgarlata, (in press); Annex). Me and my supervisor, Dr. Lounès Chikhi, gathered more than 210 articles, reviewed the use and interpretation of genetic and genomic data generated over the last decades on terrestrial Malagasy mammals (primates, rodents, carnivorans, bats, tenrec), and provided an overview of this rich and diverse literature. We summarised i) early genetic studies on taxonomy and phylogenetics, used for delimiting species and uncovering the origin of Malagasy mammals; ii) more recent population genetic studies aiming at characterising species genetic diversity and structure; iii) sociogenetic studies investigating different components of social systems (i.e., social organization, social structure and mating system); and iv) studies reconstructing the demographic history of species in relation to natural habitat changes and human activities.

6.2 Patch-scale processes in HL&F: edge-effect and habitat loss

The consequences of habitat loss and fragmentation on biodiversity has been a major research question in conservation biology over the last ~70 years (Curtis, 1956; Moore, 1962; Fahrig, 2003, 2019). Still, the importance of habitat fragmentation *per se* is an ongoing debate in the literature (Fahrig, 2017; Fletcher et al., 2018; Fahrig et al., 2019; Watling et al., 2020). Nevertheless, it is difficult to ignore that many studies have shown that habitat fragmentation *per se* processes, such as patch-size, edge-effect and patch-isolation, have an impact on ecological responses (Hanski et al., 1995; Ries et al., 2004; Prevedello and Vieira, 2010; Hanski et al., 2013; Rybicki and Hanski, 2013; Haddad et al., 2015; Pfeifer et al., 2017; Ries et al., 2017; Haddad et al., 2017; Betts et al., 2019; Bueno and Peres, 2019).

Regarding the consequences of HL&F on genetic diversity, recent meta-analyses have shown an overall negative impact of HL&F on species genetic diversity (Aguilar et al., 2008; Almeida-Rocha et al., 2020; Honnay and Jacquemyn, 2007; Lino et al., 2019; Miraldo et al., 2016; Monteiro et al., 2019; Rivera Ortíz et al., 2015; Schlaepfer et al., 2018; Vranckx et al., 2012). However, it is still unclear what is the relative effect of each HL&F process on genetic diversity. For instance, how important is patch-size compared to edge-effect, patch connectivity or *landscape-scale* habitat amount? This is perhaps linked to the fact that population genetic studies are somehow skewed towards patterns of genetic differentiation (e.g., F_{st}) at the *landscape-scale* (Cushman et al., 2012; Jackson and Fahrig, 2014; Mona et al., 2014; Jackson and Fahrig, 2016), thus leaving open questions concerning the effect of *patch-scale* processes on genetic diversity. The study of Almeida-Rocha et al., (2020) has attempted to address this knowledge gap. They compared the effect size of several HL&F processes at the *landscape*- and *patch-scale* on the genetic diversity of 69 terrestrial taxa (animals and plants), and stated that

landscape-scale processes were the factors with most detrimental effect on genetic diversity (forest cover effect size: -0.67). Nevertheless, they also showed that *patch-scale* processes have a non-negligible effect on genetic diversity (patch-size effect size: -0.42; fragments isolation effect size: -0.65). It appears therefore somehow approximative to exclude *patch-scale* processes among those processes with most detrimental effect on genetic diversity. Almeida-Rocha et al., (2020) have two important messages for the questions addressed in this PhD thesis. First, both *landscape-scale* and *patch-scale* processes impact genetic diversity probably to a similar extent. Second, a major knowledge gap exists concerning the influence of edge-effect on genetic diversity. Concerning this second observation, note that, instead, many ecological studies have revealed a pervasive detrimental effect of edges on species abundance and ecosystems biodiversity (e.g., Haddad et al., 2015; Pfeifer et al., 2017).

Theoretical population genetic studies have in several occasions anticipated that habitat edges should have a significant negative effect on genetic diversity (Maruyama, 1972; Tajima, 1990; Wilkins and Wakeley, 2002; Wilkins, 2004; Ringbauer et al., 2018). These studies, however, were considering the impact of edges in habitat of simple shape (e.g., rectangular), which is rather simplistic compared to what it is observed in real habitat fragments (Malcolm, 1994, 1998; Fletcher, 2005). In fact, these approaches can partially include the additive effects of multiple edges that arise in fragments with irregular shapes. This point was first raised by Malcolm, (1994, 2001) and later Pfeifer et al., (2017) proposed a statistical approach for quantifying the impact of habitat edges on species abundance in heterogeneous fragmented landscapes. These three studies were concerned with ecological responses to edge-effect, leaving open the question of how edge-impacted changes in genetic diversity could be reliably estimated in real complex fragments. Indeed, statistical approaches from ecology could be applied on genetic data (e.g., Keyghobadi et al., 2005), however there are differences between ecological and genetic responses that may be important

to consider. In fact, although both responses rely on similar species life-history traits such as dispersal, they also differ in some of the causal processes underlying observed patterns. For instance, early theoretical studies on spatial genetic population structure have shown that the influence of neighbouring populations on local genetic diversity (genetic correlation) depends on the relative contribution of mutation (μ) and dispersal (m) rate, that is the genetic correlation between individuals at a given distance *k* can be expressed as $r(k) \sim \exp\left(-k\sqrt{4\mu/m}\right)/\sqrt{k}$ (Kimura and Weiss, 1964). This is not the case for ecological responses, since mutation rate is typically not a relevant factor for ecological responses emerging at short-time scale, relative to the species generation time. Therefore, the approach of Pfeifer et al., (2017) may not be as accurate for genetic responses as it is for ecological responses. In Chapter 2, we aimed at filling this gap by: i) proposing an approach similar to the one developed in Pfeifer et al., (2017), but that includes the relative effect of mutation and dispersal rate on genetic response, and by ii) using spatial population genetic theory to provide a framework for predicting changes in genetic diversity due to edge-effect and habitat contraction at equilibrium or iii) non-equilibrium conditions. The objective of Chapter 2 was not to provide an inferential method, but rather show that we could theoretically quantify and predict the impact of edge-effect and habitat contraction on genetic diversity in space and time.

The study in Chapter 2 contributes also to describe the coalescence process underlying the decay in genetic diversity caused by habitat contraction. Analytical results on the decay of genetic diversity in a finite two-dimensional model date back to the work of Maruyama, (1972). In that study, Maruyama, (1972) derived the exact value of the rate of decay in genetic variability, but since that required a complicated expression (recursive approach), he also proposed an approximation describing the long-term change in genetic variability (largest eigenvalue). In a spatially-explicit simulation study investigating the genetic consequences of recent habitat

contraction, Leblois et al., (2006) noted that the Maruyama's approximation ($\sim D_e\sigma^2/N_{tot}$, if $D_e\sigma^2 \ll 1$) was valid for long-term decay in genetic variability. They also observed that decay in genetic diversity due to recent habitat contraction (e.g., 20 generations) were more in agreement with the short-term rate of decrease in local genetic diversity given by Barton and Wilson, (1995) ($\sim 1/8t\pi D_e\sigma^2$). In Chapter 2 we have used a close approximation to the coalescence probability given by Barton and Wilson, (1995), which does not require recursive calculation (as derived by Wilkins, 2004). Accordingly, the numerical framework presented in Chapter 2 could describe the decay in genetic diversity at both short- and long-term time scales rather efficiently. Note that, although I used expressions from Wilkins, (2004), our study differs in two ways from the work of Wilkins, (2004). First, I propose a statistical approach that could estimate the impact of edge-effect on local genetic diversity considering complex irregular habitat shapes, while the approach of Wilkins, (2004) is limited to simple shapes (i.e., rectangular) due to the computational burden of the method there used. Second, I was interested in describing the coalescent process in a finite two-dimensional habitat undergoing contraction, thus at non-equilibrium, whereas Wilkins, (2004) investigated only one habitat conformation (no contraction).

In both equilibrium and non-equilibrium scenarios, I modelled the genealogical process of a spatially structured population using the 'separation-of-timescales' approach, which considers that the genealogy of genes in a structured model can be divided in two phases: '*scattering phase*' and '*collecting phase*' (Wakeley, 1999; Wilkins, 2004; Barton et al., 2013). The '*scattering phase*' (shallow branches of the genealogies) would be influenced by local geographical processes, whereas the '*collecting phase*' (deep branches of the genealogies) would be no longer dependent from local processes, since it mostly includes lineages that have crossed the population multiple times. It is the '*collecting phase*', according to several studies (Wakeley, 1999; Austerlitz et al., 2000; Rousset, 2004; Barton et al., 2013), the part of the genealogical process that is mostly influenced by non-

equilibrium demographic events (e.g., habitat contraction or expansion). Moreover, Wilkins, (2004) showed that the '*collecting phase*', under certain conditions, can be modelled as the genealogical process of a panmictic population with a given effective population size (standard neutral coalescent model). Thus, in Chapter 2 I modelled the contribution of the '*collecting phase*' for a population undergoing habitat contraction using the expression for the effective population size of a panmictic population in a varying environment (Griffiths and Tavare, 1994). For both equilibrium and non-equilibrium scenarios, the results of the numerical frameworks were in close agreement with the simulated data. Therefore, I suggest that the frameworks proposed in Chapter 2 could be used for assessing the impact of habitat edges on the genetic diversity of natural populations in both equilibrium and non-equilibrium conditions.

The work carried out in Chapter 2 can be extended in various ways. Note, for instance, that the genetic response used in Chapter 2 consisted in a statistic (heterozygosity) that depends on allele frequencies, which in turn can be confounded by deeper ancestral patterns (i.e., large-scale demographic events) (Wilkins, 2004; Ringbauer et al., 2017). A possible solution to this problem is to use, as genetic response, the distribution of identity-by-descent (IBD) segments. IBD segments consist in DNA segments that are co-inherited by two individuals since their most recent common ancestor (MRCA) (Thompson, 2013). The expected number of IBD segments of a given map length L decreases with the time since the most recent common ancestor (TMRCA), because of recombination occurring during each meiotic event since TMRCA. One of the advantages of using IBD segments is that statistical inference can be constrained on shared IBD segments of large map length, which contain information on recent demography given that those segments have undergone few and recent recombination events (Barton et al., 2013). Therefore, IBD segments-based approaches may reveal patterns that with allele frequency-based statistics may be difficult to observe,

due to the influence of historical non-equilibrium demographic events on allele frequencies.

Several approaches based on IBD segments have been developed for inferring recent demographic history (e.g., Ralph and Coop, 2013; Palamara and Pe'er, 2013; Gravel et al., 2013; Ringbauer et al., 2017). For instance, Ralph and Coop, (2013) and Ringbauer et al., (2017) have shown that the expected number of shared DNA segments of map length L between two individuals at geographical distance r that coalesce at time t, given demographic parameters θ, can be expressed as:

$$E[N_L^t \mid r, \theta] = E[K_L^t]\, \psi(t, r, \theta) \qquad (Eq.\ 1)$$

where $\psi(t, r, \theta)$ is the probability that two lineages separated by a geographical distance r have a common ancestor at time t in the past and $E[K_L^t]$ is the expected number of shared DNA segments of map length L obtained by recombination events occurring over time since the common ancestor at time t. Ringbauer et al., (2017) showed that assuming a chromosome map length G much greater than the map length of the shared DNA segment L, $E[K_L^t]$ can be approximated by:

$$E[K_L^t] \approx \mathrm{G}\, 4\, t^2 e^{-2Lt} \qquad (Eq.\ 2)$$

Since in Chapter 2 I was interested in comparing the expected genetic response of individuals sampled in locations at various distances from habitat edges, we should consider the case of two individuals sampled at $r \approx 0$. Thus, we could use *Eq. 11* of Chapter 2 (derived in *Eq. A7* by Wilkins, 2004) for modelling the probability that two lineages sampled in the same subpopulation ($r \approx 0$) coalesce at time t, $\psi(t, r \approx 0, \theta)$. A statistical approach based on IBD segments could be developed under the following scenarios: i) assuming equilibrium, that is considering that the surveyed habitat fragment

has not changed in configuration for relatively long time; or ii) assuming non-equilibrium, namely that the surveyed habitat fragment has changed in configuration in the recent past.

Equilibrium: Assuming that the sampled population is at mutation-drift-dispersal equilibrium, then we can obtain the expected number of shared DNA segments of map length L between two individuals at $r \approx 0$ as:

$$E[N_L \mid r \approx 0, \theta] = \int_0^\infty E[N_L^t \mid r \approx 0, \theta]\, dt \qquad (Eq.\ 3)$$

As shown in *Eq. 11* of Chapter 2, $\psi(t, r \approx 0, \theta)$, and thus also $E[N_L \mid r \approx 0, \theta]$, depends on effective population density D_e and dispersal rate σ, that is $\theta = (\sigma, D_e)$. Therefore, we could use estimates of σ and edge-corrected population density (D_{edge}), calculated as in Chapter 2, for each subpopulation sampled at various distances from habitat edges, and compute $E[N_L \mid r \approx 0, \theta]$. Then, we could correlate the expected $E[N_L \mid r \approx 0, \theta]$ for each subpopulation with the corresponding observed number of shared DNA segments of map length L and assess the significance of such correlation. Such analysis could be performed either for a specific map length L, $E[N_L \mid r \approx 0, \theta]$, or for map lengths higher than a given threshold L_0, $E\left[N_{>L_0} \mid r \approx 0, \theta\right]$, or using the whole distributions of map length $\vec{L} = L_1, L_2, L_3, \ldots, L_n$, $E\left[N_{\vec{L}} \mid r \approx 0, \theta\right]$, where n is the number of map length bins differing by a small enough map length ΔL.

Non-Equilibrium: The sampled population may not be at mutation-drift-dispersal equilibrium, especially if the habitat contraction has occurred in the recent past. If we assume that the habitat has contracted instantaneously at time t_* in the past, we could ask when the contraction event has occurred, or more precisely when the sampled subpopulations have decreased in D_e

due to edge-effect. In this case, one could express $\psi(t, r, \theta)$ as the sum of two integrals, each referring to the probability of coalescence either before or after contraction. Before contraction, the sampled subpopulation would have a higher effective population density ($D_{edge\,2}$) compared to after contraction ($D_{edge\,1}$). Accordingly, we would need to model $E[N_L \mid r \approx 0, \theta, t_*]$ as:

$$E[N_L \mid r \approx 0, \theta, t_*] = \int_0^{t_*} E[K_L^t]\, \psi(t, r \approx 0, \theta_{edge\,1})\, dt + \int_{t_*}^{\infty} E[K_L^t]\, \psi(t, r \approx 0, \theta_{edge\,2})\, dt \qquad \textit{(Eq. 4)}$$

As in Ringbauer et al., (2017), we could use a likelihood method for inferring the time of contraction, t_*. The likelihood function $f(X|\theta)$, which is the probability of observing the data $\Pr(X|\theta)$ given the demographic parameters $\theta = (\sigma, t_*, D_{edge\,1}, D_{edge\,2})$, can be approximated by a composite likelihood $\tilde{f}(X \mid \theta)$. The composite likelihood consists in the multiplication of the component likelihoods, which in this case correspond to the multiplication of the likelihoods of the expected number of shared segments for each bin of map length L_i, with $i \in (1, n)$. Under the assumption that the number of shared segments can be modelled independently within each bin i, and assuming that L_i differ from L_{i+1} or L_{i-1} by a small enough ΔL map length, the likelihood for the expected number of shared segments of map length L_i can be modelled by a Poisson distribution with mean:

$$\lambda_i(r \approx 0, \theta, t_*) = E\left[N_{L_i} \mid r \approx 0, \theta, t_*\right]\Delta L \qquad \textit{(Eq. 5)}$$

Therefore, the composite likelihood of a pair of individuals at distance $r \approx 0$ sharing segments of length $\vec{L} = L_1, L_2, L_3, \ldots, L_n$ is given by:

$$\tilde{f}(\vec{L} \mid \theta) = Pr(\vec{L} \mid \theta) \approx \lambda_1\lambda_2\lambda_3 \ldots \lambda_n \exp\left(-\sum_{i=1}^{n} \lambda_i\right) \qquad (Eq.\ 6)$$

The maximum likelihood estimates of $\theta = (\sigma, t_*, D_{edge\,1}, D_{edge\,2})$ could then be obtained numerically using for instance the Nelder-Mead method (as in Ringbauer et al., (2017)) or an expectation-maximization (EM) algorithm. The $D_{edge\,1}$ and $D_{edge\,2}$ values inferred for each subpopulation could then be correlated with estimates of edge-corrected population density following the procedure in Chapter 2, and assessed for significance. Alternatively, we could fix the $D_{edge\,1}$, $D_{edge\,2}$ and σ parameters by using available estimates and apply the maximum-likelihood approach only for the estimation of t_*.

6.3 Landscape-scale processes in HL&F: isolation-by-distance in the context of HL&F

While in Chapter 2 I investigated the genetic consequences of HL&F at the *patch-scale* (habitat fragment), in Chapter 3 I mostly focused on *landscape-scale* genetic patterns. I used the word 'mostly' because *landscape-scale* HL&F does indeed imply processes at both *patch*- and *landscape-scale*. In fact, HL&F occurring at *landscape-scale*, not only impair movement of individuals among habitat patches and reduce the total habitat amount of a landscape, but also reduce habitat amount within patches (i.e., patch size). The reasons that drove me to explore the genetic consequences of HL&F at the *landscape-scale* could be summarised in two main points: the lack of spatially-explicit models exploring the mechanisms behind genetic response to HL&F and the observation that present-day natural populations often exhibit significant IBD patterns, while clearly rejecting the assumption in the IBD model of a continuous habitat (e.g., Schneider et al., 2010; Zhao et al., 2013; Roser et al., 2017; Aleixo-Pais et al., 2019; Pertoldi et al., 2021).

Regarding the first point, consider that our current understanding of how between-fragments genetic diversity increases after HL&F is limited to i) correlation between observed present-day genetic patterns and landscape structure (e.g., Keyghobadi et al., 2005; Vranckx et al., 2012; Schlaepfer et al., 2018; Almeida-Rocha et al., 2020), ii) theoretical studies based on non-spatial structured models (island model) (e.g., Alcala et al., 2013), or iii) to spatially-explicit simulation studies investigating stochastic landscapes varying in habitat amount and fragmentation in a relatively complex manner (e.g., Cushman et al., 2012; Jackson and Fahrig, 2016). Although we should recognize that these studies have indeed provided useful insights on the subject, they also have limitations that may hamper our understanding of the effects of HL&F on genetic diversity. For instance, meta-analyses have shown that there is a negative effect of HL&F on genetic diversity across many taxa worldwide (Vranckx et al., 2012; Rivera Ortíz et al., 2015; Miraldo et al., 2016; Schlaepfer et al., 2018; Lino et al., 2019; Almeida-Rocha et al., 2020). Such methodological approaches point to a rather general detrimental effect of HL&F on genetic diversity. However, they seldom examine the mechanisms behind the observed patterns (e.g., mutation, dispersal). On the other hand, Alcala et al., (2013) went further, and considered a finite island model for analytically describing the dynamics of genetic diversity after an abrupt change in dispersal (proxy for HL&F). Their results showed that when $2\mu \ll 1/2N$ (μ: mutation rate, N: island population size) within-island genetic diversity decreases faster than between-island genetic diversity, whereas in the case of $2\mu \gg 1/2N$ both reach equilibrium values in approximatively the same amount of time. However, the model analysed in Alcala et al., (2013) is a non-spatial structured model, whereas data across many taxa have shown that spatial structure may be the most common form of population subdivision (Meirmans, 2012; Sexton et al., 2014). Accordingly, spatially-explicit simulations were carried out by Jackson and Fahrig, (2016) to examine how habitat area or fragmentation *per se* influences spatial genetic structure. Their results suggest that habitat area has a higher effect on spatial genetic

structure than habitat fragmentation *per se*. Note however that Jackson and Fahrig, (2016) investigated spatial genetic patterns from complex stochastic landscape configurations, while sampling genetic data only at the end of the simulations (after 1000 generations). Thus, from Jackson and Fahrig, (2016) study, it still remains unclear i) whether the simulated populations had reached quasi-equilibrium, ii) what is the temporal dynamics of genetic diversity across these landscapes, and iii) whether *patch-scale* processes, such as edge-effect, could have had a not negligible influence on simulated genetic diversity, given the complex geometries of the modelled landscapes.

The second aspect that motivated our research on *landscape-scale* genetic response to HL&F concerns the observation that present-day fragmented populations typically exhibit significant IBD pattern. This is somehow surprising, given that IBD models rely on the assumption that habitat is continuous, such that an equilibrium between dispersal, mutation and drift is reached over the whole habitat (Wright, 1943; Malécot, 1948). In fact, in the absence of gene-flow among habitat fragments, it is qualitatively expected that genetic drift would be much more influential than historical gene-flow, resulting in no correlation between geographical and genetic distance (Case III in Hutchison and Templeton, 1999). A clear interpretation of such pattern, however, should go beyond qualitative expectations and would require a characterisation of the temporal dynamics of between and within-fragment genetic diversity in a spatial context.

In Chapter 3, we used a 'hybrid' approach to fill these knowledge gaps. We carried out spatially-explicit simulations to explore non-equilibrium isolation-by-distance (IBD) in landscapes undergoing HL&F and introduced a theoretical framework based on classical results in a torus-like space (Slatkin, 1991, 1993) for analysing IBD in the context of HL&F. Unlike Jackson and Fahrig, (2016), we decided to simulate a simple HL&F configuration where a homogenous squared grid would be subjected to HL&F generating a fragmented landscape composed of nine patches with equal geometry (square), such that edge-effect acts similarly within each fragment. We

explored three scenarios of HL&F, aiming at reproducing alternative models that are likely to occur in natural populations: i) instantaneous HL&F, ii) gradual (two-steps) HL&F, and iii) instantaneous HL&F following range expansion (Fig. 6). It is worth to note that scenarios i) and ii) were at mutation–dispersal–drift equilibrium prior to HL&F, whereas scenario iii) considers a non-equilibrium initial condition generated by a recent range expansion. Overall, our simulation results showed that IBD is maintained for thousands of generations after HL&F, depending to a large extent on dispersal rate.

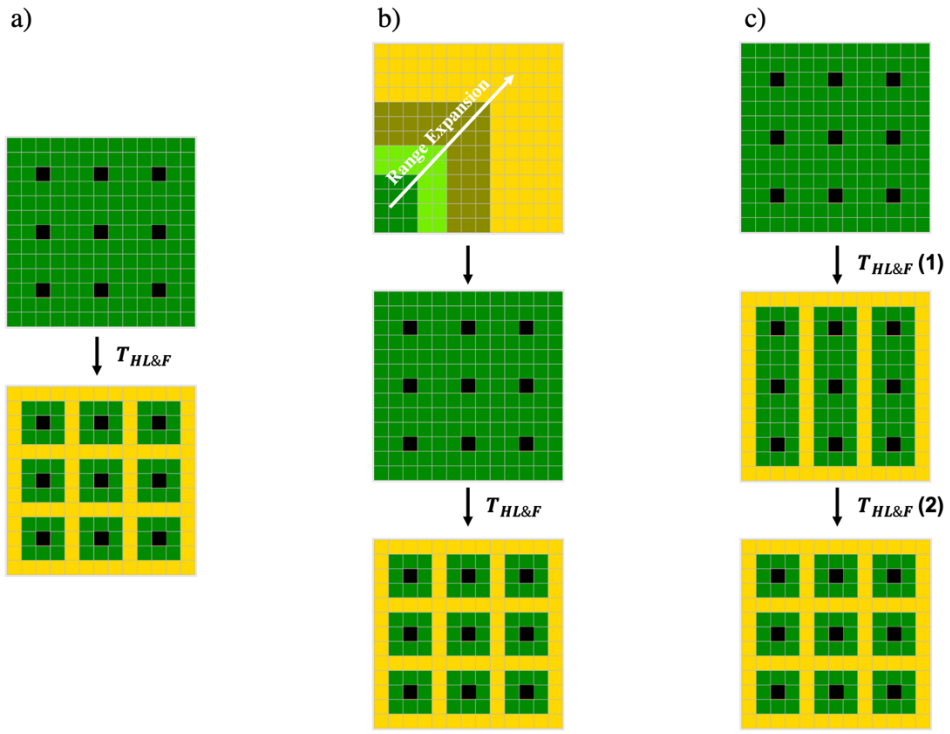


Figure 6. Simulated scenarios of *landscape-scale* HL&F. a) Instantaneous HL&F; **b)** instantaneous HL&F following range expansion; **c)** gradual (two-steps) HL&F. Green: suitable habitat. Yellow: unsuitable habitat. Black square: sampled demes.

The 'Two-steps HL&F' scenario consisted in a first HL&F event, forming three mid-size forest fragments, which after a time interval $T_{HL\&F}(2)$

undergo a second HL&F event, leading to the nine-fragment landscape configuration of the 'instantaneous HL&F' scenario. The results of the 'Two-steps HL&F' scenario showed that longer is $T_{HL\&F}(2)$, larger is the time at which IBD will be lost after HL&F. Since natural and anthropogenic HL&F are likely to occur as a spatially-correlated gradual process (Mandelbrot, 1982; Krummel et al., 1987; Milne, 1988; Solé and Manrubia, 1995), our results imply that fine-scale HL&F between neighbouring forests needs to be considered for accurately interpreting observed genetic differentiation in fragmented landscapes. To put it in another way, it is more likely that HL&F occurred in the more recent past for neighbouring fragmented habitats than for fragments far apart. Consequently, populations occurring in neighbouring fragments should exhibit smaller genetic differentiation than fragments far apart, to an extent that is larger than differences expected in a continuous spatial model.

Range expansion is another evolutionary process that have shaped the genetic diversity of many taxa. A well-known example is the post-glacial colonisation of continental Europe after the Last Glacial Period (LGM), about 20 thousand years ago (Hewitt, 1999, 2000). During the cold LGM period (33K-20K years ago) (Clark et al., 2009), the growth of ice sheets constrained species into habitat patches (refugia) in continental Europe: mainly Iberia, Italy and the Balkans (Hewitt, 2000, 1999; Lattin, 1967; Taberlet et al., 1998; Juřičková et al., 2014). The end of the LGM was marked by increase in temperature, the retreat of ice sheets and the expansion of species from the southern refugia towards the north. Evidence for the post-glacial species expansion come from fossil data as well as genetic data (Hewitt, 2000, 1999; Lattin, 1967; Taberlet et al., 1998; Juřičková et al., 2014). The fact that present-day populations still hold traces of range expansions starting less than 20K years ago in their genes, raises the question on how ancient and recent demographic events jointly influence the genealogical process of sampled genes, or as others have put it "Is it ancient or modern history that we can read in our genes?" (Nichols, 1996; Hewitt, 1999). Following this

rationale, we thus investigated a scenario where a species colonises a continuous landscape, which is then subjected to instantaneous HL&F (nine-patches configuration). In agreement with previous theoretical studies (Meirmans, 2012; Mona et al., 2014), our results showed that IBD is maintained for longer in populations that have undergone a range expansion than in populations that are at mutation-drift-dispersal equilibrium prior to HL&F. The influence of a recent range expansion on the IBD of fragmented populations decreases with the age of the expansion. This is explained by the fact that older is the range expansion, more probable is that a given population had time to reach mutation-drift-dispersal equilibrium prior to HL&F.

Altogether, the 'Two-steps HL&F' and 'Range expansion + instantaneous HL&F' scenarios point out that the IBD patterns estimated from natural populations are likely influenced by several processes. Therefore, we argue that if our objective is to make use of spatial genetic pattern for estimating the time of HL&F, the best way to proceed might be to consider only neighbouring habitat fragments. We justify this conclusion using two arguments.

The first argument considers that HL&F is typically a spatially-correlated process, which implies that geographically close fragments have lower genetic differentiation than expected under a scenario of instantaneous HL&F (see Fig. 7 and Chapter 3). In fact, in the 'instantaneous HL&F', pairwise genetic differentiation between neighbouring fragments increases at the same rate as pairs of fragments that are far apart. Instead, we have shown that for the 'Two-steps HL&F', genetic differentiation between fragments far apart increases much faster than neighbouring fragments.

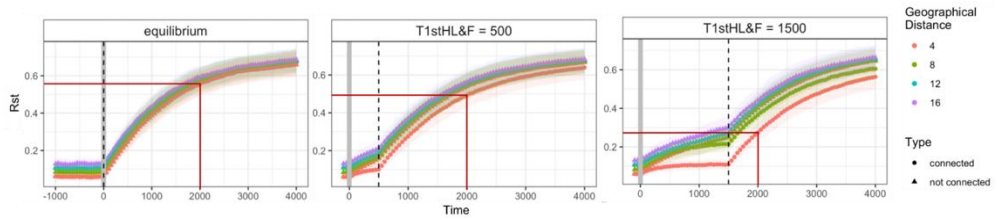


Figure 7. Pairwise genetic differentiation (R_{st}) during 'Instantaneous HL&F' and 'Two steps HL&F'. equilibrium: 'Instantaneous HL&F'. T1stHL&F: time of the first HL&F event (see Fig. 6). Thick grey vertical line: first HL&F event. Dashed black vertical line: second HL&F event. Red line indicates R_{st} value at 2000 generations after the first HL&F event across all scenarios.

The second argument is based on the prediction that the genealogy of many natural populations is likely influenced by past non-equilibrium demographic events (e.g., range expansion) (Wilkins, 2004). This means that many wild fragmented populations were probably far from mutation-dispersal-drift equilibrium at the moment of HL&F, as it can be, for instance, a species that has undergone recent range expansion, where genetic diversity typically decreases with increasing distance from the origin of expansion (e.g., Austerlitz et al., 1997; Edmonds et al., 2004; Wegmann et al., 2006; Slatkin and Excoffier, 2012).

Considering these two arguments, a natural question to ask would therefore be: how can we overcome the 'complication' of a non-equilibrium initial condition prior to HL&F? As proposed by Wilkins, (2004), one possible approach would be to model the non-equilibrium demography only in the deepest branches of the genealogy (*collecting phase*), since the shorter branches of the genealogy (*scattering phase*) are mostly influenced by local evolutionary processes, such as population density and dispersal rate. An alternative approach, when analysing real data, could be to remove those genetic markers showing large-scale geographic pattern in allele frequency that are attributable to non-equilibrium processes (as suggested in Ringbauer et al., 2018). However, we should note that, under certain conditions,

ancestral non-equilibrium processes may also be ignored. In fact, Slatkin and Barton, (1989) and Barton et al., (2013) have shown that at intermediate spatial scale (e.g., neighbouring populations), 'quasi-equilibrium' can be reached over relatively short temporal scales, depending on mutation (μ) and dispersal (σ) rate, as given by $Diam(A)^2/\sigma^2 \ll t^* \ll \mu^{-1}$ where A is the sampling area and t^* is the time period within which the allelic distribution in A has remained approximately constant.

Nevertheless, even if all necessary precautions are taken in order to get rid of confounding signatures in the deeper branches of the genealogy, still we do not have the proper theoretical framework to correctly interpret genetic patterns emerging in landscapes that have undergone HL&F. In fact, one would like to have a framework that jointly implement the effect of geographically-limited dispersal and time of divergence (equivalent to time of HL&F) on expected genetic pattern. However, models of population divergence are currently based on non-spatial population structure (Nielsen and Wakeley, 2001; Hey and Nielsen, 2004; Gutenkunst et al., 2009; Costa and Wilkinson-Herbots, 2017; but see Rodríguez et al., 2018). In Chapter 3, we proposed a theoretical framework that could be interpreted as a model of population divergence in space, more precisely a 'strict isolation in space' model. Inspired by the work of Duforet-Frebourg and Slatkin, (2016), the model of 'strict isolation in space' that we propose is based on early classical results in a toroidal stepping-stone model (Slatkin, 1991, 1993). Briefly, we assumed that an ancestral population occupying a torus-like space is at equilibrium and undergoes instantaneous HL&F at a given time t_*, leading to the breaking apart of the torus-like space into several habitat fragments of equal size, also of torus-like space. Since HL&F, lineages have moved within each habitat fragment. At the time of genetic sampling, which we assumed to be at time $t = 0$, we do not know where the ancestors of the sampled genes were at t_*. We find a solution to this problem by using a discrete random-walk, backwards in time, which traces back the location of the ancestral lineages at the time of HL&F. The model provides quantity that are of interest in

population genetics, that is the expected number of pairwise differences between genes at a given geographical distance r, collected in two habitat fragments A and B, given demographic parameters θ:

$$\mathrm{E}(H_r|\, r, \theta) = 2\mu(t_{00} + \bar{S}_r + t_*) \qquad (Eq.\ 7)$$

and the corresponding expected pairwise genetic differentiation,

$$F_{st}(r, \theta) = \frac{t_{00} + \bar{S}_r + t_* - \tilde{t}_{00}(t_*)}{t_{00} + \bar{S}_r + t_*} \qquad (Eq.\ 8)$$

where t_{00} is the expected coalescence time, before HL&F, for two alleles sampled in the same deme, $\bar{S}_r$ is the average amount of time required for two alleles to be in the same subpopulation given that they are at r distance apart and considering all possible paths travelled by the ancestral lineages of each allele within habitat fragments, t_* is the time since HL&F event and $\tilde{t}_{00}(t_*)$ is the expected coalescence time for two alleles sampled in the same deme, within each habitat fragment, given that habitat fragments have contracted since t_*.

To summarize our findings, analyses of the proposed model showed that geography does matter for thousands of generations after HL&F, depending on several factors such as the dimension of the torus-like space prior to HL&F, dispersal rate, deme size and age of HL&F. The relative impact of geography on measures of pairwise F_{st} relies additionally on the extent of habitat contraction, which determines the rate and intensity of decay in genetic diversity within habitat fragments, $\tilde{t}_{00}(t_*)$. Our results have also shown that, if the configuration of the fragment is symmetric with respect to the sampling site (i.e., center of the fragment), pairwise genetic differences between isolated habitat fragments increases linearly with time since HL&F, simplifying some of our calculations.

Regarding the controversy on the relative role of habitat amount or fragmentation *per se* in determining genetic responses to HL&F (Cushman et al., 2012; Jackson and Fahrig, 2016), the theoretical framework presented in Chapter 3 can perhaps help to better understand the consequences of HL&F on genetic diversity. The simulation study of Jackson and Fahrig, (2016) concluded that habitat amount may often be a better predictor of spatial genetic structure than habitat configuration (fragmentation *per se*). The model proposed in Chapter 3 supports the conclusions of Jackson and Fahrig, (2016). In fact, from *Eq. 8*, we can observe that the numerator of $F_{st}(r,\theta)$ increases with time since HL&F due to i) increase in t_*, and ii) increase in $\Delta t_{00} = t_{00} - \tilde{t}_{00}(t_*)$, given that $\tilde{t}_{00}(t_*)$ decreases with t_*. Since $t_{00} - \tilde{t}_{00}(t_*)$ is likely much greater than t_* in real populations, from *Eq. 8* we can conclude that increase in $F_{st}(r,\theta)$ after HL&F is mostly related to changes in habitat amount $\tilde{t}_{00}(t_*)$. However, it is also theoretically justified to state that the total amount of habitat in a landscape, as considered in Jackson and Fahrig, (2016), is not always the best predictor of spatial genetic structure.

To better understand such statement, let's consider two extreme cases of HL&F (Fig. 8). One implies decrease of habitat amount by means of habitat contraction, namely actual loss of habitat. The second considers that, even if a landscape is fragmented by infinitesimally thin barriers, such that a not significant amount of habitat is lost, the population size of the species occurring within the habitat fragment would be effectively reduced. For better picturing the second scenario, we could imagine a relatively realistic scenario where a large landscape gets fragmented by a 'grid' of roads. From a human and landscape-scale perspective, we would be inclined to assume that the total habitat amount across the landscape has not changed after HL&F. However, from the perspective of a species, which is strictly associated to its habitat, this may not be true. In fact, after HL&F, movement and reproduction of individuals will be confined within the habitat boundary of the fragment, which results into an effective decrease in within-fragment genetic diversity, $\tilde{t}_{00}(t_*)$.

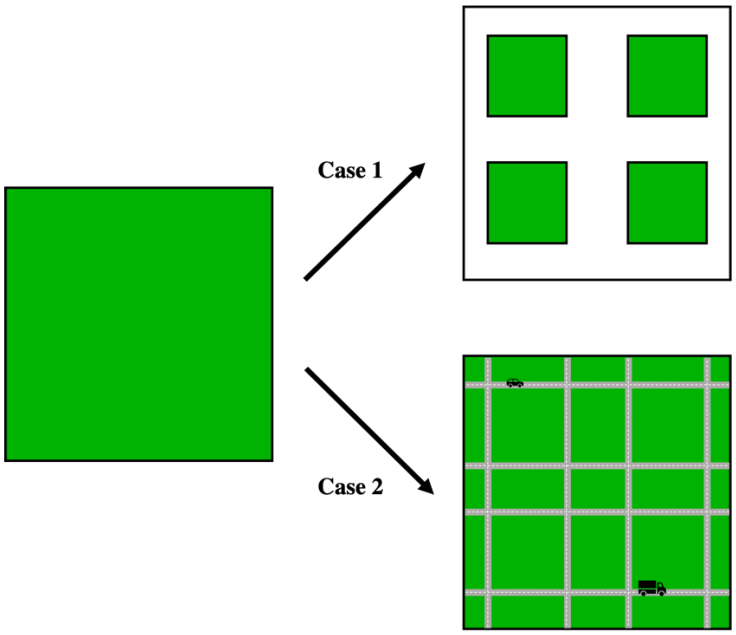


Figure 8. Two possible scenarios of HL&F. Case 1: HL&F by means of habitat contraction; Case 2: HL&F following the establishment of strict barriers to gene-flow.

We hope that the simulation and numerical results presented in Chapter 3 will be useful for the scientific community as a general overview of how spatial genetic pattern changes in the context of HL&F. The study of Hutchison and Templeton, (1999) have contributed much in that sense so far. In that work, the authors proposed four hypothetical relationships between F_{st} and geographical distance, classified on the basis of how drift and gene-flow influence spatial genetic structure in species with geographically limited dispersal. However, Hutchison and Templeton, (1999) lacked a quantitative framework on which these four hypotheses could rely on and being tested. With our study, we aimed to extend previous knowledge on the behaviour of IBD in time and space, and provide an analytical framework that evolutionary biologists can use to explore expected patterns according to the study case of interest.

The work developed in Chapter 3 could be extended as an inferential framework in several ways. In the following subsections, I describe three future directions of research on this regard: i) HL&F in a torus-like space (stepping-stone model); ii) HL&F in continuous and finite space; iii) HL&F as barrier to gene-flow in continuous space.

6.3.1 HL&F in a torus-like space (stepping-stone model)

The model is equivalent to the one investigated in Chapter 3 and summarised in Fig. 9.

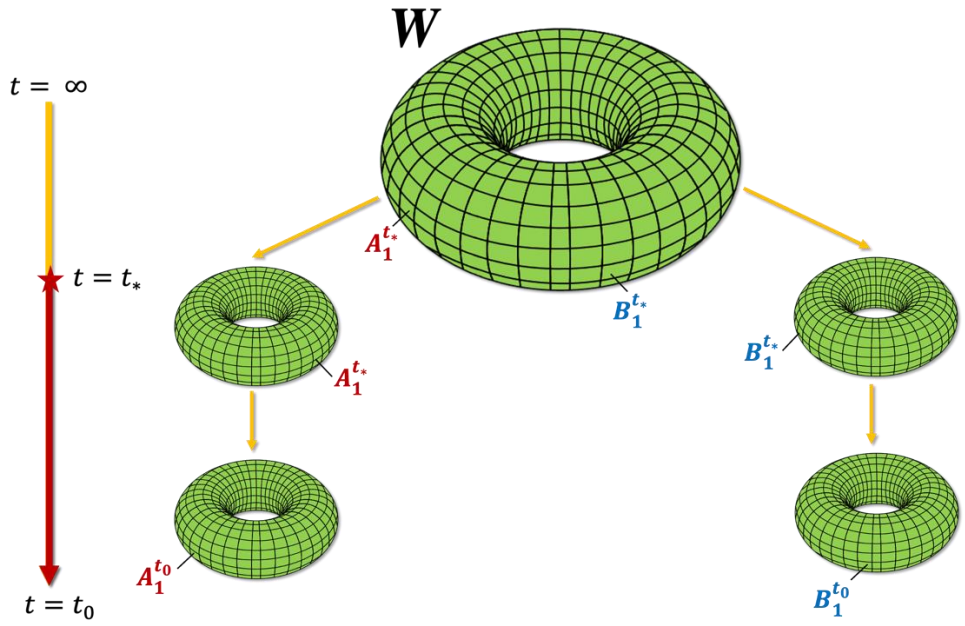


Figure 9. HL&F in a torus-like space (stepping-stone model). $\boldsymbol{A}_1^{t_0}$ and $\boldsymbol{B}_1^{t_0}$ indicate the location of individuals sampled in habitat fragments A and B, respectively, at time $t = t_0$. $\boldsymbol{A}_1^{t_*}$ and $\boldsymbol{B}_1^{t_*}$ indicate one possible location of the ancestral lineages at the time of HL&F ($t = t_*$) for individuals sampled in habitat fragments A and B, respectively. W: finite two-dimensional habitat prior to HL&F. $t = \infty$ specifies that population W is at mutation-drift-dispersal equilibrium.

Consider that the average coalescence time for two lineages sampled *i* and *j* steps apart after HL&F can be expressed as:

$$T_{A,B}(t_*) = t_* + \sum_{x_{A_2}=0}^{d_{x(A)}-1} \sum_{y_{A_2}=0}^{d_{y(A)}-1} \sum_{x_{B_2}=0}^{d_{x(B)}-1} \sum_{y_{B_2}=0}^{d_{x(B)}-1} S_{ij}\, P\left(\boldsymbol{A}_1^{t_*}, \boldsymbol{B}_1^{t_*} \mid \boldsymbol{A}_1^{t_0}, \boldsymbol{B}_1^{t_0}\right) + t_{00} \qquad \text{(Eq. 9)}$$

where t_* is the time since HL&F; $d_{x(A)}$ and $d_{y(A)}$ are the number of demes at x- and y-directions in fragment A, respectively; $d_{x(B)}$ and $d_{y(B)}$ are the number of demes at x- and y- directions in fragment B, respectively; $\boldsymbol{A}_1^{t_0} = (x_{A_1}^{t_0},\ y_{A_1}^{t_0})$ and $\boldsymbol{B}_1^{t_0} = (x_{B_1}^{t_0},\ y_{B_1}^{t_0})$ are the geographical coordinates of the sampled lineages in habitat fragment A and B, respectively; $\boldsymbol{A}_1^{t_*} = (x_{A_1}^{t_*},\ y_{A_1}^{t_*})$ and $\boldsymbol{B}_1^{t_*} = (x_{B_1}^{t_*},\ y_{B_1}^{t_*})$ are the geographical locations of the ancestral lineages at the time of HL&F (t_*); S_{ij} is the average amount of time required for two lineages at $i = \left|x_{A_1}^{t_*} - x_{B_1}^{t_*}\right|$ and $j = \left|y_{A_1}^{t_*} - y_{B_1}^{t_*}\right|$ steps apart to be in the same subpopulation; $t_{00} = 2N_e d_x d_y$ is the average time to coalesce once the two lineages occupy the same subpopulation before HL&F (t_*); N_e is the subpopulation population size, which is assumed to be constant before and after HL&F; and d_x and d_y are the number of demes at x- and y-directions in the torus-like space prior to HL&F. The model here considered have then the following parameters to estimate, $\theta = (N_e, d_x, d_y, d_{x(A)}, d_{y(A)}, d_{x(B)}, d_{y(B)}, t_*)$. The number of parameters could be reduced if for instance we assume equal number of demes at x- and y-directions, and equal number of demes in fragments A and B, $\theta = (N_e,\ d_{before\ HL\&F},\ d_{after\ HL\&F}, t_*)$.

Using genome-wide DNA sequence data and assuming independent loci and absence of recombination rate within locus, one could fit the

expression $T_{A,B}(t_*)$ with an estimate of observed $T_{A,B}$ for any pairwise comparison between individuals sampled in fragment A and B. The observed $T_{A,B}$ can be obtained, for instance, using the frequentist estimator of Tang et al., (2002): $\hat{T}_{Freq} = D_{A,B}/2\mu$, where $D_{A,B}$ is the number of nucleotide differences observed between two individuals sampled in fragment A and B, and μ is the mutation rate. A coarse-grid search of the whole parameter space, within reasonable parameter ranges, could be performed for identifying the best combination of parameters θ, namely the parameter values that minimize the sum of square distances between observed and expected data.

6.3.2 HL&F in continuous and finite space

We could model HL&F by approximating the spatial movement of lineages back in time using classical results on Brownian motion. We consider that an initial infinite two-dimensional population W at mutation-drift-dispersal equilibrium undergoes HL&F at time t_*, generating two finite 2D-habitat fragments A and B (Fig. 10).

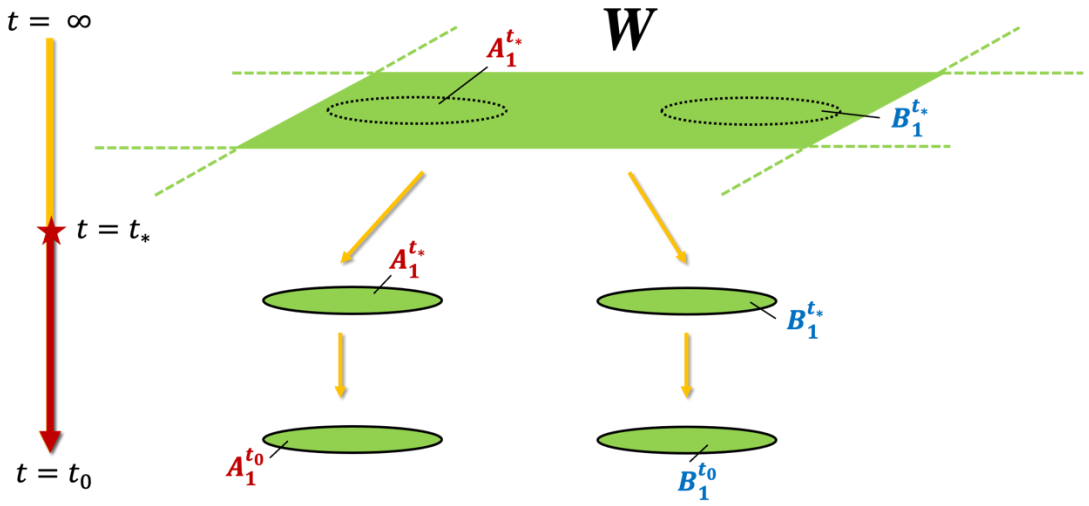


Figure 10. HL&F in continuous and finite space. $\boldsymbol{A}_1^{t_0}$and $\boldsymbol{B}_1^{t_0}$ indicate the location of individuals sampled in habitat fragments A and B, respectively, at time $t = t_0$. $\boldsymbol{A}_1^{t_*}$ and $\boldsymbol{B}_1^{t_*}$ indicate one possible location of the ancestral lineages at the time of HL&F ($t = t_*$) for individuals sampled in habitat fragments A and B, respectively. W: infinite two-dimensional habitat prior to HL&F. $t = \infty$ specifies that population W is at mutation-drift-dispersal equilibrium.

Assume that two individuals $\boldsymbol{A}_1^{t_0} = (x_{A_1}^{t_0},\ y_{A_1}^{t_0})$ and $\boldsymbol{B}_1^{t_0} = (x_{B_1}^{t_0},\ y_{B_1}^{t_0})$ are sampled in fragment A and B, respectively. A quantity of interest is the probability that lineages $\boldsymbol{A}_1^{t_0}$ and $\boldsymbol{B}_1^{t_0}$ coalesce at time T. Note that no coalescence is possible until the two lineages of $\boldsymbol{A}_1^{t_0}$ and $\boldsymbol{B}_1^{t_0}$ are in the ancestral continuous population W, given that fragments A and B are isolated from each other since HL&F. As in (Duforet-Frebourg and Slatkin, 2016) and in Chapter 3, the coalescence process between lineages $\boldsymbol{A}_1^{t_0}$ and $\boldsymbol{B}_1^{t_0}$ can be divided in three phases, going backward in time: i) the genealogy that traces back to the position of the ancestral lineages $\boldsymbol{A}_1^{t_*}$ and $\boldsymbol{B}_1^{t_*}$ within fragment A and B, respectively, at the time of HL&F; ii) the time until the ancestral lineages $\boldsymbol{A}_1^{t_*}$ and $\boldsymbol{B}_1^{t_*}$ are in the same subpopulation within the continuous

habitat W; and iii) time to the most common ancestor, once the lineages of $\boldsymbol{A}_1^{t_*}$ and $\boldsymbol{B}_1^{t_*}$ are in the same subpopulation. Therefore, the probability that lineages $\boldsymbol{A}_1^{t_0}$ and $\boldsymbol{B}_1^{t_0}$ coalesce at time $T = t + t_*$, can be written as:

$$\psi(T|r^{t_0},\theta) = \int_{\min x_A}^{\max x_A}\int_{\min y_A}^{\max y_A}\int_{\min x_B}^{\max x_B}\int_{\min y_B}^{\max y_B} P(\boldsymbol{A}_1^{t_*}|\boldsymbol{A}_1^{t_0},\theta)\, P(\boldsymbol{B}_1^{t_*}|\boldsymbol{B}_1^{t_0},\theta)\, \psi_w(t|r^{t_*},\theta)\, d_{x_A} d_{y_A} d_{x_B} d_{y_B} \quad \textit{(Eq. 10)}$$

where, $P(\boldsymbol{A}_1^{t_*}|\boldsymbol{A}_1^{t_0},\theta)$ and $P(\boldsymbol{B}_1^{t_*}|\boldsymbol{B}_1^{t_0},\theta)$ are the probabilities that the ancestral lineages of $\boldsymbol{A}_1^{t_0}$ and $\boldsymbol{B}_1^{t_0}$ were at position $\boldsymbol{A}_1^{t_*}$ and $\boldsymbol{B}_1^{t_*}$ at the time of HL&F, $\psi_w(t|r^{t_*},\theta)$ is the probability that the ancestral lineages $\boldsymbol{A}_1^{t_*}$ and $\boldsymbol{B}_1^{t_*}$ have a common ancestor at time t, given that they are at a distance $r^{t_*} = \sqrt{\left(x_{A_1}^{t_*} - x_{B_1}^{t_*}\right)^2 + \left(y_{A_1}^{t_*} - y_{B_1}^{t_*}\right)^2}$. Expressions on the distribution of coalescence times for a pair of sequences in a rectangular habitat (Wilkins, 2004) could be used for $P(\boldsymbol{A}_1^{t_*}|\boldsymbol{A}_1^{t_0},\theta)$ and $P(\boldsymbol{B}_1^{t_*}|\boldsymbol{B}_1^{t_0},\theta)$. Alternatively, we could for instance use the probability density function (PDF) of a Brownian motion in confined geometries, such as a disc of radius a (Bickel, 2007; Mortensen et al., 2021) and modify *Eq. 10* accordingly. In this case, the geographical position of each lineage is expressed in polar coordinates, $\boldsymbol{A}_1^{t_0} = \boldsymbol{A}_1^{t_0}(d_{A_1}^{t_0}, \phi)$ with $0 \leq d_{A_1} < a$ and $0 \leq \phi < 2\pi$. The corresponding PDF for finding the lineage of $\boldsymbol{A}_1^{t_0}$ in $\boldsymbol{A}_1^{t_*}$ at time t_* is expressed as:

$$P\left(\boldsymbol{A}_1^{t_*}\middle|\boldsymbol{A}_1^{t_0},\theta\right) = \frac{1}{\pi a^2} + \frac{1}{\pi a^2}\sum_{l=-\infty}^{\infty}\cos[l(\phi - \phi_0)]\sum_{m=1}^{\infty}\frac{\alpha_{lm}^2}{\alpha_{lm}^2 - l^2}\exp\left[-\alpha_{lm}^2\frac{t_*}{\tau}\right]\frac{J_l\left(\alpha_{lm}\frac{d_{A_1}^{t_*}}{a}\right)J_l\left(\alpha_{lm}\frac{d_{A_1}^{t_0}}{a}\right)}{J_l(\alpha_{lm})^2} \quad \textit{(Eq. 11)}$$

where $\tau = a^2/D = 2a^2/\sigma^2$ and $\alpha_{lm} > 0$ is the solution of the derivative of the Bessel function of the first kind of order l equal to 0, $J_l'(x) = 0$, which are ordered in ascending order of magnitude: $0 < \alpha_{l1} < \alpha_{l2} < \cdots \alpha_{lm}$. Instead, $\psi_w(t|r^{t_*},\theta)$ in *Eq. 10* would correspond to:

$$\psi_w(t|r^{t_*},\theta) = \frac{1}{2D_e}\frac{1}{4\pi\sigma^2 t}\,e^{-\frac{\left(\boldsymbol{A}_1^{t_*}-\boldsymbol{B}_1^{t_*}\right)^2}{4\sigma^2 t}} \quad \textit{(Eq. 12)}$$

where D_e is the subpopulation density, or number of diploid individuals in a deme. The cumulative density distribution (CDF) can then be expressed as:

$$CDF(r^{t_0},\theta) = \int_{\min x_A}^{\max x_A}\int_{\min y_A}^{\max y_A}\int_{\min x_B}^{\max x_B}\int_{\min y_B}^{\max y_B} P\left(\boldsymbol{A}_1^{t_*}\middle|\boldsymbol{A}_1^{t_0},\theta\right)P\left(\boldsymbol{B}_1^{t_*}\middle|\boldsymbol{B}_1^{t_0},\theta\right)\left[\int_0^{\infty}\psi_w(t|r^{t_*},\theta)\,dt\right]d_{x_A}d_{y_A}d_{x_B}d_{y_B} \quad \textit{(Eq. 13)}$$

The $CDF(r^{t_0}, \theta)$ can then be used to compute the probability of identity-by-descent of two individuals $\boldsymbol{A}_1^{t_0}$ and $\boldsymbol{B}_1^{t_0}$ as the probability that two lineages coalesce before a mutation event occurs along each of the two lineages:

$$F(r^{t_0}, \theta) = \int_{\min x_A}^{\max x_A} \int_{\min y_A}^{\max y_A} \int_{\min x_B}^{\max x_B} \int_{\min y_B}^{\max y_B} P\left(\boldsymbol{A}_1^{t_*} \middle| \boldsymbol{A}_1^{t_0}, \theta\right) P\left(\boldsymbol{B}_1^{t_*} \middle| \boldsymbol{B}_1^{t_0}, \theta\right) \left[\int_0^{\infty} \psi_w(t|r^{t_*}, \theta)\, e^{-2\mu(t+t_*)} dt\right] d_{x_A} d_{y_A} d_{x_B} d_{y_B} \quad \textit{(Eq. 14)}$$

where $e^{-2\mu(t+t_*)}$ is the probability of not mutating since the common ancestor at time $T = t + t_*$. A relatively simple inferential approach could be developed using the probability of identity-by-descent $F(r^{t_0}, \theta)$ for modelling the expected proportion of pairwise homozygosity between two individuals sampled in A and B, given that HL&F has occurred at time t_* in the past:

$$\mathrm{E}(H_{r^{t_0}}|\theta) = F(r^{t_0}, t_*) + [1 - F(r^{t_0}, t_*)]\, \frac{1}{n} \sum_i \left(\bar{p}_i^2 + (1 - \bar{p}_i)^2\right) \quad \textit{(Eq. 15)}$$

where $i = 1 \ldots n$ is the number of independent loci and $\bar{p}_i$ is the mean allele frequency for marker i. As in Ringbauer et al., (2018), the observed proportion of pairwise homozygosity $H_{r^{t_0}}$ between two individuals sampled in A and B, with genotypes $g_{a_1}^i$ and $g_{b_1}^i$ can be computed as:

$$H_{r^{t_0}} = \frac{1}{n} \sum_i \left(g_{a_1}^i g_{b_1}^i + \left(1 - g_{a_1}^i\right)\left(1 - g_{b_1}^i\right)\right) \quad \textit{(Eq. 16)}$$

Using numerical optimization procedures, one could find the parameter values $\theta = (\sigma, t_*, D_e, \mu)$, or $\theta = (\sigma, t_*, D_e)$ if mutation rate is known, that minimize the sum of all squared deviations between expected and observed pairwise homozygosity.

6.3.3 HL&F as barrier to gene-flow in continuous space

We could use a simplification of the HL&F process by considering a two-dimensional habitat that is suddenly interrupted by a barrier (e.g., savannah, road). This model was first developed by Ringbauer et al., (2018), based on analytical results from Grebenkov et al., (2014). Here, I propose an extension of the Ringbauer et al., (2018) framework by modelling non-equilibrium isolation-by-distance pattern assuming an impermeable barrier to gene-flow ($K \approx 0$) (Annex). In this model, an initial infinite two-dimensional population W at dispersal-mutation-drift equilibrium is interrupted by a vertical barrier at time t_* in the past, such that individuals occurring at the left side A and right side B of the barrier cannot disperse across the barrier after HL&F (Fig. 11).

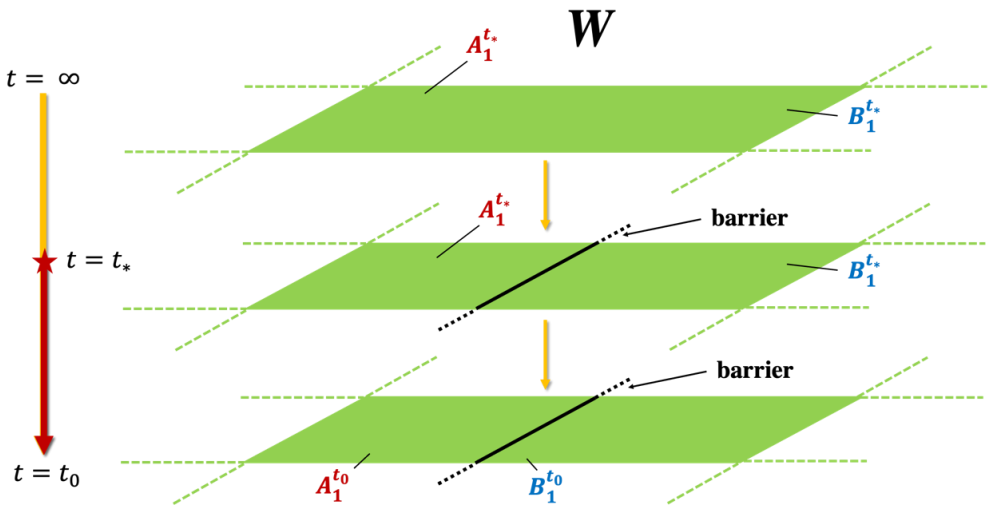


Figure 11. HL&F as barrier to gene-flow in continuous space. $\boldsymbol{A}_1^{t_0}$ and $\boldsymbol{B}_1^{t_0}$ indicate the location of individuals sampled at the left side A and right side B of the barrier at time $t = t_0$. $\boldsymbol{A}_1^{t_*}$ and $\boldsymbol{B}_1^{t_*}$ indicate one possible location of the ancestral lineages at the time of HL&F ($t = t_*$). W: infinite two-dimensional habitat prior to HL&F. $t = \infty$ specifies that population W is at mutation-drift-dispersal equilibrium.

Assuming that we sample two genes at $\boldsymbol{A}_1^{t_0} = (x_{A_1}^{t_0}, y_{A_1}^{t_0})$ and $\boldsymbol{B}_1^{t_0} = (x_{B_1}^{t_0}, y_{B_1}^{t_0})$ at the present time ($t = t_0$), we are interested in deriving the probability that these two sampled genes are identical-by-descent. As shown above, probability of identical-by-descent can be derived by first modelling the cumulative probability distribution of coalescence times. Again, the coalescence process between the genes sampled in $\boldsymbol{A}_1^{t_0}$ and $\boldsymbol{B}_1^{t_0}$ is described as a three phases process: i) the possible paths travelled by $\boldsymbol{A}_1^{t_0}$ and $\boldsymbol{B}_1^{t_0}$ lineages within each side of the barrier since t_*; ii) the probability that the lineages of the ancestors $\boldsymbol{A}_1^{t_*}$ and $\boldsymbol{B}_1^{t_*}$ come close at some point in the past before HL&F; and iii) the probability that the two lineages coalesce once they are in the same location. Thus, the probability that lineages $\boldsymbol{A}_1^{t_0}$ and $\boldsymbol{B}_1^{t_0}$ coalesce at time $T = t + t_*$, can be written as:

$$\psi(T|r^{t_0},\theta) = \int_{-\infty}^{0}\int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty} G\left(\boldsymbol{A}_1^{t_*}|\boldsymbol{A}_1^{t_0},\theta\right) G\left(\boldsymbol{B}_1^{t_*}|\boldsymbol{B}_1^{t_0},\theta\right) \psi_w(t|r^{t_*},\theta)\, d_{x_A} d_{y_A} d_{x_B} d_{y_B} \quad \textit{(Eq. 17)}$$

where $\psi_w(t|r^{t_*},\theta)$ is the probability that the ancestral lineages $\boldsymbol{A}_1^{t_*}$ and $\boldsymbol{B}_1^{t_*}$ have a common ancestor at time t, given that they are at a distance $r^{t_*} = \sqrt{\left(x_{A_1}^{t_*} - x_{B_1}^{t_*}\right)^2 + \left(y_{A_1}^{t_*} - y_{B_1}^{t_*}\right)^2}$ (*Eq. 12*), whereas $G\left(\boldsymbol{A}_1^{t_*}|\boldsymbol{A}_1^{t_0},\theta\right)$ or $G\left(\boldsymbol{B}_1^{t_*}|\boldsymbol{B}_1^{t_0},\theta\right)$ is the probability of moving from $\boldsymbol{A}_1^{t_0}$ or $\boldsymbol{B}_1^{t_0}$, to $\boldsymbol{A}_1^{t_*}$ or $\boldsymbol{B}_1^{t_*}$, respectively. In particular:

$$G\left(\boldsymbol{A}_1^{t_*}|\boldsymbol{A}_1^{t_0},\theta\right) = G_{ss}\left(x_{A_1}^{t_*}|x_{A_1}^{t_0},\theta\right) G\left(y_{A_1}^{t_*}|y_{A_1}^{t_0},\theta\right) \quad \textit{(Eq. 18)}$$

In *Eq. 18*, $G\left(y_{A_1}^{t_*}|y_{A_1}^{t_0},\theta\right)$ is the probability that the ancestral lineages of $\boldsymbol{A}_1^{t_0}$ were at position $y_{A_1}^{t_*}$ on the y-direction, which is equal to $G\left(y_{A_1}^{t_*}|y_{A_1}^{t_0},\theta\right) = \frac{1}{\sqrt{2\pi\sigma^2 t_*}} \exp\left(-\frac{\left(y_{A_1}^{t_*} - y_{A_1}^{t_0}\right)^2}{2\sigma^2 t_*}\right)$ given that movement along the y-direction is not affected by the presence of any barrier, and $G_{ss}\left(x_{A_1}^{t_*}|x_{A_1}^{t_0},\theta\right)$ is the probability that the ancestral lineages of $\boldsymbol{A}_1^{t_0}$ were at position $x_{A_1}^{t_*}$ on the x-direction, which in this case is affected by the presence of a vertical barrier. Expressions for $G_{ss}\left(x_{A_1}^{t_*}|x_{A_1}^{t_0},\theta\right)$ are detailed in Ringbauer et al., (2018) and Chapter 1 (*Eq. 20*). We can then write the cumulative probability of identity-by-descent $F(r^{t_0},\theta)$ as:

$$F(r^{t_0},\theta) = \int_{-\infty}^{0}\int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty} G\left(\boldsymbol{A}_1^{t_*}|\boldsymbol{A}_1^{t_0},\theta\right) G\left(\boldsymbol{B}_1^{t_*}|\boldsymbol{B}_1^{t_0},\theta\right) \left[\int_{0}^{\infty} \psi_w(t|r^{t_*},\theta) e^{-2\mu(t+t_*)} dt\right] d_{x_A} d_{y_A} d_{x_B} d_{y_B} \quad \textit{(Eq. 19)}$$

As in the section above (*Eq. 15*), $F(r^{t_0}, \theta)$ could be used for computing the expected proportion of pairwise homozygosity between two individuals sampled in A and B, which is then fitted to the observed proportion of pairwise homozygosity for inferring $\theta = (\sigma, t_*, D_e, \mu)$, or $\theta = (\sigma, t_*, D_e)$ if mutation rate is known.

An alternative inferential framework, which might be more accurate for recent HL&F events, would focus on the distribution of shared identity-by-descent segments between individuals sampled in A and B. Following Ralph and Coop, (2013) and Ringbauer et al., (2017), the expected number of shared DNA segments of map length L between two individuals at geographical distance $r^{t_0} = \sqrt{\left(x_{A_1}^{t_0} - x_{B_1}^{t_0}\right)^2 + \left(y_{A_1}^{t_0} - y_{B_1}^{t_0}\right)^2}$ can be written as:

$$E[N_L \mid r^{t_0}, \theta] = \int_0^\infty E[N_L^t \mid r^{t_0}, \theta]\, dt \qquad \textit{(Eq. 20)}$$

where $E[N_L^t \mid r^{t_0}, \theta]$ is the expected number of shared DNA segments of map length L between two individuals at geographical distance r^{t_0} that coalesce at time t, which is modelled by *Eq. 1*. In the context of HL&F, *Eq. 1* becomes:

$$E[N_L \mid r^{t_0}, \theta]_{barrier} = \int_{-\infty}^{0}\int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{\infty} G\left(\boldsymbol{A}_1^{t_*} \middle| \boldsymbol{A}_1^{t_0}, \theta\right) G\left(\boldsymbol{B}_1^{t_*} \middle| \boldsymbol{B}_1^{t_0}, \theta\right) \left[\int_0^\infty \psi_w(t \mid r^{t_*}, \theta) E\left[K_L^{t+t_*}\right] dt\right] d_{x_A} d_{y_A} d_{x_B} d_{y_B} \qquad \textit{(Eq. 21)}$$

if HL&F is modelled with an impermeable barrier to gene-flow, or

$$E[N_L \mid r^{t_0}, \theta]_{fragments} = \int_{\min x_A}^{\max x_A}\int_{\min y_A}^{\max y_A}\int_{\min x_B}^{\max x_B}\int_{\min y_B}^{\max y_B} P\left(\boldsymbol{A}_1^{t_*} \middle| \boldsymbol{A}_1^{t_0}, \theta\right) P\left(\boldsymbol{B}_1^{t_*} \middle| \boldsymbol{B}_1^{t_0}, \theta\right) \left[\int_0^\infty \psi_w(t \mid r^{t_*}, \theta) E\left[K_L^{t+t_*}\right] dt\right] d_{x_A} d_{y_A} d_{x_B} d_{y_B} \qquad \textit{(Eq. 22)}$$

if HL&F is modelled with finite habitat fragments (Section 6.3.2). As described in Section 6.2, one could approximate a composite likelihood using a Poisson distribution with mean $\lambda_i(r^{t_0}, \theta) = E\left[N_{L_i} | r^{t_0}, \theta\right]\Delta L$ and use numerical procedure to find the maximum-likelihood demographic parameters $\theta = (\sigma, t_*, D_e)$.

6.4 Landscape genomics of forest-dwelling species in northern Madagascar: insights of past forest cover changes

Natural and anthropogenic habitat loss and fragmentation have shaped the distribution and genetic diversity of many species across the world (e.g., Hewitt, 2000; Sandel et al., 2011; Weigelt et al., 2016). Paleoenvironmental and fossil data have helped to reconstruct past vegetational changes and how species have responded to those changes (Nogués-Bravo et al., 2018; Fordham et al., 2020). However, palaeoecological data have limitations that hamper our understanding of species response to HL&F, such as sparse and uneven spatial coverage (Kummel and Raup, 1965; Gorham et al., 2001; Berglund, 2003; Croft et al., 2018).

Genetic data contain information on past changes in connectivity and population size for a given species (Kuhner et al., 1995; Beerli and Felsenstein, 1999; Nielsen, 2000; Goldstein and Chikhi, 2002; Hey and Nielsen, 2004; Beaumont, 2010; Mazet et al., 2015, 2016; Chikhi et al., 2018; Wang et al., 2020; Arredondo et al., 2021). Under the assumptions that i) a given species is strictly associated to its habitat and ii) species' preferences for this habitat have not changed over relatively recent times, genetic data from this species represent a source of information, complementary to paleoenvironmental data, that can provide fine-scale details on past habitat

cover changes (Hewitt, 2000; Hu et al., 2009). Alternatively, if information on past habitat changes are available and reliable, one could use genetic data from a habitat-associated species to infer how these changes have influenced species connectivity and population size, and thus infer how this species have responded to such habitat changes (Fordham et al., 2014).

In Madagascar, paleoenvironmental data have played a fundamental role in documenting changes in climate and vegetation over the last tens of thousands of years (Burney, 1987; Burns et al., 2016; Voarintsoa et al., 2017, 2019; Teixeira et al., 2021; Domic et al., 2021). Nevertheless, palaeoecological evidences are still constrained to few ecoregions across the island. Moreover, it is unclear the extent on which past vegetation reconstructions retrieved from localized palaeoecological samples can be generalized to entire regions or are representative of local past environmental changes (e.g., Teixeira et al., 2021; Voarintsoa et al., 2021). Northern Madagascar is a peculiar ecoregion in the island, since it is a transition zone between the eastern humid and western dry ecoregions, containing forest patches of either dry or humid vegetation (Gautier et al., 2006). Palaeoecological data from northern Madagascar are still limited. Note, for instance, that the first paleoenvironmental study in this region was published in 2021 by Teixeira et al., (2021). Previous studies have attempted to fill this gap by carrying out demographic inferences, as proxy for reconstructing past changes in forest cover, using genetic data collected from forest-dwelling species (e.g., Quéméré et al., 2012; Rakotoarisoa et al., 2013a, b; Salmona et al., 2017). With the exception of Salmona et al., (2017), these studies were focusing on one species. However, studies based on only one species might be somehow confounded by species-specific evolutionary processes that are unrelated to past vegetation changes. A multi-species approach may reduce such hypothetical bias, since it can help identifying the evolutionary processes, environmental events or landscape features that played a major role in shaping genetic diversity across species (Cushman and Landguth,

2012; López-Duarte et al., 2012; Fourtune et al., 2016; Ashrafzadeh et al., 2020).

Following a multi-species approach, in Chapter 4, we investigated spatial population genomic patterns across five forest-dwelling endemic rodents (*Eliurus* spp.) occurring in northern Madagascar: *E. carletoni*, *E. myoxinus*, *E. ellermani*, *E. minor* and *Eliurus* sp. nova. These species are typically associated to two major vegetation types in northern Madagascar: *E. carletoni* to dry deciduous forests; *E. ellermani*, *E. minor*, *E. myoxinus* and *Eliurus* sp. nova to humid evergreen forests (e.g., Jansa et al., 2019; Chapter 4; Chapter 5). As forest-dwelling, ecologically diverse and co-occurring species, these five *Eliurus* spp. represent a good model for assessing i) how ecologically diverse species respond to the same landscape features; ii) identify major barrier to gene-flow across northern Madagascar; and iii) provide some insights on the humid and dry forest cover changes that may have occurred during the late Pleistocene-Holocene in northern Madagascar.

To put it briefly, our results on *Eliurus* spp. and the findings of previous studies on other co-occurring taxa suggest that both dry-associated and humid-associated species have undergone range expansion at some point in the recent past (Rakotoarisoa et al., 2013a, b; Shi et al., 2013), probably following the climatic oscillations that occurred during the last 25,000 years.

During the African Humid Period (15,200 - 5,000 years ago), climate and vegetation types were more humid than today in northern Madagascar (Jungers et al., 1995; Simons et al., 1995; Teixeira et al., 2021). These environmental conditions may have induced the range contraction of dry-associated species, such as the rodent *Eliurus carletoni* and the lemurs *Microcebus tavaratra* and *Lepilemur ankaranensis*, and promoted the expansion of humid-associated species, such as the rodents *E. myoxinus*, *E. minor* and *E. ellermani*, and the lemur *Microcebus arnholdi* (Sgarlata et al., 2019; Chapter 4; Chapter 5; Annex). This interpretation is supported by genetic pattern (mtDNA and/or nuclear genomic data) from humid- and dry-associated species:

- The humid-associated rodent *Eliurus ellermani*: i) nuclear genomic data exhibit a steady decline in H_{obs} with geographical distance from the forest complex ANTSB-BIN (Chapter 4); ii) neutrality tests and mismatch distribution analyses performed on mtDNA from individuals sampled in ANTSB-BIN point out to a historical population expansion (Rakotoarisoa et al., 2013a);
- The northern humid-associated rodent *E. myoxinus*: i) nuclear genomic data exhibit a steady decline in H_{obs} with geographical distance from the forest complex ANTSB-BIN (Chapter 4); ii) spatial phylogeographic analyses and neutrality tests carried out on individuals sampled in BIN showed evidence of recent expansion (Shi et al., 2013);
- The humid-associated rodent *E. minor*: nuclear genomic data showed a steady decline in H_{obs} with geographical distance from ANDR forest (Chapter 4);
- The humid-associated mouse lemur *Microcebus arnholdi*: i) nuclear genomic data exhibit a steady decline in H_{obs} with geographical distance from South Loky-Manambato region to BIN forest, suggesting an expansion from the northeast to the north (Annex);
- The dry-associated medium size lemurs *Propithecus tattersalli* and *Propithecus perrieri*: studies using genetic data from individuals sampled across the dry deciduous forests of LM and NLM regions, respectively, have dated a population decline and fragmentation between 10,000 and 4,000 years ago, a time interval included approximatively in the African Humid Period (Quéméré et al., 2012; Salmona et al., 2017).

This scenario is also supported by non-genetic data, such as subfossil-data and present-day dry *versus* humid-associated species distributions. Palaeoecological studies carried out in the present-day dry NLM region (Ankarana, Andavakoera and Montagne des Français) found subfossils of extant lemur species (*Propithecus cf. diadema*, *Indri indri*, and *Prolemur simus*) which currently occur only in the humid forests of eastern Madagascar (Jungers et al., 1995; Simons et al., 1995; Godfrey et al., 1999). The radiocarbon dating analysis of *P. simus* and *I. indri* specimens estimated that those individuals were alive about 4,560 ± 70 Cal. YBP and 4,400 ± 60 Cal. YBP, respectively, suggesting that northern Madagascar was more humid than today until at least ~4,500 years ago (Jungers et al., 1995; Simons et al., 1995). Further support for this scenario is given by the distribution of humid-associated species within isolated humid patches scattered across northern Madagascar. In particular, the fact that isolated humid patches are separated by dry vegetation and are occupied by similar species community, suggest that at a certain point in the past they were somehow connected. Evidence for such conclusions comes, in particular, from the distribution of the two endemic rodents *Eliurus myoxinus*, *Eliurus ellermani* (Chapter 4 and Chapter 5), the mouse lemur *Microcebus arnholdi* (Sgarlata et al., 2019) and several other taxa (tenrecs: *Microgale fotsyfotsy*, *Nesogale talazaci*; rodents: *Eliurus majori*; frogs: *Mantidactylus ambreensis*; *Boophis blommersae*, *B. septentrionalis, Gephyromantis ambohitra*; snake: *Pseudoxyrhopus ambreensis*; and chameleons: *Brookesia minima*), which point out to a biogeographic link between the southern and northern edge of northern Madagascar (e.g., Andreone et al., 2009; Everson et al., 2020; Rasolonjatovo et al., 2020).

At the end of the African Humid Period (~5,000 years ago), northern Madagascar experienced a reduction in precipitations, associated to warmer and drier climatic conditions (Teixeira et al., 2021). During this period, dry forests may have increased their spatial coverage across most of northern Madagascar and promoted the expansion of dry-associated species.

Conversely, humid vegetation and its associated species may have undergone HL&F, and shifted upward their lower altitudinal limit, similar to what it has been reported for eastern Africa (Ivory and Russell, 2018). Evidences for an expansion of dry-associated species come from:

- The dry-associated rodent *Eliurus carletoni*: nuclear genomic data exhibit a steady decline in H_{obs} with geographical distance from the far north to the southern edge of northern Madagascar (Chapter 4);
- The dry-associated mouse lemur *Microcebus tavaratra*: i) nuclear genomic data exhibit a steady decline in H_{obs} with geographical distance from the far north, as observed in *E. carletoni* (Annex); ii) neutrality tests and mismatch distribution analyses performed on mtDNA of individuals sampled across LM region, pointed out to a historical population expansion (Sgarlata et al., 2018);
- The dry-associated sportive lemur *Lepilemur ankaranensis* i) nuclear genomic data exhibit a steady decline in H_{obs} with geographical distance from the far north, as observed in *E. carletoni* and *M. tavaratra* (Annex);

Continuous reduction in precipitations ~900 years ago coincides with an increase in *Poaceae* and charcoal particles in the Maudit lake of Montagne d'Ambre (Teixeira et al., 2021), suggesting that lowland ecosystems shifted to open grasslands, which were also more susceptible to fire ignition. We hypothesize that the climatic shifts occurred within these last 5,000 years, and in particular the fast increase in drought of the last ~900 years, contributed the most to present-day landscape configuration.

Altogether, genetic data, species distribution and palaeoecological data suggest that humid vegetations were covering northern Madagascar in

the recent past, until ~4,000-5,000 years ago. If so, by looking at present-day landscape configuration, it is reasonable to consider that humid vegetations were gradually shifting into dry vegetation, thus allowing the expansion of dry-associated species. About 900 years ago, something changed, as suggested from paleoenvironmental data (Teixeira et al., 2021). The gradual and continuous reduction in precipitation was accompanied and contrasted by a rapid increase in arid environments. This scenario provides some insights on the recent history of northern Madagascar, that differ in some aspects from other proposed scenarios. In fact, although previous genetic studies have associated evidence of population decline of dry associated species with present-day fragmentation configuration (i.e., Quéméré et al., 2012; Salmona et al., 2017), we may perhaps consider that such genetic signatures are actually indicative of a population decline that has occurred when humid forests, unsuitable habitat type for dry-associated species, have expanded across Madagascar. The proposed scenario does not deny the importance of recent HL&F in shaping population genetic structure, as shown by the correlation between present-day forest configuration and genetic or genomic structure across various taxa (Quéméré et al., 2010; Rakotoarisoa et al., 2013a; Sgarlata et al., 2018; Aleixo-Pais et al., 2019; Salmona et al., 2021), but states that present-day fragmentation configuration is perhaps too recent (not more than 900 years ago) for being detected with classical methods of demographic inference.

Future analyses are required to test these alternative scenarios. The theoretical work carried out in Chapter 2, 3 and the species genomic pattern identified in Chapter 4, 5, point out to two major directions of research on the subject.

First, it is time to explicitly consider spatial genetic structure for reconstructing recent changes in population connectivity and populations size (Bradburd and Ralph, 2019). Since the recent part of the gene genealogy is strongly influenced by geography-dependent local processes (Wakeley, 1999; Wilkins, 2004; Barton et al., 2013), ignoring spatial population structure

have important implications on demographic inference (e.g., Battey et al., 2020). The frameworks in Chapter 2, 3 and the improvements that could be carried out (as proposed in Section 6.3) may help to correctly infer time since HL&F and distinguish recent from more ancient events.

Second, the spatial genomic analyses carried out across *Eliurus*, *Microcebus* and *Lepilemur* species have documented genetic patterns that can be attributed to range expansion. Appropriate modelling and dating of the expansion onset for each of those species is likely going to clarify how reliable is the expansion signature recovered from genetic data and when did it occur. To do so, a spatially explicit framework may be warranted. An example is the model developed by Austerlitz et al., (1997) and applied by Liu et al., (2006) to date the time since the onset of expansion in humans. In Austerlitz et al., (1997) (*Eq. 13*), only the case of a one-dimensional stepping-stone model with d demes is considered, such that the recurrence equation for expected coalescence times of two genes from demes i and j sampled at t generations since the onset of expansion, is:

$$T_{i,j}(t) = 1 + \sum_{k=1}^{d}\sum_{l=1}^{d} p(k\,|i,t)\; p(l\,|j,t)\; T_{k,l}(t-1)\; \left(1 - \frac{\delta_{k,l}}{2N_{k(t-1)}}\right) \qquad \text{(Eq. 23)}$$

where $p(k\,|i,t)$ and $p(l\,|j,t)$ are the probabilities of a lineage sampled in demes i and j, at time t since the onset of expansion, of being in deme k and l in the previous generation $(t-1)$, $T_{k,l}(t-1)$ is the expected coalescence times for two lineages that were in deme k and l in the previous generation, $\delta_{k,l} = 1$ if $k = l$ (that is, if lineages in demes i and j were in the same deme k in the previous generation) and $\delta_{k,l} = 0$ if $k \neq l$, and $N_{k(t-1)}$ is the diploid population size in deme k in the previous generation. Note, therefore, that $1 - \delta_{k,l}/2N_{k(t-1)}$ is equal to 1 when the lineages are in different demes in the

previous generation and $1 - 1/2N_{k(t-1)}$ when they are in the same deme in the previous generation. In other words, the last term means that if the two lineages are in the same deme in the previous generation, there is a probability $1 - 1/2N_{k(t-1)}$ of not coalescing and therefore the expected coalescence time will be given by $T_{k,l}(t-1)$ for $k = l$. The formulas for computing $N_{k(t-1)}$, $p(k\,|i,t)$ and $p(l\,|j,t)$ are detailed in Austerlitz et al., (1997) and are thus not reported in here. In Liu et al., (2006), the expression for $T_{i,j}(t)$ was fitted to microsatellite data from human populations. In particular, they considered that, under the Stepwise Mutation Model of microsatellite evolution, the variance of allele-repeat size within each population i (X_i), for $i \in (1,d)$, depends on expected coalescence time: $Var(X_i) = 2\mu T_{i,i}(t)$. Liu et al., (2006) carried out a coarse-grid search of the whole parameter space, by comparing observed $T_{i,i}(t)$ values, estimated from observed $Var(X_i)$, with expected values computed from *Eq. 23*, using known mutation rate value. The parameter values that minimized the sum of square distances between observed and expected data, were selected.

Similarly, using DNA sequence data obtained with RAD-seq approach and assuming independent loci and absence of recombination rate within locus, one could estimate T_i using the expression for the expected number of pairwise genetic differences between individuals sampled at i under an infinite allele model, $\theta_i = 2\mu T_{i,i}(t)$. Practically, we could estimate $T_{i,i}(t)$ using a frequentist estimator as used by Tang et al., (2002), $\hat{T}_{Freq} = D_i/2\mu$, where D_i is the number of nucleotide differences between any two sequences in deme i. One of the advantages of this estimator, beyond its easy calculation, is the fact that it does not depend on a specific demographic model. Alternative approaches for estimating T_i from nucleotide sequence data could be considered and are discussed, for instance, in King and Wakeley, (2016).

6.5 Implications for conservation

There are different ways in which the results of this PhD thesis contribute to the conservation of species in Madagascar and elsewhere. The theoretical work carried out in Chapter 2 and Chapter 3 examines the temporal trajectory of genetic diversity in spatially structured populations undergoing habitat loss and fragmentation (HL&F), i) providing a statistical approach able to quantify edge-impacted changes in genetic diversity, ii) proposing a theoretical framework describing the rate of decay in genetic diversity in a habitat patch subjected to habitat contraction, iii) characterising how HL&F influences *landscape-scale* spatial genetic structure under various demographic scenarios, likely to occur in natural populations; and iv) proposing a spatially-explicit theoretical framework modelling spatial genetic structure during HL&F.

HL&F includes processes acting at the *patch*- and *landscape-scale*, consisting in: edge-effect, patch-size, patch-isolation, and landscape-level habitat amount and fragmentation (Fahrig, 2003, 2019). The theoretical work in Chapter 2 and 3 attempt to investigate the role of those processes on the genetic diversity of species, by also proposing ways for quantitatively estimating how landscape configuration impacts the genetic diversity of populations. Therefore, we can imagine that the outcomes of this research may aid to understand how past, present and future anthropogenic or natural HL&F influence the genetic diversity of species.

Since genetic diversity represents the raw material for evolutionary processes that support survival and adaptation of populations, preserving genetic diversity is a key conservation goal in the current biodiversity crisis (Díaz et al., 2020; Kardos et al., 2021). Accordingly, the theoretical frameworks, such those proposed in Chapter 2 and 3, may be of assistance in the identification of the best ways to preserve species genetic diversity in the habitats that are continuously modified by human activities (Fordham et

al., 2014). Note, however, that the work developed in this PhD thesis assume that genetic variation is neutral, that is, it has no effect on the number of offspring of individuals. As such, the proposed theoretical frameworks are not a full representation of all relevant evolutionary process, but still useful i) since they consider processes common to species characterized by geographically-limited dispersal; and ii) because neutral processes play major roles in determining adaptive genetic variation (Kardos et al., 2021).

In Chapter 4 and 5 (as well as in Annex), we focused on endemic rodents and lemur species occurring in the fragmented landscapes of northern Madagascar, updating their distributional range, documenting their genetic and genomic diversity and describing previously unknown taxa. The natural forests of Madagascar have been subjected to severe HL&F over the last decades and predictions are not optimistic (Vieilledent et al., 2018). Documenting present-day biodiversity in such threatened natural environments is fundamental for raising awareness on the extraordinary species diversity present in the island, as well as providing information that will hopefully motivate the implementation or reinforcement of conservation strategies within the surveyed areas. As an example, note for instance, that the results presented in Chapter 4 and Annex have already contributed to the last update on the conservation status of lemur species for the IUCN Red List, which took place on May 2018 in Antananarivo, Madagascar (Blanco et al., 2020a, 2020e, 2020f, 2020g, 2020c, 2020d, 2020b; Borgerson et al., 2020; Ganzhorn et al., 2020; Razafindramanana et al., 2020; Sgarlata et al., 2020c, 2020a, 2020d, 2020b). Also, the fieldwork carried out in often remote areas of northern Madagascar provided information on species presence/absence to institutions and organizations (e.g., WWF, Fanamby, DEF) working directly for the conservation of natural forests in Madagascar. In particular, note that often these organizations do not have the logistical means to survey the still wide and largely unexplored forests of Madagascar, thus relying on the work of researchers for documenting the biodiversity of these areas.

6.6 References

Aguilar, R., Quesada, M., Ashworth, L., Herrerias-Diego, Y., Lobo, J., 2008. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. Molecular Ecology 17, 5177–5188.

Alcala, N., Streit, D., Goudet, J., Vuilleumier, S., 2013. Peak and Persistent Excess of Genetic Diversity Following an Abrupt Migration Increase. Genetics 193, 953–971.

Aleixo-Pais, I., Salmona, J., Sgarlata, G.M., Rakotonanahary, A., Sousa, A.P., Parreira, B., Kun-Rodrigues, C., Ralantoharijaona, T., Jan, F., Rasolondraibe, E., Minhós, T., Zaonarivelo, J.R., Andriaholinirina, N.V., Chikhi, L., 2019. The genetic structure of a mouse lemur living in a fragmented habitat in Northern Madagascar. Conserv Genet 20, 229–243.

Almeida-Rocha, J.M., Soares, L.A.S.S., Andrade, E.R., Gaiotto, F.A., Cazetta, E., 2020. The impact of anthropogenic disturbances on the genetic diversity of terrestrial species: A global meta-analysis. Molecular Ecology 29, 4812–4822.

Andreone, F., Glaw, F., Mattioli, F., Jesu, R., Schimmenti, G., Randrianirina, J.E., Vences, M., 2009. The peculiar herpetofauna of some Tsaratanana rainforests and its affinities with Manongarivo and other massifs and forests of northern Madagascar. Italian Journal of Zoology 76, 92–110.

Arredondo, A., Mourato, B., Nguyen, K., Boitard, S., Rodríguez, W., Noûs, C., Mazet, O., Chikhi, L., 2021. Inferring number of populations and changes in connectivity under the n-island model. Heredity 126, 896–912.

Ashrafzadeh, M.R., Khosravi, R., Adibi, M.A., Taktehrani, A., Wan, H.Y., Cushman, S.A., 2020. A multi-scale, multi-species approach for assessing effectiveness of habitat and connectivity conservation for endangered felids. Biological Conservation 245, 108523.

Austerlitz, F., Jung-Muller, B., Godelle, B., Gouyon, P.-H., 1997. Evolution of Coalescence Times, Genetic Diversity and Structure during Colonization. Theoretical Population Biology 51, 148–164.

Austerlitz, F., Mariette, S., Machon, N., Gouyon, P.H., Godelle, B., 2000. Effects of colonization processes on genetic diversity: differences between annual plants and tree species. Genetics 154, 1309–1321.

Barton, N.H., Etheridge, A.M., Kelleher, J., Véber, A., 2013. Inference in two dimensions: Allele frequencies versus lengths of shared sequence blocks. Theoretical Population Biology, Coalescent Theory 87, 105–119.

Barton, N.H., Wilson, I., 1995. Genealogies and geography. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 349, 49–59.

Battey, C.J., Ralph, P.L., Kern, A.D., 2020. Space is the Place: Effects of Continuous Spatial Structure on Analysis of Population Genetic Data. Genetics 215, 193–214.

Beaumont, M.A., 2010. Approximate Bayesian Computation in Evolution and Ecology. Annual Review of Ecology, Evolution, and Systematics 41, 379–406.

Beerli, P., Felsenstein, J., 1999. Maximum-Likelihood Estimation of Migration Rates and Effective Population Numbers in Two Populations Using a Coalescent Approach. Genetics 152, 763–773.

Berglund, B.E., 2003. Handbook of Holocene Palaeoecology and Palaeohydrology, Illustrated edition. ed. The Blackburn Press, Caldwell, N.J.

Betts, M.G., Wolf, C., Pfeifer, M., Banks-Leite, C., Arroyo-Rodríguez, V., Ribeiro, D.B., Barlow, J., Eigenbrod, F., Faria, D., Fletcher, R.J., Hadley, A.S., Hawes, J.E., Holt, R.D., Klingbeil, B., Kormann, U., Lens, L., Levi, T., Medina-Rangel, G.F., Melles, S.L., Mezger, D., Morante-Filho, J.C., Orme, C.D.L., Peres, C.A., Phalan, B.T., Pidgeon, A., Possingham, H., Ripple, W.J., Slade, E.M., Somarriba, E., Tobias, J.A., Tylianakis, J.M., Urbina-Cardona, J.N., Valente, J.J., Watling, J.I., Wells, K., Wearn, O.R., Wood, E., Young, R., Ewers, R.M., 2019. Extinction filters mediate the global effects of habitat fragmentation on animals. Science 366, 1236–1239.

Bickel, T., 2007. A note on confined diffusion. Physica A: Statistical Mechanics and its Applications 377, 24–32.

Blanco, M., Borgerson, C., Dolch, R., Donati, G., Ganzhorn, J., Greene, L., Le Pors, B., Lewis, R., Louis, E.E., Rafalinirina, H.A., Raharivololona, B., Rakotoarisoa, G., Ralison, J., Randriahaingo, H., Rasoloarison, R., Razafindrasolo, M., Sgarlata, G.M., Wright, P., Youssouf, J., Zaonarivelo, J., 2020a. IUCN Red List of Threatened Species: Cheirogaleus major. IUCN Red List of Threatened Species.

Blanco, M., Borgerson, C., Dolch, R., Ganzhorn, J., Greene, L., Le Pors, B., Lewis, R., Rafalinirina, H.A., Raharivololona, B., Rakotoarisoa, G., Ralison, J., Randriahaingo, H., Rasoloarison, R., Razafindrasolo, M., Sgarlata, G.M., Wright, P., Zaonarivelo, J., 2020b. IUCN Red List of Threatened Species: Microcebus simmonsi. IUCN Red List of Threatened Species.

Blanco, M., Dolch, R., Ganzhorn, J., Greene, L., Le Pors, B., Lewis, R., Louis, E.E., Radespiel, U., Rafalinirina, H.A., Raharivololona, B., Rakotoarisoa, G., Ralison, J., Randriahaingo, H., Rasoloarison, R., Razafindrasolo, M., Sgarlata, G.M., Wright, P., Zaonarivelo, J., 2020c. IUCN Red List of Threatened Species: Microcebus mittermeieri. IUCN Red List of Threatened Species.

Blanco, M., Dolch, R., Ganzhorn, J., Greene, L., Le Pors, B., Lewis, R., Radespiel, U., Rafalinirina, H.A., Raharivololona, B., Rakotoarisoa, G., Ralison, J., Randriahaingo, H., Rasoloarison, R., Razafindrasolo, M., Sgarlata, G.M., Wright, P., Zaonarivelo, J., 2020d. IUCN Red List of Threatened Species: Microcebus ravelobensis. IUCN Red List of Threatened Species.

Blanco, M., Dolch, R., Ganzhorn, J., Greene, L., Le Pors, B., Lewis, R., Radespiel, U., Rafalinirina, H.A., Raharivololona, B., Rakotoarisoa, G., Ralison, J., Randriahaingo, H., Rasoloarison, R., Razafindrasolo, M., Sgarlata, G.M., Wright, P., Zaonarivelo, J., 2020e. IUCN Red List of Threatened Species: Microcebus danfossi. IUCN Red List of Threatened Species.

Blanco, M., Dolch, R., Ganzhorn, J., Greene, L., Le Pors, B., Lewis, R., Radespiel, U., Rafalinirina, H.A., Raharivololona, B., Rakotoarisoa, G., Ralison, J., Randriahaingo, H., Rasoloarison, R., Razafindrasolo, M., Sgarlata, G.M., Wright, P., Zaonarivelo, J., 2020f. IUCN Red List of Threatened Species: Microcebus macarthurii. IUCN Red List of Threatened Species.

Blanco, M., Dolch, R., Ganzhorn, J., Greene, L., Le Pors, B., Lewis, R., Rafalinirina, H.A., Raharivololona, B., Rakotoarisoa, G., Ralison, J., Randriahaingo, H., Rasoloarison, R., Razafindrasolo, M., Sgarlata, G.M.,

Wright, P., Zaonarivelo, J., 2020g. IUCN Red List of Threatened Species: Microcebus mamiratra. IUCN Red List of Threatened Species.

Borgerson, C., Dolch, R., Ganzhorn, J., Greene, L., Guthrie, N.K., Johnson, S., Le Pors, B., Lewis, R., Louis, E.E., Rafalinirina, H.A., Raharivololona, B., Rakotoarisoa, G., Ralison, J., Randriahaingo, H., Rasoloarison, R., Raoelinjanakolona, N.N., Sgarlata, G.M., Schwitzer, C., Wright, P., Zaonarivelo, J., 2020. IUCN Red List of Threatened Species: Microcebus jollyae. IUCN Red List of Threatened Species.

Bradburd, G.S., Ralph, P.L., 2019. Spatial Population Genetics: It's About Time. Annual Review of Ecology, Evolution, and Systematics 50, 427–449.

Bueno, A.S., Peres, C.A., 2019. Patch-scale biodiversity retention in fragmented landscapes: Reconciling the habitat amount hypothesis with the island biogeography theory. Journal of Biogeography 46, 621–632.

Burney, D.A., 1987. Late Quaternary stratigraphic charcoal records from Madagascar. Quaternary Research 28, 274–280.

Burns, S.J., Godfrey, L.R., Faina, P., McGee, D., Hardt, B., Ranivoharimanana, L., Randrianasy, J., 2016. Rapid human-induced landscape transformation in Madagascar at the end of the first millennium of the Common Era. Quaternary Science Reviews 134, 92–99.

Chikhi, L., Rodríguez, W., Grusea, S., Santos, P., Boitard, S., Mazet, O., 2018. The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: insights into demographic inference and model choice. Heredity 120, 13–24.

Chikhi, L., Sgarlata, G.M., in press. Genetics of Malagasy Mammals: An Overview, in: The Natural History of Madagascar.

Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W., McCabe, A.M., 2009. The Last Glacial Maximum. Science 325, 710–714.

Costa, R.J., Wilkinson-Herbots, H., 2017. Inference of Gene Flow in the Process of Speciation: An Efficient Maximum-Likelihood Method for the Isolation-with-Initial-Migration Model. Genetics 205, 1597–1618.

Croft, D.A., Su, D.F., Simpson, S.W., 2018. Methods in Paleoecology: Reconstructing Cenozoic Terrestrial Environments and Ecological Communities. Springer.

Curtis, J.T., 1956. The modification of mid-latitude grasslands and forests by man., in: Man's Role in Changing the Face of the Earth. University of Chicago Press., Chicago, IL, pp. 721–736.

Cushman, S.A., Landguth, E.L., 2012. Multi-taxa population connectivity in the northern Rocky Mountains. Ecological Modelling. 231: 101-112. 101–112.

Cushman, S.A., Shirk, A., Landguth, E.L., 2012. Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex landscapes. Landscape Ecol 27, 369–380.

Díaz, S., Zafra-Calvo, N., Purvis, A., Verburg, P.H., Obura, D., Leadley, P., Chaplin-Kramer, R., Meester, L.D., Dulloo, E., Martín-López, B., Shaw, M.R., Visconti, P., Broadgate, W., Bruford, M.W., Burgess, N.D., Cavender-Bares, J., DeClerck, F., Fernández-Palacios, J.M., Garibaldi, L.A., Hill, S.L.L., Isbell, F., Khoury, C.K., Krug, C.B., Liu, J., Maron, M., McGowan, P.J.K., Pereira,

H.M., Reyes-García, V., Rocha, J., Rondinini, C., Shannon, L., Shin, Y.-J., Snelgrove, P.V.R., Spehn, E.M., Strassburg, B., Subramanian, S.M., Tewksbury, J.J., Watson, J.E.M., Zanne, A.E., 2020. Set ambitious goals for biodiversity and sustainability. Science.

Domic, A.I., Hixon, S.W., Velez, M.I., Ivory, S.J., Douglass, K.G., Brenner, M., Curtis, J.H., Culleton, B.J., Kennett, D.J., 2021. Influence of Late Holocene Climate Change and Human Land Use on Terrestrial and Aquatic Ecosystems in Southwest Madagascar. Frontiers in Ecology and Evolution 9, 833.

Duforet-Frebourg, N., Slatkin, M., 2016. Isolation-by-distance-and-time in a stepping-stone model. Theoretical Population Biology 108, 24–35.

Edmonds, C.A., Lillie, A.S., Cavalli-Sforza, L.L., 2004. Mutations arising in the wave front of an expanding population. PNAS 101, 975–979.

Everson, K.M., Jansa, S.A., Goodman, S.M., Olson, L.E., 2020. Montane regions shape patterns of diversification in small mammals and reptiles from Madagascar's moist evergreen forest. Journal of Biogeography 47, 2059–2072.

Fahrig, L., 2019. Habitat fragmentation: A long and tangled tale. Global Ecology and Biogeography 28, 33–41.

Fahrig, L., 2017. Ecological Responses to Habitat Fragmentation Per Se. Annual Review of Ecology, Evolution, and Systematics 48, 1–23.

Fahrig, L., 2003. Effects of Habitat Fragmentation on Biodiversity. Annual Review of Ecology, Evolution, and Systematics 34, 487–515.

Fahrig, L., Arroyo-Rodríguez, V., Bennett, J.R., Boucher-Lalonde, V., Cazetta, E., Currie, D.J., Eigenbrod, F., Ford, A.T., Harrison, S.P., Jaeger, J.A.G., Koper, N., Martin, A.E., Martin, J.-L., Metzger, J.P., Morrison, P., Rhodes, J.R., Saunders, D.A., Simberloff, D., Smith, A.C., Tischendorf, L., Vellend, M., Watling, J.I., 2019. Is habitat fragmentation bad for biodiversity? Biological Conservation 230, 179–186.

Fletcher, R.J., 2005. Multiple edge effects and their implications in fragmented landscapes. Journal of Animal Ecology 74, 342–352.

Fletcher, R.J., Didham, R.K., Banks-Leite, C., Barlow, J., Ewers, R.M., Rosindell, J., Holt, R.D., Gonzalez, A., Pardini, R., Damschen, E.I., Melo, F.P.L., Ries, L., Prevedello, J.A., Tscharntke, T., Laurance, W.F., Lovejoy, T., Haddad, N.M., 2018. Is habitat fragmentation good for biodiversity? Biological Conservation 226, 9–15.

Fordham, D.A., Brook, B.W., Moritz, C., Nogués-Bravo, D., 2014. Better forecasts of range dynamics using genetic data. Trends in Ecology & Evolution 29, 436–443.

Fordham, D.A., Jackson, S.T., Brown, S.C., Huntley, B., Brook, B.W., Dahl-Jensen, D., Gilbert, M.T.P., Otto-Bliesner, B.L., Svensson, A., Theodoridis, S., Wilmshurst, J.M., Buettel, J.C., Canteri, E., McDowell, M., Orlando, L., Pilowsky, J., Rahbek, C., Nogues-Bravo, D., 2020. Using paleo-archives to safeguard biodiversity under climate change. Science.

Fourtune, L., Paz-Vinas, I., Loot, G., Prunier, J.G., Blanchet, S., 2016. Lessons from the fish: a multi-species analysis reveals common processes underlying similar species-genetic diversity correlations. Freshwater Biology 61, 1830–1845.

Ganzhorn, J., Blanco, M., Borgerson, C., Dolch, R., Donati, G., Greene, L., Le Pors, B., Lewis, R., Louis, E.E., Rafalinirina, H.A., Raharivololona, B., Rakotoarisoa, G., Ralison, J., Randriahaingo, H., Rasoloarison, R., Razafindrasolo, M., Sgarlata, G.M., Wright, P., Youssouf, J., Zaonarivelo, J., 2020. IUCN Red List of Threatened Species: Microcebus griseorufus. IUCN Red List of Threatened Species.

Gautier, L., Ranirison, P., Nusbaumer, L., Wohlhauser, S., 2006. Aperçu des massifs forestiers de la région Loky-Manambato, in: Inventaires de La Faune et de La Flore Du Nord de Madagascar Dans La Région Loky-Manambato, Analamerana et Andavakoera. Rech. Développement. Sér Sci Biol.

Godfrey, L.R., Jungers, W.L., Simons, E.L., Chatrath, P.S., Rakotosamimanana, B., 1999. Past and Present Distributions of Lemurs in Madagascar, in: Rakotosamimanana, B., Rasamimanana, H., Ganzhorn, J.U., Goodman, S.M. (Eds.), New Directions in Lemur Studies. Springer US, Boston, MA, pp. 19–53.

Goldstein, D.B., Chikhi, L., 2002. Human migrations and population structure: what we know and why it matters. Annu Rev Genomics Hum Genet 3, 129–152.

Gorham, E., Brush, G.S., Graumlich, L.J., Rosenzweig, M.L., Johnson, A.H., 2001. The value of paleoecology as an aid to monitoring ecosystems and landscapes, chiefly with reference to North America. Environ. Rev. 9, 99–126.

Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J.K., Muzzio, M., Rodriguez-Flores, J.L., Kenny, E.E., Gignoux, C.R., Maples, B.K., Guiblet, W., Dutil, J., Via, M., Sandoval, K., Bedoya, G., Oleksyk, T.K., Ruiz-Linares, A., Burchard, E.G., Martinez-Cruzado, J.C., Bustamante, C.D., 2013.

Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data. PLoS Genet 9, e1004023.

Grebenkov, D.S., Nguyen, D.V., Li, J.-R., 2014. Exploring diffusion across permeable barriers at high gradients. I. Narrow pulse approximation. J Magn Reson 248, 153–163.

Griffiths, R.C., Tavare, S., 1994. Sampling Theory for Neutral Alleles in a Varying Environment. Philosophical Transactions: Biological Sciences 344, 403–410.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., Bustamante, C.D., 2009. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. PLOS Genetics 5, e1000695.

Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D., Cook, W.M., Damschen, E.I., Ewers, R.M., Foster, B.L., Jenkins, C.N., King, A.J., Laurance, W.F., Levey, D.J., Margules, C.R., Melbourne, B.A., Nicholls, A.O., Orrock, J.L., Song, D.-X., Townshend, J.R., 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. Science Advances 1, e1500052.

Haddad, N.M., Gonzalez, A., Brudvig, L.A., Burt, M.A., Levey, D.J., Damschen, E.I., 2017. Experimental evidence does not support the Habitat Amount Hypothesis. Ecography 40, 48–55.

Hanski, I., Pakkala, T., Kuussaari, M., Lei, G., 1995. Metapopulation Persistence of an Endangered Butterfly in a Fragmented Landscape. Oikos 72, 21–28.

Hanski, I., Zurita, G.A., Bellocq, M.I., Rybicki, J., 2013. Species–fragmented area relationship. PNAS 110, 12715–12720.

Hending, D., Sgarlata, G.M., Le Pors, B., Rasolondraibe, E., Jan, F., Rakotonanahary, A.N., Ralantoharijaona, T.N., Debulois, S., Andrianiaina, A., Cotton, S., Rasoloharijaona, S., Zaonarivelo, J.R., Andriaholinirina, N.V., Chikhi, L., Salmona, J., 2020. Distribution and conservation status of the endangered Montagne d'Ambre fork-marked lemur (Phaner electromontis). J Mammal 101, 1049–1060.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907–913.

Hewitt, G.M., 1999. Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society 68, 87–112.

Hey, J., Nielsen, R., 2004. Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, With Applications to the Divergence of Drosophila pseudoobscura and D. persimilis. Genetics 167, 747–760.

Honnay, O., Jacquemyn, H., 2007. Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. Conserv Biol 21, 823–831.

Hu, F.S., Hampe, A., Petit, R.J., 2009. Paleoecology meets genetics: deciphering past vegetational dynamics. Frontiers in Ecology and the Environment 7, 371–379.

Hutchison, D.W., Templeton, A.R., 1999. Correlation of Pairwise Genetic and Geographic Distance Measures: Inferring the Relative Influences of Gene

Flow and Drift on the Distribution of Genetic Variability. Evolution 53, 1898–1914.

Ivory, S.J., Russell, J., 2018. Lowland forest collapse and early human impacts at the end of the African Humid Period at Lake Edward, equatorial East Africa. Quaternary Research 89, 7–20.

Jackson, N.D., Fahrig, L., 2016. Habitat amount, not habitat configuration, best predicts population genetic structure in fragmented landscapes. Landscape Ecol 31, 951–968.

Jackson, N.D., Fahrig, L., 2014. Landscape context affects genetic diversity at a much larger spatial extent than population abundance. Ecology 95, 871–881.

Jansa, S.A., Carleton, M.D., Soarimalala, V., Rakotomalala, Z., Goodman, S.M., 2019. A review of the Eliurus tanala complex (Rodentia, Muroidea, Nesomyidae), with description of a new species from dry forests of western Madagascar. (Bulletin of the American Museum of Natural History, no. 430).

Jungers, W.L., Godfrey, L.R., Simons, E.L., Chatrath, P.S., 1995. Subfossil Indri indri from the Ankarana Massif of northern Madagascar. Am J Phys Anthropol 97, 357–366.

Juřičková, L., Horáčková, J., Ložek, V., 2014. Direct evidence of central European forest refugia during the last glacial period based on mollusc fossils. Quaternary Research 82, 222–228.

Kardos, M., Armstrong, E.E., Fitzpatrick, S.W., Hauser, S., Hedrick, P.W., Miller, J.M., Tallmon, D.A., Funk, W.C., 2021. The crucial role of genome-wide genetic variation in conservation. PNAS 118.

Keyghobadi, N., Roland, J., Matter, S.F., Strobeck, C., 2005. Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation: an empirical demonstration. Proceedings of the Royal Society B: Biological Sciences 272, 553–560.

Kimura, M., Weiss, G.H., 1964. The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. Genetics 49, 561–576.

King, L., Wakeley, J., 2016. Empirical Bayes Estimation of Coalescence Times from Nucleotide Sequence Data. Genetics 204, 249–257.

Krummel, J.R., Gardner, R.H., Sugihara, G., O'Neill, R.V., Coleman, P.R., 1987. Landscape Patterns in a Disturbed Environment. Oikos 48, 321–324.

Kuhner, M.K., Yamato, J., Felsenstein, J., 1995. Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. Genetics 140, 1421–1430.

Kummel, B., Raup, D. (Eds.), 1965. Handbook of Paleontological Techniques, First Edition. ed. W.H. Freeman & Co.

Lattin, G. de, 1967. Grundriss der Zoogeographie. VEB Gustav Fischer Verlag, Jena.

Le Pors, B., Sgarlata, G.M., Salmona, J., Rasolondraibe, E., Ralantoharijaona, T., Randriamaroson, J., Rasoloharijaona, S., Andriaholinirina, N.V., Chikhi, L., 2020. Lemur inventory of the Analabe forest, SAVA region, and observation of an al-bino Eulemur. Lemur News 56–61.

Leblois, R., Estoup, A., Streiff, R., 2006. Genetics of recent habitat contraction and reduction in population size: does isolation by distance matter? Mol Ecol 15, 3601–3615.

Lino, A., Fonseca, C., Rojas, D., Fischer, E., Ramos Pereira, M.J., 2019. A meta-analysis of the effects of habitat loss and fragmentation on genetic diversity in mammals. Mammalian Biology 94, 69–76.

Liu, H., Prugnolle, F., Manica, A., Balloux, F., 2006. A Geographically Explicit Genetic Model of Worldwide Human-Settlement History. Am J Hum Genet 79, 230–237.

López-Duarte, P.C., Carson, H.S., Cook, G.S., Fodrie, F.J., Becker, B.J., DiBacco, C., Levin, L.A., 2012. What Controls Connectivity? An Empirical, Multi-Species Approach. Integrative and Comparative Biology 52, 511–524.

Malcolm, J., 1998. A model of conductive heat flow in forest edges and fragmented landscapes. Climatic Change 39, 487–502.

Malcolm, J.R., 2001. Extending models of edge effects to diverse landscape configurations, with a Neotropical test case, in: Lessons from Amazonia - The Ecology and Conservation of a Fragmented Forest. Yale University Press, New Haven, pp. 347–357.

Malcolm, J.R., 1994. Edge Effects in Central Amazonian Forest Fragments. Ecology 75, 2438–2445.

Malécot, G., 1948. Les mathématiques de l'hérédité. Barnéoud frères.

Mandelbrot, B.B., 1982. The Fractal Geometry of Nature, 2nd prt. edition. ed. Times Books, San Francisco.

Maruyama, T., 1972. Rate of Decrease of Genetic Variability in a Two-Dimensional Continuous Population of Finite Size. Genetics 70, 639–651.

Mazet, O., Rodríguez, W., Chikhi, L., 2015. Demographic inference using genetic data from a single individual: Separating population size variation from population structure. Theoretical Population Biology 104, 46–58.

Mazet, O., Rodríguez, W., Grusea, S., Boitard, S., Chikhi, L., 2016. On the importance of being structured: instantaneous coalescence rates and human evolution--lessons for ancestral population size inference? Heredity (Edinb) 116, 362–371.

Meirmans, P.G., 2012. The trouble with isolation by distance. Molecular Ecology 21, 2839–2846.

Mercier, J.-L., Wilmé, L., 2013. The Eco-Geo-Clim model: explaining Madagascar's endemism. Madagascar Conservation & Development 8, 63–68.

Milne, B.T., 1988. Measuring the fractal geometry of landscapes. Applied Mathematics and Computation 27, 67–79.

Miraldo, A., Li, S., Borregaard, M.K., Flórez-Rodríguez, A., Gopalakrishnan, S., Rizvanovic, M., Wang, Z., Rahbek, C., Marske, K.A., Nogués-Bravo, D., 2016. An Anthropocene map of genetic diversity. Science 353, 1532–1535.

Mona, S., Ray, N., Arenas, M., Excoffier, L., 2014. Genetic consequences of habitat fragmentation during a range expansion. Heredity 112, 291–299.

Monteiro, W.P., Veiga, J.C., Silva, A.R., Carvalho, C. da S., Lanes, É.C.M., Rico, Y., Jaffé, R., 2019. Everything you always wanted to know about gene flow in tropical landscapes (but were afraid to ask). PeerJ 7, e6446.

Moore, N.W., 1962. The Heaths of Dorset and their Conservation. Journal of Ecology 50, 369–391.

Mortensen, K.I., Flyvbjerg, H., Pedersen, J.N., 2021. Confined Brownian Motion Tracked With Motion Blur: Estimating Diffusion Coefficient and Size of Confining Space. Frontiers in Physics 8, 601.

Nichols, R., 1996. Is it ancient or modern history that we can read in the genes? pp. 69–87.

Nielsen, R., 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics 154, 931–942.

Nielsen, R., Wakeley, J., 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics 158, 885–896.

Nogués-Bravo, D., Rodríguez-Sánchez, F., Orsini, L., de Boer, E., Jansson, R., Morlon, H., Fordham, D.A., Jackson, S.T., 2018. Cracking the Code of Biodiversity Responses to Past Climate Change. Trends in Ecology & Evolution 33, 765–776.

Palamara, P.F., Pe'er, I., 2013. Inference of historical migration rates via haplotype sharing. Bioinformatics 29, i180–i188.

Pertoldi, C., Ruiz-Gonzalez, A., Bahrndorff, S., Renee Lauridsen, N., Nisbeth Henriksen, T., Eskildsen, A., Høye, T.T., 2021. Strong isolation by distance

among local populations of an endangered butterfly species (Euphydryas aurinia). Ecology and Evolution 11, 12790–12800.

Pfeifer, M., Lefebvre, V., Peres, C.A., Banks-Leite, C., Wearn, O.R., Marsh, C.J., Butchart, S.H.M., Arroyo-Rodríguez, V., Barlow, J., Cerezo, A., Cisneros, L., D'Cruze, N., Faria, D., Hadley, A., Harris, S.M., Klingbeil, B.T., Kormann, U., Lens, L., Medina-Rangel, G.F., Morante-Filho, J.C., Olivier, P., Peters, S.L., Pidgeon, A., Ribeiro, D.B., Scherber, C., Schneider-Maunoury, L., Struebig, M., Urbina-Cardona, N., Watling, J.I., Willig, M.R., Wood, E.M., Ewers, R.M., 2017. Creation of forest edges has a global impact on forest vertebrates. Nature 551, 187–191.

Prevedello, J.A., Vieira, M.V., 2010. Does the type of matrix matter? A quantitative review of the evidence. Biodivers Conserv 19, 1205–1223.

Quéméré, E., Amelot, X., Pierson, J., Crouau-Roy, B., Chikhi, L., 2012. Genetic data suggest a natural prehuman origin of open habitats in northern Madagascar and question the deforestation narrative in this region. PNAS 109, 13028–13033.

Quéméré, E., Crouau-Roy, B., Rabarivola, C., Jr, E.E.L., Chikhi, L., 2010. Landscape genetics of an endangered lemur (Propithecus tattersalli) within its entire fragmented range. Molecular Ecology 19, 1606–1621.

Rakotoarisoa, J.E., Raheriarisena, M., Goodman, S.M., 2013. Late Quaternary climatic vegetational shifts in an ecological transition zone of northern Madagascar: insights from genetic analyses of two endemic rodent species. Journal of Evolutionary Biology 26, 1019–1034.

Rakotoarisoa, J. E., Raheriarisena, M., Goodman, S.M., 2013. A phylogeographic study of the endemic rodent Eliurus carletoni (Rodentia:

Nesomyinae) in an ecological transition zone of Northern Madagascar. J. Hered. 104, 23–35.

Ralph, P., Coop, G., 2013. The Geography of Recent Genetic Ancestry across Europe. PLOS Biology 11, e1001555.

Rasolonjatovo, S.M., Scherz, M.D., Hutter, C.R., Glaw, F., Rakotoarison, A., Razafindraibe, J.H., Goodman, S.M., Raselimanana, A.P., Vences, M., 2020. Sympatric lineages in the Mantidactylus ambreensis complex of Malagasy frogs originated allopatrically rather than by in-situ speciation. Molecular Phylogenetics and Evolution 144, 106700.

Razafindramanana, J., Salmona, J., King, T., Roullet, D., Eppley, T., Sgarlata, G.M., Schwitzer, C., 2020. IUCN Red List of Threatened Species: Propithecus coronatus. IUCN Red List of Threatened Species.

Ries, L., Fletcher, R.J., Battin, J., Sisk, T.D., 2004. Ecological Responses to Habitat Edges: Mechanisms, Models, and Variability Explained. Annu. Rev. Ecol. Evol. Syst. 35, 491–522.

Ries, L., Murphy, S.M., Wimp, G.M., Fletcher, R.J., 2017. Closing Persistent Gaps in Knowledge About Edge Ecology. Curr Landscape Ecol Rep 2, 30–41.

Ringbauer, H., Coop, G., Barton, N.H., 2017. Inferring Recent Demography from Isolation by Distance of Long Shared Sequence Blocks. Genetics 205, 1335–1351.

Ringbauer, H., Kolesnikov, A., Field, D.L., Barton, N.H., 2018. Estimating Barriers to Gene Flow from Distorted Isolation-by-Distance Patterns. Genetics 208, 1231–1245.

Rivera Ortíz, F.A., Aguilar, R., Arizmendi, M.D.C., Quesada, M., Oyama, K., 2015. Habitat fragmentation and genetic variability of tetrapod populations.

Rodríguez, W., Mazet, O., Grusea, S., Arredondo, A., Corujo, J.M., Boitard, S., Chikhi, L., 2018. The IICR and the non-stationary structured coalescent: towards demographic inference with arbitrary changes in population structure. Heredity (Edinb) 121, 663–678.

Roser, L.G., Ferreyra, L.I., Ewens, M., Vilardi, J.C., Saidman, B.O., 2017. Isolation by distance and stochastic migration in a fragmented population of Prosopis alba. American Journal of Botany 104, 313–321.

Rousset, F., 2004. Genetic Structure and Selection in Subdivided Populations. Princeton University Press.

Rybicki, J., Hanski, I., 2013. Species–area relationships and extinctions caused by habitat loss and fragmentation. Ecology Letters 16, 27–38.

Sabater Pi, J., 1967. An albino lowland gorilla from Rio Muni, West Africa, and notes on its adaptation to captivity. Folia Primatol (Basel) 7, 155–160.

Salmona, J., Dresen, A., Ranaivoson, A.E., Manzi, S., Pors, B.L., Hong-Wa, C., Razanatsoa, J., Andriaholinirina, N.V., Rasoloharijaona, S., Vavitsara, M.-E., Besnard, G., 2021. How ancient forest fragmentation and riparian connectivity generate high levels of genetic diversity in a micro-endemic Malagasy tree.

Salmona, J., Heller, R., Quéméré, E., Chikhi, L., 2017. Climate change and human colonization triggered habitat loss and fragmentation in Madagascar. Mol. Ecol. 26, 5203–5222.

Salmona, J., Rasolondraibe, E., Jan, F., Rakotonanahary, A.N., Ralantoharijaona, T., Le Pors, B., Ousseni, D.S.A., Aleixo-Pais, I., Marques, A.J., Sgarlata, G.M., Teixeira, H., Gabillaud, V., Miller, A., Ibouroi, M.T., Zaonarivelo, J.R., Andriaholinirina, N., Chikhi, L., 2018. Re-discovering the forgotten Phaner population of the small and isolated Analafiana forest (Vohémar, SAVA). Lemur News 31–36.

Sandel, B., Arge, L., Dalsgaard, B., Davies, R.G., Gaston, K.J., Sutherland, W.J., Svenning, J.-C., 2011. The Influence of Late Quaternary Climate-Change Velocity on Species Endemism. Science 334, 660–664.

Schlaepfer, D.R., Braschler, B., Rusterholz, H.-P., Baur, B., 2018. Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. Ecosphere 9, e02488.

Schneider, N., Chikhi, L., Currat, M., Radespiel, U., 2010. Signals of recent spatial expansions in the grey mouse lemur (Microcebus murinus). BMC Evolutionary Biology 10, 105.

Sexton, J.P., Hangartner, S.B., Hoffmann, A.A., 2014. Genetic Isolation by Environment or Distance: Which Pattern of Gene Flow Is Most Common? Evolution 68, 1–15.

Sgarlata, G.M., Le Pors, B., Blanco, M., Salmona, J., Chikhi, L., Louis, E.E.Jr., Cotton, S., Hending, D., Dolch, R., Ganzhorn, J., Greene, L., Lewis, R., Rafalinirina, H.A., Raharivololona, B., Rakotoarisoa, G., Ralison, J., Randriahaingo, H., Rasoloarison, R., Razafindrasolo, M., Wright, P., Zaonarivelo, J., 2020a. IUCN Red List of Threatened Species: Microcebus tavaratra. IUCN Red List of Threatened Species.

Sgarlata, G.M., Le Pors, B., Blanco, M., Salmona, J., Chikhi, L., Louis, E.E.Jr., Cotton, S., Hending, D., Frasier, C.L., Dolch, R., Ganzhorn, J., Greene, L., Lewis, R., Rafalinirina, H.A., Raharivololona, B., Rakotoarisoa, G., Ralison, J., Randriahaingo, H., Rasoloarison, R., Razafindrasolo, M., Wright, P., Zaonarivelo, J., 2020b. IUCN Red List of Threatened Species:Cheirogaleus shethi. IUCN Red List of Threatened Species.

Sgarlata, G.M., Le Pors, B., Blanco, M., Salmona, J., Chikhi, L., Louis, E.E.Jr., Dolch, R., Ganzhorn, J., Greene, L., Lewis, R., Rafalinirina, H.A., Raharivololona, B., Rakotoarisoa, G., Ralison, J., Randriahaingo, H., Rasoloarison, R., Razafindrasolo, M., Wright, P., Zaonarivelo, J., 2020c. IUCN Red List of Threatened Species: Microcebus arnholdi. IUCN Red List of Threatened Species.

Sgarlata, G.M., Le Pors, B., Salmona, J., Hending, D., Chikhi, L., Cotton, S., 2020d. IUCN Red List of Threatened Species: Phaner electromontis. IUCN Red List of Threatened Species.

Sgarlata, G.M., Salmona, J., Aleixo-Pais, I., Rakotonanahary, A., Sousa, A.P., Kun-Rodrigues, C., Ralantoharijaona, T., Jan, F., Zaranaina, R., Rasolondraibe, E., Zaonarivelo, J.R., Andriaholinirina, N.V., Chikhi, L., 2018. Genetic Differentiation and Demographic History of the Northern Rufous Mouse Lemur (Microcebus tavaratra) Across a Fragmented Landscape in Northern Madagascar. Int J Primatol 39, 65–89.

Sgarlata, G.M., Salmona, J., Pors, B.L., Rasolondraibe, E., Jan, F., Ralantoharijaona, T., Rakotonanahary, A., Randriamaroson, J., Marques, A.J., Aleixo-Pais, I., Zoeten, T. de, Ousseni, D.S.A., Knoop, S.B., Teixeira, H., Gabillaud, V., Miller, A., Ibouroi, M.T., Rasoloharijaona, S., Zaonarivelo, J.R., Andriaholinirina, N.V., Chikhi, L., 2019. Genetic and morphological diversity of mouse lemurs (Microcebus spp.) in northern Madagascar: The

discovery of a putative new species? American Journal of Primatology 81, e23070.

Sgarlata, G.M., Salmona, J., Razanaparany, T.P., Rabarivola, C., Jan, F., Rasolondraibe, E., Andriaholinirina, N.V., Lafosse, S., Chikhi, L., Manni, F., Bayart, F., 2016. Mitochondrial genetic diversity in the crowned sifaka (Propithecus coronatus) in a fragmented landscape. Primate Conservation 39–57.

Shi, J.J., Chan, L.M., Rakotomalala, Z., Heilman, A.M., Goodman, S.M., Yoder, A.D., 2013. Latitude drives diversification in Madagascar's endemic dry forest rodent Eliurus myoxinus (subfamily Nesomyinae). Biological Journal of the Linnean Society 110, 500–517.

Simons, E.L., Burney, D.A., Chatrath, P.S., Godfrey, L.R., Jungers, W.L., Rakotosamimanana, B., 1995. AMS 14C Dates for Extinct Lemurs from Caves in the Ankarana Massif, Northern Madagascar. Quaternary Research 43, 249–254.

Slatkin, M., 1993. Isolation by Distance in Equilibrium and Non-Equilibrium Populations. Evolution 47, 264–279.

Slatkin, M., 1991. Inbreeding coefficients and coalescence times. Genetics Research 58, 167–175.

Slatkin, M., Barton, N.H., 1989. A COMPARISON OF THREE INDIRECT METHODS FOR ESTIMATING AVERAGE LEVELS OF GENE FLOW. Evolution 43, 1349–1368.

Slatkin, M., Excoffier, L., 2012. Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. Genetics 191, 171–181.

Solé, R.V., Manrubia, S.C., 1995. Are rainforests self-organized in a critical state? Journal of Theoretical Biology 173, 31–40.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G., Cosson, J.-F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology 7, 453–464.

Tajima, F., 1990. Relationship between migration and DNA polymorphism in a local population. Genetics 126, 231–234.

Tang, H., Siegmund, D.O., Shen, P., Oefner, P.J., Feldman, M.W., 2002. Frequentist estimation of coalescence times from nucleotide sequence data using a tree-based partition. Genetics 161, 447–459.

Teixeira, H., Montade, V., Salmona, J., Metzger, J., Bremond, L., Kasper, T., Daut, G., Rouland, S., Ranarilalatiana, S., Rakotondravony, R., Chikhi, L., Behling, H., Radespiel, U., 2021. Past environmental changes affected lemur population dynamics prior to human impact in Madagascar. Commun Biol 4, 1–10.

Thompson, E.A., 2013. Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations. Genetics 194, 301–326.

Vieilledent, G., Grinand, C., Rakotomalala, F.A., Ranaivosoa, R., Rakotoarijaona, J.-R., Allnutt, T.F., Achard, F., 2018. Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. Biological Conservation 222, 189–197.

Voarintsoa, N.R.G., Matero, I.S.O., Railsback, L.B., Gregoire, L.J., Tindall, J., Sime, L., Cheng, H., Edwards, R.L., Brook, G.A., Kathayat, G., Li, X., Rakotondrazafy, A.F.M., Razanatseheno, M.O.M., 2019. Investigating the 8.2 ka event in northwestern Madagascar: Insight from data–model comparisons. Quaternary Science Reviews 204, 172.

Voarintsoa, N.R.G., Ratovonanahary, A.L.J., Rakotovao, A.Z.M., Bouillon, S., 2021. Understanding the linkage between regional climatology and cave geochemical parameters to calibrate speleothem proxies in Madagascar. Science of The Total Environment 784, 147181.

Voarintsoa, N.R.G., Wang, L., Railsback, L.B., Brook, G.A., Liang, F., Cheng, H., Edwards, R.L., 2017. Multiple proxy analyses of a U/Th-dated stalagmite to reconstruct paleoenvironmental changes in northwestern Madagascar between 370 CE and 1300 CE. Palaeogeography, Palaeoclimatology, Palaeoecology 469, 138–155.

Vranckx, G., Jacquemyn, H., Muys, B., Honnay, O., 2012. Meta-analysis of susceptibility of woody plants to loss of genetic diversity through habitat fragmentation. Conserv Biol 26, 228–237.

Wakeley, J., 1999. Nonequilibrium Migration in Human History. Genetics 153, 1863–1871.

Wang, K., Mathieson, I., O'Connell, J., Schiffels, S., 2020. Tracking human population structure through time from whole genome sequences. PLOS Genetics 16, e1008552.

Watling, J.I., Arroyo-Rodríguez, V., Pfeifer, M., Baeten, L., Banks-Leite, C., Cisneros, L.M., Fang, R., Hamel-Leigue, A.C., Lachat, T., Leal, I.R., Lens, L., Possingham, H.P., Raheem, D.C., Ribeiro, D.B., Slade, E.M., Urbina-

Cardona, J.N., Wood, E.M., Fahrig, L., 2020. Support for the habitat amount hypothesis from a global synthesis of species density studies. Ecology Letters 23, 674–681.

Wegmann, D., Currat, M., Excoffier, L., 2006. Molecular Diversity After a Range Expansion in Heterogeneous Environments. Genetics 174, 2009–2020.

Weigelt, P., Steinbauer, M.J., Cabral, J.S., Kreft, H., 2016. Late Quaternary climate change shapes island biodiversity. Nature 532, 99–102.

Wilkins, J.F., 2004. A Separation-of-Timescales Approach to the Coalescent in a Continuous Population. Genetics 168, 2227–2244.

Wilkins, J.F., Wakeley, J., 2002. The Coalescent in a Continuous, Finite, Linear Population. Genetics 161, 873–888.

Wilmé, L., Goodman, S.M., Ganzhorn, J.U., 2006. Biogeographic Evolution of Madagascar's Microendemic Biota. Science 312, 1063–1065.

Wilmé, L., Ravokatra, M., Dolch, R., Schuurman, D., Mathieu, É., Schuetz, H., Waeber, P., 2012. Toponyms for centers of endemism in Madagascar.

Wright, S., 1943. Isolation by Distance. Genetics 28, 114–138.

Zhao, Y., Vrieling, K., Liao, H., Xiao, M., Zhu, Y., Rong, J., Zhang, W., Wang, Y., Yang, J., Chen, J., Song, Z., 2013. Are habitat fragmentation, local adaptation and isolation-by-distance driving population divergence in wild rice Oryza rufipogon? Mol Ecol 22, 5531–5547.

7 Annex

7.1 Genetic Differentiation and Demographic History of the Northern Rufous Mouse Lemur (*Microcebus tavaratra*) Across a Fragmented Landscape in Northern Madagascar

Gabriele Maria Sgarlata[1], Jordi Salmona[1,2,3], Isa Aleixo-Pais[1,4,5], Ando Rakotonanahary[6], Ana Priscila Sousa[1], Célia Kun-Rodrigues[1], Tantely Ralantoharijaona[6], Fabien Jan[1], Radavison Zaranaina[6], Emmanuel Rasolondraibe[6], John Rigobert Zaonarivelo[7], Nicole Volasoa Andriaholinirina[6], Lounès Chikhi[1,2,3]

1- Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, P-2780-156 Oeiras, Portugal
2- CNRS, Université Paul Sabatier, ENFA; UMR 5174 EDB (Laboratoire Évolution & Diversité Biologique), 118 route de Narbonne, F-31062 Toulouse, France
3- Université Toulouse 3 Paul Sabatier, CNRS; UMR 5174 EDB, F-31062 Toulouse, France
4- Organisms and Environment Division, Cardiff School of Biosciences, Cardiff University, CF10 3AX, Cardiff, UK
5- Centre for Research in Anthropology (CRIA-FCSH/NOVA), 1069-061 Lisbon, Portugal
6- Faculté des Sciences, Université de Mahajanga, BP 652 401, Mahanjanga, Madagascar
7- Département des Sciences de la Nature et de l'Environnement, Université d'Antsiranana, 201 Antsiranana, Madagascar

Keywords Geneticstructure·Habitat fragmentation·Isolation by distance·Population genetics · Small mammals

Abstract

Phylogeographic barriers, together with habitat loss and fragmentation, contribute to the evolution of a species' genetic diversity by limiting gene flow and increasing genetic differentiation among populations. Changes in connectivity can thus affect the genetic diversity of populations, which may influence the evolutionary potential of species and the survival of populations in the long term. We studied the genetic diversity of the little known Northern rufous mouse lemur (*Microcebus tavaratra*), endemic to Northern Madagascar. We focused on the population of *M. tavaratra* in the Loky–Manambato region, Northern Madagascar, a region delimited by two permanent rivers and characterized by a mosaic of fragmented forests. We genotyped 148 individuals at three mitochondrial loci (D- loop, cytb, and cox2) in all the major forests of the study region. Our analyses suggest that *M. tavaratra* holds average genetic diversity when compared to other mouse lemur species, and we identified two to four genetic clusters in the study region, a pattern similar to that observed in another lemur endemic to the region (*Propithecus tattersalli*). The main cluster involved samples from the two mountain forests in the study region, which were connected until recently. However, the river crossing the study region does not appear to be a strict barrier to gene flow in *M. tavaratra*. Finally, the inferred demographic history of *M. tavaratra* suggests no detectable departure from stationarity over the last millennia. Comparisons with codistributed species (*P. tattersalli* and two endemic rodents, *Eliurus* spp.) suggest both differences and similarities in the genetic clusters identified (i.e., barriers to species dispersal) and in the inferred demographic history. These comparisons suggest that studies of codistributed species are important to understand the effects of landscape features on species and to reconstruct the history of habitat changes in a region.

7.2 The unique and isolated *Phaner* population of Analafiana (Vohémar, SAVA)

Jordi Salmona[1,2,3], Emmanuel Rasolondraibe[4], Fabien Jan[1], Ando Nantenaina Rakotonanahary[4], Tantely Ralantoharijaona[4], Barbara Le Pors[1], Dhurham Said Ali Ousseni[4], Isa Aleixo-Pais[1,8,9], Adam Joseph Doncheff Marques[1], Gabriele Maria Sgarlata[1], Helena Teixeira[1,5], Vivien Gabillaud[1], Alexandra Miller[1,6], Mohamed Thani Ibouroi[4,10,11], John Rigobert Zaonarivelo[9], Nicole Volasoa Andriaholinirina[4], Lounès Chikhi[1,2,3]

1 Instituto Gulbenkian de Ciênca, Rua da Quinta Grande, 6, P-2780-156 Oeiras, Portugal;
2 CNRS, Université Paul Sabatier, IRD ; UMR5174 EDB (Laboratoire Évolution & Diversité Biologique) ; 118 route de Narbonne, F-31062 Toulouse, France.
3 Université Toulouse 3 Paul Sabatier, CNRS ; UMR5174 EDB, F-31062 Toulouse, France
4 Faculté des Sciences, de Technologie et de l'Environnement, Université de Mahajanga, BP 652, (401) Mahajanga, Madagascar.
5 Institute of Zoology, University of Veterinary Medicine Hanover, Hanover, Germany
6 School of Human Sciences, The University of Western Australia, Crawley, WA 6009
7 EPHE, PSL Research University, CNRS, UM, SupAgro, IRD, INRA, UMR 5175 CEFE, F-34293 Montpellier, FRANCE
8 Organisms and Environment Division, Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3AX, United Kingdom
9 Centre for Research in Anthropology (CRIA-FCSH/NOVA), 1069-061, Lisbon, Portugal
10 Institut des Sciences et Industrie du Vivant et de l'environnement, AgroParisTech, 19 Avenue de Maine 75015 Paris, France
11 Département des Sciences de la Nature et de l'Environnement, Université d'Antsiranana, 201 Antsiranana, Madagascar.

Abstract

Despite the fact that lemurs are the most threatened group of mammals and live in one of the world's richest biodiversity hotspots, there are still lemur lineages and regions of Madagascar that suffer from a lack of naturalistic data. This is particularly striking for the genus *Phaner* as well as for many unprotected forests of Northeastern Madagascar. We detected the presence of *Phaner* in the isolated lowland dry forest of Analafiana, located south of the Manambato River and close to the city of Vohémar. Little is known about the presence of *Phaner* in this region. The only record of *Phaner* in the area around Analafiana comes from museum specimens (Groves and Tattersall 1991) geographically assigned to Vohémar and identified as belonging to a "*doubtful subspecies*". Here we provide morphological and genetic (mitochondrial DNA) data for a single juvenile individual captured in Analafiana. We compare our sequence data to the only four *Phaner* sequences publicly available: two *P. electromontis* individuals sampled in Montagne d'Ambre and Ankarana, and two other from *P. pallescens* and *P. parienti* species. We find that the new mtDNA sequence is divergent from the two *P. electromontis* sequences, but closer to these than other *Phaner* sequences. We also note, that the Analafiana *Phaner* individual is morphologically more similar to *P. electromontis* than to the other species. Altogether, this data suggests this unique and isolated population of *Phaner* survives in the unprotected Analafiana forest in northeastern Madagascar. It is urgent to carry out more genetic, behavioral, acoustic, and demographic research in this region and on this population, whose exact taxonomic status should be urgently clarified. Our results strongly suggest that Analafiana's *Phaner* population likely holds the memory of a unique evolutionary history.

7.3 The genetic structure of a mouse lemur living in a fragmented habitat in Northern Madagascar

Isa Aleixo-Pais[1,4,5], Jordi Salmona[1,2,3], Gabriele Maria Sgarlata[1], Ando Rakotonanahary[7], Ana Priscilla Sousa[1], Bárbara Parreira[1], Célia Kun-Rodrigues[1], Tantely Ralantoharijaona[7], Fabien Jan[1], Emmanuel Rasolondraibe[7], Tânia Minhós[1,5,6], John Rigobert Zaonarivelo[8], Nicole Volasoa Andriaholinirina[2], Lounès Chikhi[1,2,3]

1- Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal
2- Laboratoire Évolution & Diversité Biologique (EDB UMR 5174), Université de Toulouse Midi-Pyrénées, CNRS, IRD, UPS, 31062 Toulouse, France EDB, 31062 Toulouse, France
3- Université Toulouse 3 Paul Sabatier, CNRS; UMR 5174
4- Organisms and Environment Division, Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3AX, UK
5- Centre for Research in Anthropology (CRIA-FCSH/NOVA), 1069-061 Lisbon, Portugal
6- Department of Anthropology, School of Social Sciences and Humanities, Universidade Nova de Lisboa, 1069-061 Lisbon, Portugal
7- Faculté des Sciences, Université de Mahajanga, BP 652, 401 Mahajanga, Madagascar
8- Département des Sciences de la Nature et de l'Environnement, Université d'Antsiranana, 201 Antsiranana, Madagascar

Abstract

Habitat loss and fragmentation are major threats to biodiversity worldwide. Madagascar is among the top biodiversity hotspots and in the past 100 years several species became endangered on the island as a consequence of anthropogenic activities. In this study, we assessed the levels of genetic diversity and variation of a population of mouse lemurs (*Microcebus tavaratra*) inhabiting the degraded forests of the Loky-Manambato region (Northern Madagascar). We used a panel of 15 microsatellite markers to genotype 149 individuals. Our aim was to understand if the elements contributing to the heterogeneity of the landscape, such as forest fragmentation, roads, rivers and open habitat, influence the genetic structure of this population. The results showed that geographic distance along with open habitat, vegetation type and, to some extent, the Manankolana River, seem to be the main factors responsible for *M. tavaratra* population structure in this region. We found that this species still maintains substantial levels of genetic diversity within each forest patch and at the overall population, with low genetic differentiation observed between patches. This seems to suggest that the still existing riparian forest network connecting the different forest patches in this region, facilitates dispersal and maintains high levels of gene flow. We highlight that special efforts targeting riparian forest maintenance and reforestation might be a good strategy to reduce the effect of habitat frag- mentation on the genetic diversity of extant *M. tavaratra* populations.

7.4 Lemur inventory of the Analabe forest, SAVA region, and observation of an albino *Eulemur*

Barbara Le Pors[1*], Gabriele Maria Sgarlata[1], Jordi Salmona[2], Emmanuel Rasolondraibe[3], Tantely Ralantoharijaona[3], Jacquis Randriamaroson[3], Solofonirina Rasoloharijaona[3], Nicole Volasoa Andriaholinirina[3], Lounes Chikhi[1,2]

1-Instituto Gulbenkian de Ciência, rua da Quinta Grande, Oeiras, Portugal

2-Laboratoire Évolution & Diversité Biologique (EDB UMR 5174), Université de Toulouse Midi-Pyrénées, CNRS, IRD, UPS, 118 route de Narbonne, Bât. 4R1, 31062 Toulouse cedex 9, France

3-Faculté des Sciences Technologie et de l'Environnement, Université de Mahajanga, Mahajanga, Madagascar

Abstract

Analabe is a fragment (1064ha) of a dry forest located in northern Madagascar where we observed five lemur species and an adult albino *Eulemur* holding an infant accompanied by a male crowned lemur (*Eulemur coronatus*) in that forest. These observations suggest that small forests are still able to harbour diversity in Madagascar today. However, even when they are located close to larger protected forests, they may suffer from increasing isolation. We thus urge the community to increase the effort to identify and survey isolated forests before their unique diversity is gone.

7.5 Distribution and conservation status of the Endangered Montagne d'Ambre fork-marked lemur (*Phaner electromontis*)

Daniel Hending[1,2,#], Gabriele Maria Sgarlata[3,#], Barbara Le Pors[3], Emmanuel Rasolondraibe[4], Fabien Jan[3], Ando Rakotonanahary[4], Tantely Ralantoharijaona[4], Stéphane Debulois[5], Angelo Andrianiaina[6], Sam Cotton[2], Solofonirina Rasoloharijaona[4], John Rigobert Zaonarivelo[7], Nicole Volasoa Andriaholinirina[4], Lounès Chikhi[3,5,†], Jordi Salmona[3,5,†]

1- School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, United Kingdom (DH)
2- Department of Field Conservation and Science, Bristol Zoological Society, Bristol BS8 3HA, United Kingdom (DH, SC)
3- Instituto Gulbenkian de Ciênca, P-2780-156 Oeiras, Portugal (GS, BLP, FJ, LC, JS)
4- Faculté des Sciences, Université de Mahajanga, BP 652 401, Mahajanga, Madagascar (ER, AR, TR, SR, NA)
5- CNRS, Université Paul Sabatier, IRD; UMR5174 EDB (Laboratoire Évolution & Diversité Biologique), F-31062 Toulouse, France (SD, LC, JS)
6- Mention: Zoologie et Biodiversité Animale, Université d'Antananarivo, BP 906 Antananarivo 101, Madagascar (AA)
7- Département des Sciences de la Nature et de l'Environnement, Université d'Antsiranana, 201 Antsiranana, Madagascar (JZ)

[#]These two authors contributed equally [†]Shared equally in directing the research

Abstract

The geographic distribution of a species can provide insights into its population size, ecology, evolution, and how it responded to past (and may respond to future) environmental change. Improving our knowledge of the distribution of threatened species thus is a high priority in assessing their conservation status. However, there are few data available for many recently described yet understudied and potentially threatened primate taxa, making their conservation difficult. Here, we investigated the distribution of the Montagne d'Ambre fork-marked lemur, *Phaner electromontis*, a threatened nocturnal primate endemic to northern Madagascar and classified as Endangered by the IUCN. Because fork-marked lemurs are highly vocal, we used acoustic surveys to assess the species' presence- absence and relative population density within 66 distinct forest survey sites in northern Madagascar. Further, we compared data among five forest types within the study area and investigated the relationship between relative population density and climate variables. We report the presence of *P. electromontis* in 22 study sites; several of these populations were unknown previously. Although we found *P. electromontis* most frequently in dry- transitional forests, our results suggest that geography (spatial autocorrelation) rather than environmental variables explains the species' distribution. We hypothesize that environmental unpredictability and gummivory, combined with the presence of several distinct *Phaner* species in the studied area, could explain the observed distribution.

7.6 Population density estimates of northern rufuous mouse lemur (*Microcebus tavaratra*) and Arnhold's mouse lemur (*Microcebus arnholdi*) across northern Madagascar

Gabriele Maria Sgarlata[1], Jordi Salmona[1,2], Barbara Le Pors[1], Emmanuel Rasolondraibe[3], Fabien Jan[1], Tantely Ralantoharijaona[3], Ando Rakotonanahary[3], Jacquis Randriamaroson[3], Solofonirina Rasoloharijaona[3], John Rigobert Zaonarivelo[4], Nicole Volasoa Andriaholinirina[3], Lounès Chikhi[1,2]

1. Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, 2780-156 Oeiras, Portugal;
2. Laboratoire Evolution & Diversité Biologique (EDB UMR 5174), CNRS, IRD, UPS, Université de Toulouse Midi-Pyrénées, Toulouse, France;
3. Département de Biologie Animale et Ecologie, Faculté des Sciences, Université de Mahajanga, Mahajanga, Madagascar;
4. Département des Sciences de la Nature et de l'Environnement, Université d'Antsiranana, 201 Antsiranana, Madagascar

Brief Introduction and Methods

Since 2010, we have visited more than 40 forest sites across northern Madagascar, gathering a unique and extensive collection of mouse lemur (*Microcebus spp.*) observations (Fig. 1 and Table 1). In the present study, we complement *Microcebus* presence data from previously known locations (Sgarlata et al. 2019 and references therein) with new data collected in areas that have been poorly studied.

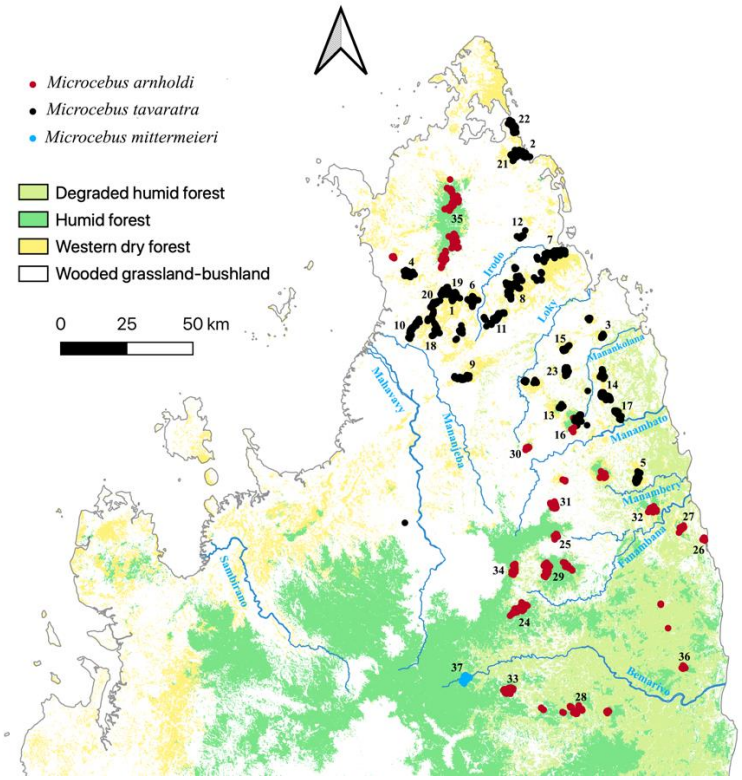


Figure 9. Distribution of *Microcebus spp.* observations across northern Madagascar between 2010 and 2018. Dots indicate locations of observation during line-transect distance sampling. Numbers represent location names, but refer to Table 1 for details on the visited locations.

In the visited sites, we carried out line transect distance sampling surveys (Buckland, 2001) to estimate population density of *Microcebus tavaratra*, *M. arnholdi* and *M. mittermeieri*. The surveyed transects were nearly straight and ranged between 600 m and ~1 km. Distance sampling surveys were conducted at night, between 18.30h and 21.30h, as it is the time in which generally *Microcebus* are active for feeding. In some cases, transects were travelled multiple times, waiting at least 24 hours to avoid disturbance of the resident animals and therefore limiting bias in population density estimates. The surveying teams were composed of two/three people and each time only one observer was recording animal sightings, therefore, reducing the visibility bias due to variable number of team members across years. During distance sampling survey, animals were sighted with the help of head torches and a more powerful hand-held torch to confirm the observation and genus identity (i.e., *Microcebus*). Because of the phenotypic crypticism of the *Microcebus* genus, species could not be recognized by simple observation, thus, information from previous genetic studies and genetic analyses on fewer captured individuals were used to claim the presence of one *Microcebus* species or another. Following the animal sighting, the team members were recording the number of observed individuals, perpendicular distance from the animal location to the transect with the aid of a digital laser rangefinder (Bosch® PLR 50; 50 m range; ± 2 mm accuracy), the time and the GPS position (Garmin®Etrex-H GPS). Since 2010, we have surveyed in total 912 transects for a total effort of 2449 km, in particular 254 transects (total effort = 387 km) for the *M. arnholdi* group and 582 transects (total effort = 1977 km) for *M. tavaratra*, and 17 transects (total effort = 16 km) for *M. mittermeieri* (found only in one of the surveyed sites). In one site (Binara forest), previous genetic data have shown the co-occurrence of two *Microcebus* species (*M. arnholdi* and *M. tavaratra*; Sgarlata et al. 2019), therefore population density values from this forest will refer to the *Microcebus* genus. In a few forests, no *Microcebus* observations were reported, although 37 transects were visited (total effort = 32 km). Details on

the transects length, total effort (km) and number of observations for each visited site are shown in Table 1.

We estimated site population densities using the conventional distance sampling (CDS) method implemented in the R package *Distance* (Miller et al., 2019). This method estimates population densities from the perpendicular distance of each animal sighting, the total length of the travelled transects and the effective strip-width (ESW). It models the visibility performance of an observer (probability of detection) by fitting a detection function to the observed distribution of recorded distances, and thus, infer the number of un-observed animals due to limitations in human visibility. We tested three detection functions (*hazard-rate*, *half-normal* and *uniform*), adding also adjustments terms (*cosine*, *Hermite polynomial* and *simple polynomial*). We truncated our datasets using a 5% threshold, that is removing the 5% more distant observations, to limit the impact of faraway observations on the fitting of the detection function. Model selection was performed using the AIC criteria and assessing goodness of fit with Cramér-von Mises test.

Results

During line-transect distance sampling surveys we observed in total 2053 *Microcebus* individuals, of which 1638 *M. tavaratra*, 295 *M. arnholdi*, 88 *Microcebus* sp. and 11 *M. mittermeieri* (Tab. 1). We covered in total 2449 km, corresponding to 912 transects. We obtained more than 30 recorded individuals for 23 sites, out of 40 surveyed forests. In order to compute accurate density estimates and pool data years, we made the assumption of no temporal variation in population density within the time range of the present study (2010 - 2018). In Fig. 2, Fig. 3 and Table 1, we show the selected detection function and the goodness-of-fit assessment for each of the surveyed sites. The Cravér-Mises test suggests that, within each surveyed

site, fitted data were not significantly different from the empirical distribution of animal sightings. In *M. tavaratra*, population density estimates ranged between 49 ind/km^2 (Andavakoera) to 396 ind/km^2 (Ambondromifehy), with a median value of 191 ind/km^2. In *M. arnholdi*, population density estimates ranged between 108 ind/km^2 (Ankijanala) to 376 ind/km^2 (Bezavona_Ankirendrina), with a median value of 179 ind/km^2. The comparison of population density values between the dry-associated mouse lemur (*M. tavaratra*) and the humid-associated (*M. arnholdi*) suggest that i) there are no major differences in population density between the two species; and ii) that contrary to previous studies (Setash et al., 2017), we did not find higher population density in species associated to western-dry habitat compared to eastern-humid habitat types.

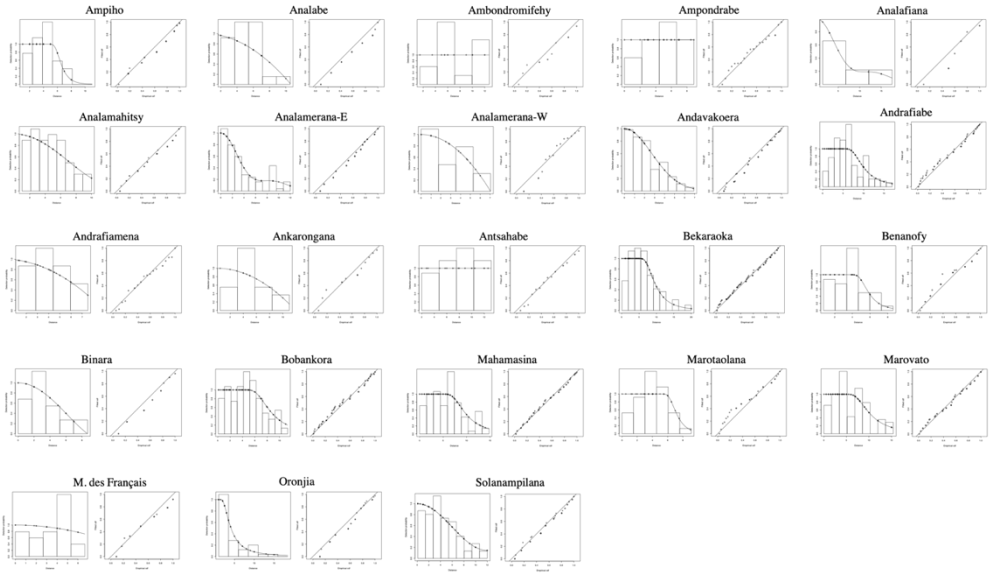


Figure 10. Detection probability and goodness-of-fit QQ plot for each surveyed site in *Microcebus tavaratra* distribution. Left panel: it is represented the fitted detection function overlayed on the histogram of observed distances. x-axis: perpendicular distance of animal sighting with respect to the transect. y-axis: number of observed animals at each distance bin. Right panel: the comparison of quantile-quantile plots for the selected detection function. X-axis: empirical cumulative distribution function. Y-axis: fitted cumulative distribution function.

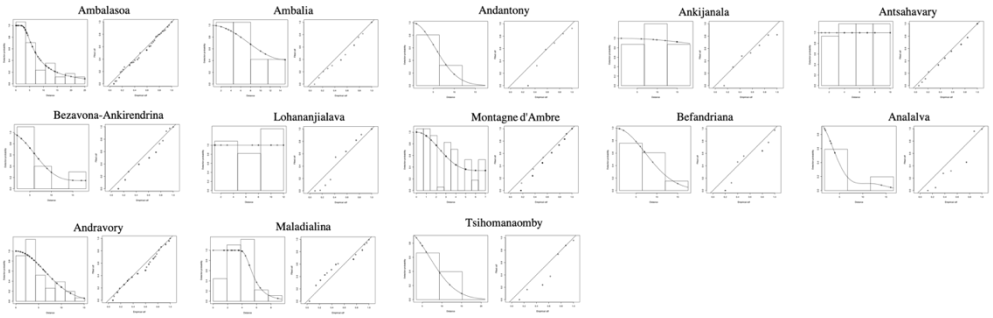


Figure 11. Detection probability and goodness-of-fit QQ plot for each surveyed site in *Microcebus arnholdi* distribution. Left panel: it is represented the fitted detection function overlayed on the histogram of observed distances. x-axis: perpendicular distance of animal sighting with respect to the transect. y-axis: number of observed animals at each distance bin. Right panel: the comparison of quantile-quantile plots for the selected detection function. X-axis: empirical cumulative distribution function. Y-axis: fitted cumulative distribution function.

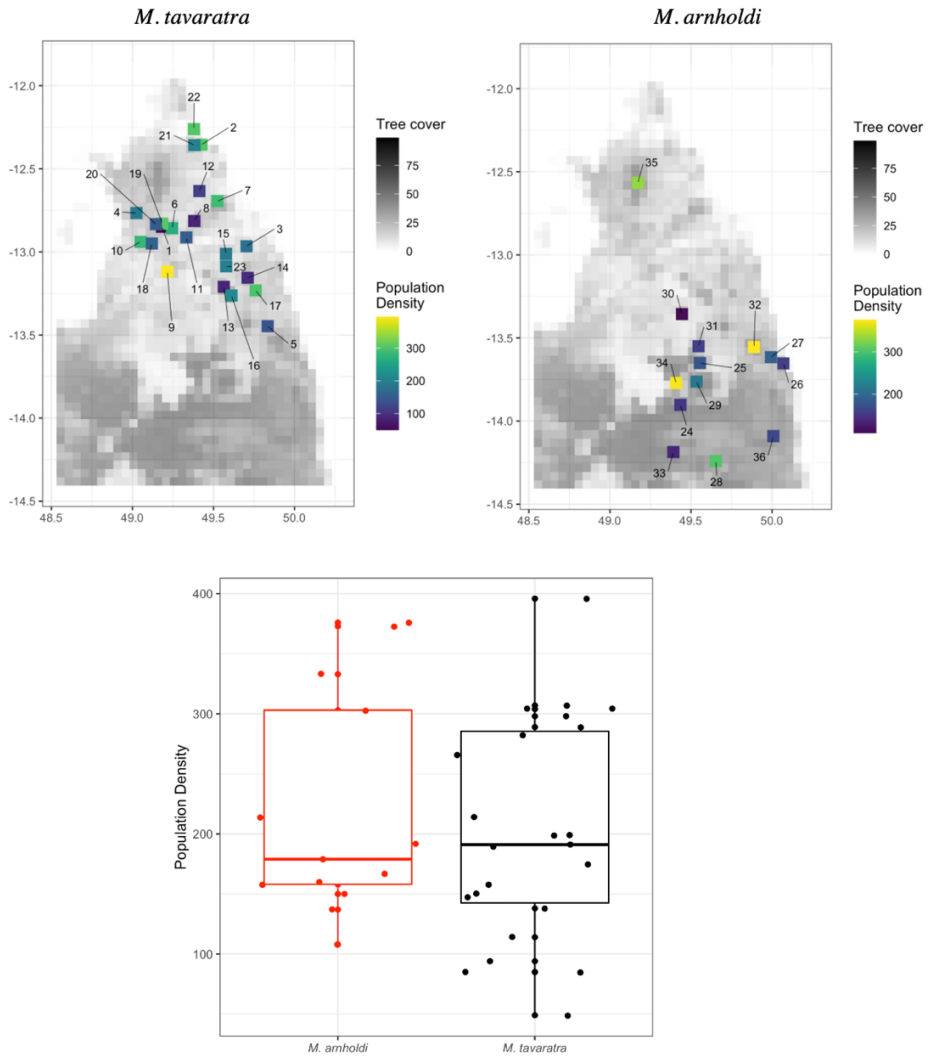


Figure 12. Population density of *M. tavaratra* and *M. arnholdi* across northern Madagascar. Refer to Table 1 for details on the names of the visited locations. Population density is expressed in individuals per km^2. In *M. tavaratra*, site 9 (Andavakoera) has the highest population density (396 ind/km^2) and site 1 (Ambondromifehy) the lowest population density (49 ind/km^2). In *M. arnholdi*, site 32 (Bezavona_Ankirendrina) and 34 (Maladialina) have the highest population density (~370 ind/km^2) and site 30 (Ankijanala) the lowest population density (108 ind/km^2).

Table 5. Estimated Population density through line-distance sampling in *Microcebus tavaratra*, *M. arnholdi* and *M. mittermeieri*. Site: location name; CODE: location ID as represented in Fig. 1; Nb. Tr: number of surveyed transects; Effort: sum of transect lengths; Obs: number of individual observations; ESW: effective strip-width; D: population density in ind/km^2; D (sd): standard deviation of population density; Min: lower limit 95% confidence interval; Max: upper limit 95% confidence interval; Craver-Mises p-val: significant test assessing whether data from the empirical distribution function and the fitted cumulative distribution function are from the same distribution; Det. Function: fitted detection function.

M. tavaratra											
Site	CODE	Nb.Tr	Effort (Km)	Obs	ESW (m)	D	D (sd)	Min	Max	Craver-Mises p-val	Det. Function
Ambondromifehy	1	5	12.1	15	11.2	49	26	16	150	0.68	Uniform + polynomial adj.
Ampiho	2	10	12.9	46	5.8	304	124	126	730	0.34	Hazard-rate
Ampondrabe	3	3	7.4	20	8.9	175	142	14	2244	0.72	Uniform + polynomial adj.
Analabe	4	13	10.9	24	5.1	189	36	127	282	0.76	Uniform + polynomial adj.
Analafiana	5	8	6.1	10	5.4	138	78	45	428	0.64	Half-normal + cosine adj.
Analamahitsy	6	11	19.1	59	5.9	266	93	126	562	0.45	Uniform + Hermite polynomial adj.
Analamerana_E	7	16	36.5	95	4.5	289	54	199	421	0.48	Half-normal + cosine adj.
Analamerana_W	8	5	22.4	17	4.7	85	73	12	621	0.67	Uniform + polynomial adj.
Andavakoera	9	10	26.0	67	3.4	396	112	215	728	0.60	Half-normal
Andrafiabe	10	17	26.1	149	10.8	282	70	168	471	0.08	Hazard-rate
Andrafiamena	11	6	13.6	20	4.5	150	52	69	325	0.91	Uniform + polynomial adj.
Ankarongana	12	6	9.0	16	7.4	114	48	45	292	0.36	Uniform + polynomial adj.
Antsahabe	13	3	9.0	16	11.1	85	56	15	489	0.98	Half-normal
Bekaraoka	14	18	57.6	112	10.5	94	13	71	124	0.54	Hazard-rate
Benanofy	15	4	17.0	36	5.3	191	70	69	534	0.11	Hazard-rate
Binara	16	12	10.3	22	5.2	214	93	88	521	0.76	Uniform + Hermite polynomial adj.
Bobankora	17	8	25.2	128	8.7	298	62	187	475	0.23	Hazard-rate
Mahamasina	18	21	32.6	99	10.4	158	25	115	216	0.60	Hazard-rate
Marotaolana	19	3	8.4	37	7.3	307	150	47	2011	0.12	Hazard-rate
Marovato	20	20	29.5	88	10.8	147	28	101	216	0.54	Hazard-rate
Montagne des francais	21	10	10.0	23	6.0	199	65	103	383	0.44	Uniform + polynomial adj.
Oronjia	22	19	16.3	44	4.2	304	107	153	602	0.55	Hazard-rate
Solanampilana	23	11	29.1	77	6.8	199	45	122	324	0.57	Uniform + cosine adj.

M. arnholdi											
Site	CODE	Nb.Tr	Effort (Km)	Obs	ESW (m)	D	D (sd)	Min	Max	Craver-Mises p-val	Det. Function
Ambalasoa	24	17	15.8	47	9.9	150	44	84	267	0.96	Hazard-rate
Ambaliha	25	6	3.8	14	9.6	179	74	75	428	0.96	Uniform + cosine adj.
Ambohitrandrina	26	8	3.8	10	7.9	160	52	81	315	0.59	Half-normal
Analalva	27	4	5.4	8	3.5	192	143	43	864	0.43	Half-normal + cosine adj.
Andantony	28	4	1.9	7	5.8	303	123	123	750	0.71	Half-normal
Andravory	29	22	16.9	55	7.7	214	43	144	319	0.68	Half-normal
Ankijanala	30	4	1.8	7	16.6	108	83	22	528	0.86	Uniform + polynomial adj.
Antsahavary	31	8	7.0	19	9.0	158	80	59	423	0.84	Uniform + cosine adj.
Bezavona_Ankirendrina	32	8	3.3	18	6.4	376	172	152	925	0.83	Uniform + cosine adj.
Lohananjialava	33	4	3.7	12	11.0	137	50	63	296	0.27	Uniform + polynomial adj.
Maladialina	34	11	6.6	28	5.9	373	97	218	639	0.07	Hazard-rate
Montagne d'Ambre	35	37	37.9	100	4.1	333	54	243	457	0.11	Uniform + cosine adj.
Tsihomanaomby	36	4	4.3	8	5.2	167	66	71	394	0.58	Half-normal
M. mittermeieri											
Site	CODE	Nb.Tr	Effort (Km)	Obs	ESW (m)	D	D (sd)	Min	Max	Craver-Mises p-val	Det. Function
Andohanibemarivo	37	7	5.5	7	9.9	62	30	23	168	0.91	Uniform + polynomial adj.

References

Buckland, S.T., 2001. Introduction to Distance Sampling: Estimating Abundance of Biological Populations. Oxford University Press.

Miller, D.L., Rexstad, E., Thomas, L., Marshall, L., Laake, J.L., 2019. Distance Sampling in R. Journal of Statistical Software 89, 1–28.

Setash, C.M., Zohdy, S., Gerber, B.D., Karanewsky, C.J., 2017. A biogeographical perspective on the variation in mouse lemur density throughout Madagascar. Mammal Review 47, 212–229.

Sgarlata, G.M., Salmona, J., Pors, B.L., Rasolondraibe, E., Jan, F., Ralantoharijaona, T., Rakotonanahary, A., Randriamaroson, J., Marques, A.J., Aleixo-Pais, I., Zoeten, T. de, Ousseni, D.S.A., Knoop, S.B., Teixeira, H., Gabillaud, V., Miller, A., Ibouroi, M.T., Rasoloharijaona, S., Zaonarivelo, J.R., Andriaholinirina, N.V., Chikhi, L., 2019. Genetic and morphological diversity of mouse lemurs (Microcebus spp.) in northern Madagascar: The discovery of a putative new species? American Journal of Primatology 81, e23070.

7.7 Genetics of Malagasy Mammals: An Overview

Lounès Chikhi[1,2] and Gabriele Maria Sgarlata[1]

1- Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, 2780-156 Oeiras, Portugal;
2- Laboratoire Evolution & Diversité Biologique (EDB UMR 5174), CNRS, IRD, UPS, Université de Toulouse Midi-Pyrénées, Toulouse, France

Introduction

The aim of this contribution is to review the use and interpretation of genetic and genomic data generated over the last decades on Malagasy mammals, and provide an overview of this rich and diverse literature. We have explained most, but not all, technical and terminological issues in separate boxes. While we argue that genetic studies are still too limited in number, we could not do justice to the wide-ranging literature, and have limited ourselves to work focusing principally on within species diversity. Our review superficially addresses phylogeographic and phylogenetic studies as they are treated in detail in other contributions herein.

Genetic data are increasingly used across a wide range of taxa and disciplines, from ecology, to taxonomy, evolution, anthropology, biodiversity, and conservation (Goldstein and Chikhi 2002; Steiner et al. 2013; Leitwein et al. 2020; see Box 1). For Malagasy organisms, genetic data have also been used to address a wide variety of questions (e.g., Yoder 2013; Lawler 2018). Research teams interested in behavioral ecology have used genetic data to clarify patterns of diversity and relatedness at a local scale and to quantify sex-biased dispersal rates (e.g., Radespiel et al. 2001; Wimmer et al. 2002; Lawler et al. 2003). At a larger temporal and geographical scale, other researchers have investigated complex and sometimes controversial phylogenetic relationships between species that diversified over tens of millions of years and across a wide variety of environments within the island (Yoder et al. 1996, 2003; Jansa et al. 1999; Douady et al. 2002; Everson et al. 2016). For instance, genetic studies have played a central role in our understanding of the relationships between lemurs and other primates and in the demonstration of their monophyly (e.g., Yoder et al. 1996). Genetic data have clarified the former controversial phylogenetic questions, and have established the

now well accepted scenario of ancient arrival of the lemurifom clade and its diversification on Madagascar. Similar genetic studies on other terrestrial Malagasy mammals (Tenrecidae, Eupleridae, and Nesomyinae) have also helped to firmly establish that the colonization events of these three groups were each unique and that these groups diversified in Madagascar (Yoder et al. 2003; Poux et al. 2005; Everson et al. 2016). More recently, ancient DNA studies have provided a fascinating and unique perspective on extinct subfossil lemur species (Karanth et al. 2005; see Godfrey and Jungers, p. xx) showing for instance that genetic diversity of two extinct species was relatively low compared to extant lemur species (Kistler et al. 2015).

At an intermediate geographical and temporal scale (i.e. between phylogenetic and behavioral ecology studies) genetic studies have also been conducted for more than two decades on Malagasy mammals. For instance, Leipoldt et al. (1996) compared genetic diversity among *Lepilemur mustelinus ruficaudatus* (now *L*. *ruficaudatus*; see Radespiel et al., p. xx) sampled in different forests of Kirindy CNFEREF in the central west. Similarly, Rabarivola et al. (1996) compared genetic diversity among populations of *Eulemur macaco* sampled on two near-shore islands, Nosy Be and Nosy Komba, as well as the mainland, in the northwest (see Johnson et al., p. xx, for details on this genus). Phylogeographic studies linking population genetic and phylogenetic analyses have also helped to clarify the relationships between closely related species and test hypotheses concerning species diversification (e.g., Russell et al. 2007 on bats). For example, Markolf and Kappeler (2013) inferred gene-flow between *Eulemur* sister species that are now separated by rivers, and thus suggested that rivers may not always have been a major factor promoting reproductive isolation between sister taxa.

Despite this large diversity of genetic studies across time, space, and taxonomic groups, the associated literature devoted to mammals is highly biased towards lemurs. Out of the >210 articles that we identified on this subject, >170 focused on lemurs. Assuming that there are ~ 250 mammal species on Madagascar out of which 100-110 are lemurs, this suggests that lemurs represent ~40% of mammal species but ~80% of all studies. If, as some authors consider, there are only ~50 lemur species (Tattersall and Cuozzo 2018), and thus ~200 mammal

species, the over-representation of lemurs is even more dramatic. Within lemurs a bias can also be observed towards four genera (*Microcebus*, *Lepilemur*, *Propithecus*, and *Eulemur*). Among non-primate mammals, rodents and bats are the most represented. Interestingly, the most studied mammal species, outside lemurs, is introduced. The invasive Black Rat (*Rattus rattus*: see Ramasindrazana et al., p. xx) is indeed the species of interest of at least seven genetic studies, which have generally focused on aspects associated with the transmission of plague on Madagascar (e.g., Hingston et al. 2005; Tollenaere et al. 2010; see Tortosa, p. xx).

This overview suggests that although important genetic studies have been published, much more work is needed. We have nearly no population genetic data on several lemur genera such as *Phaner*, *Allocebus*, *Prolemur*, or *Cheirogaleus* and even the emblematic *Indri indri* (see Gamba et al., p. xx) or *Varecia variegata* are only represented by relatively few studies each (e.g., Louis et al. 2005; Nunziata et al. 2016; Bonadonna et al. 2019; Baden et al. 2020; see Vasey et al., p. xx, for details on *Varecia*). Beyond lemurs the situation is even more striking. Most species are either not represented at all or only represented by a couple of studies, as can be seen among the Eupleridae Carnivora (Veron et al. 2018), *Cryptoprocta ferox* (see Gerber and Hawkins, p. xx) and *Mungotictis decemlineata* (see Kappeler et al., p. xx). It would thus be fair to say that we have today nearly no understanding of the genetic diversity hosted by most Malagasy mammals. There is an absolute and urgent need to document this unique diversity across the island and across all species, in the hope that such studies will contribute to their conservation.

While our understanding of the genetic diversity of Malagasy mammals is indeed still very limited, the number of Malagasy and international teams developing high quality projects to document that diversity and to identify the corresponding evolutionary processes has increased in the last decade, even if not rapidly enough. The move from population genetic studies using a limited number of independent genetic markers to population genomic studies where tens of thousands of markers or whole genomes are used is now well underway, also for Malagasy mammals. Some examples are the recent genetic studies that have analyzed whole mitochondrial genomes and have thus expanded phylogenetic or phylogeographic work (e.g., Lei et al. 2017). While mitogenomic studies do not actually use large

numbers of independent markers, they are often included among genomic studies because they analyze much larger sequence data sets (~16,000 base pairs [=bp]) than classical mtDNA studies (typically less than 1000-2000 bp). Other genomic studies have used full genome sequences for an increasing number of species and individuals. For instance, next-generation sequencing technologies (see Box 2) have allowed researchers to investigate old questions in new ways, from the diet of *Propithecus tattersalli* (Quéméré et al. 2013) to the adaptive immune system of *P. coquereli* (Larsen et al. 2014; see Lawler and Richards, p. xx, for details on this genus); the demographic history of *Eulemur flavifrons* and *E. macaco* (Meyer et al. 2015), *P. coquereli* and *Prolemur simus* (Hawkins et al. 2018), or *Microcebus* spp. (Poelstra et al. 2020; see Kappeler et al., p. xx); the evolution of sensory receptor genes in *Microcebus* spp. (Hunnicutt et al. 2020); speciation and admixture in shrew tenrecs (*Microgale* spp.; Everson et al. 2020a; see Jenkins et al., p. xx); or ancient admixture in *Cheirogaleus* spp. (Williams et al. 2020; see Blanco, p. xx, for details on the genus).

In addition, the statistical analyses of genetic (see Box 2) and genomic data (see Box 3) from Malagasy mammal species has become increasingly sophisticated, with new modelling approaches being applied to genetic (Olivieri et al. 2008; Schneider et al. 2010; Salmona et al. 2017) or genomic (Hawkins et al. 2018; Poelstra et al. 2020; Williams et al. 2020) data. Some of the new methods allow researchers to better describe the patterns of genetic diversity and differentiation, while others allow the reconstruction of ecological or evolutionary processes, including the identification of major events in the recent evolutionary history of these animals. The distinction between the description of patterns and the reconstruction of evolutionary processes is not always clear-cut as we will discuss below, but this distinction helps us organize this review in the next two sections. We will also discuss several issues related to taxonomy without entering into much details since phylogenetic research is discussed for different mammal groups elsewhere in this chapter (see, for example, Goodman et al., p. xx, for bats; Yoder, p. xx, for lemurs; Veron et al., p. xx, for endemic carnivorans; Jansa and Carleton, p. xx, for endemic rodents).

Genetics at Different Scales: from Social and Local to Landscape and Species Wide Patterns

Early Genetic Studies: A Diversity of Molecular Approaches but Few Population Studies

The earliest genetic studies of Malagasy mammals, while limited in number and in taxonomic diversity, used a surprisingly varied set of molecular approaches demonstrating an early interest of researchers in the molecular methods of their times (see Box 3 for definitions of methods). The techniques used included immunodiffusion (Dene et al. 1976, 1980), karyotypes (Rumpler and Dutrillaux 1976; Rumpler et al. 1985), DNA-DNA hybridization (Bonner et al. 1980), restriction fragment length polymorphism (=RFLP; Crovella and Rumpler 1992; Jung et al. 1992), repetitive DNA (Montagnon et al. 1993; Crovella et al. 1994), random amplified polymorphic DNA (=RAPD; Leipoldt et al. 1996; Neveu et al. 1998), nuclear gene sequences (Koop et al. 1989), and mtDNA sequences (Hasegawa et al. 1990; Del Pero et al. 1995). However, with the exception of the study of Arnaud et al. (1992) most of the studies mentioned above were not about population diversity but rather aimed at answering taxonomic and phylogenetic questions. As we will see throughout this contribution, this is still the case for most genetic studies on Malagasy mammals more than two decades later. This does not mean that all taxonomic and phylogenetic studies address the same questions, and some aimed at clarifying lemur taxonomy (Rumpler and Albignac 1978; Crovella and Rumpler 1992; see Yoder, p. xx, for further details), whereas others examine the relationships of lemurs with primate groups outside of Madagascar (Dene et al. 1976; Dutrillaux 1979; Goodman et al. 1994).

Population studies were thus uncommon and they usually estimated genetic differences between groups of individuals sampled in different but geographically close locations. The main questions were 1) whether genetic diversity varied across locations, 2) whether individuals sampled from different locations could be considered to be from the same or different populations (or different species in Bachmann et al. 2000), and 3) whether some identified populations had lower diversity, and were thus possibly isolated. For instance, Arnaud et al. (1992) compared the genetic diversity of captive and wild samples of *Eulemur macaco*.

Leipoldt et al. (1996) used RAPDs on *Lepilemur ruficaudatus*, and detected departures from random mating but no major differences in genetic diversity across populations. Tomiuk et al. (1997) used allozymes on the same populations and confirmed strong departures from random mating with significant deficits in heterozygotes (see Box 1 and Box 2). Bachmann et al. (2000) extended these studies by using allozymes, RAPDs, and karyotypic analyses, the latter for species validation. Similar population genetic studies were carried out by Rabarivola et al. (1996) on *E. macaco*, Fausser et al. (2000) on *E. m. flavifrons* (now *E. flavifrons*), and Hafen et al. (1998), Neveu et al. (1998), and Tomiuk et al. (1998) on *Microcebus murinus*.

Population genetic publications concerning Malagasy mammals mainly used two types of genetic markers (see Box 3). Leipoldt et al. (1996), Hafen et al. (1998), and a few others used RAPDs, whereas Rabarivola et al. (1996) and Tomiuk et al. (1997) used allozymes. This is notable because allozymes had been extensively used in many regions of the world and were probably the dominant class of genetic markers typed to study diversity in many taxa from insects to birds, fish, and mammals during this early phase of genetic studies (see review by Allendorf 2017). However, most genetic markers were rapidly replaced by microsatellite and mtDNA loci in studies of Malagasy mammal. From a practical point of view, the mtDNA and microsatellite markers had several advantages. Minute quantities of samples such of an ear or tail or non-invasive samples such as hair or feces, could be used to obtain mtDNA or microsatellite data (Taberlet et al. 1996; Goossens et al. 2000, 2011). These techniques were also important for researchers working on elusive, rare or endangered species that were difficult to approach or capture. However, we remark that genetic studies using non-invasive samples have not been notably frequent among Malagasy mammals, with a few exceptions (e.g., Quéméré et al. 2010a, 2910b, 2012; Salmona et al. 2015, 2017 for *Propithecus*; Bertoncini et al. 2017 for *E. collaris*).

Before moving to the next section, it is important to stress that all these studies were focused on lemurs. For non-primate Malagasy mammals little seems to have been done in that period. Jansa et al. (1999) analyzed mtDNA sequences to reconstruct the phylogeny of Malagasy rodents and Douady et al. (2002) studied

the phylogeny of tenrecs using mtDNA and nuclear gene sequences. We did not find any published population genetic study from this early period.

Microsatellites and mtDNA: The First Studies on Social and Population Structure

When microsatellites started to be adopted by researchers working on Malagasy mammals there was a change of focus. Instead of replacing allozymes and other markers in the analysis of populations, microsatellites were mainly used with mtDNA sequences to study processes on a much more local scale. In the period between 2000 and 2005, most of these studies were carried out by research teams that had established field stations and were mainly interested in using highly polymorphic markers to identify individuals and study local (possibly sex-biased) migration patterns, differences in male and female behavior, and the details of mating systems. These sociogenetic studies sometimes established aspects suspected or expected from behavioral or comparative data from other mammal species. For instance, in general higher migration rates were expected in males compared to females based on data collected across mammals (Greenwood 1980). In other cases, genetic data allowed researchers to identify behaviors that were difficult to observe or were expected to be rare or absent. Outside of lemurs, such sociogenetic studies have been applied to very few other mammals with one study on *Cryptoprocta ferox* (Lührs et al. 2013) and two studies on *Mungotictis decemlineata* (Schneider and Kappeler 2014; Schneider et al. 2016). As the next section will show, sociogenetic studies have offered remarkable insights into the behavioral ecology of lemurs, yet there are still no such studies on rodents, tenrecs or bats.

Around the same period, but at a much slower pace, microsatellite and mtDNA data started being used to study genetic diversity at the population level. The number of population genetics studies remained limited for several years. The first of such studies was that of Mayor et al. (2002) on several species of the *Propithecus* genus. These authors provided and tested genetic markers for future population genetic studies, but did not actually carry out population genetic analyses. Three years later Louis et al. (2005) provided the first example where genetic variation was analyzed across populations. The authors developed 25

microsatellite markers for *Varecia variegata*, and compared populations from four different locations, which encompassed most of the species' distribution in the moist evergreen forests of eastern Madagascar. It would take another few years before other population genetic studies would be carried out addressing questions related to population genetic diversity and differentiation.

The next two sections examine both sociogenetics and population structure studies in greater detail, and underline the progress done in our understanding of social organization and genetic diversity in Malagasy mammals.

Sociogenetics studies: Genetic data have been used for decades to study various aspects of social systems and have helped to elucidate details of the social organization of many species (see reviews by Sugg et al. 1996; Di Fiore 2003). Sociogenetic studies have typically focused on kin composition, parentage analysis, inbreeding patterns, and sex-biased dispersal. While a few studies have stressed one particular aspect of social organization, most studies have typically investigated several components of social systems. Fietz et al. (2000) developed and applied microsatellite data to examine the mating system of *Cheirogaleus medius* in central west populations and provided the first estimates of genetic diversity for this species. In *Microcebus murinus* from Ankarafantsika, Radespiel et al. (2001) found that females from the same sleeping group were genetically more closely related than females from different sleeping groups ($r_{co\text{-}sleep}$ = 0.362; $r_{not\ co\text{-}sleep}$ = 0.006). In *Hapalemur griseus alaotrensis* (now *H. alaotrensis*; see Tan et al., p. xx, for details on this genus), Nievergelt et al. (2002) found that individuals of the same social group were genetically much more related to each other than individuals from different social groups (r_{same} = 0.327 and $r_{between}$ = 0.054). They also found that the individuals of the same social group were all sharing the same mtDNA haplotype. Interestingly, Nievergelt et al. (2002) used microsatellites developed for *Eulemur mongoz* and *E. rufus* one decade earlier (Merenlender 1993). In the southwest, Lawler et al. (2003) measured relatedness in *Propithecus verreauxi* and found that females from the same social group had higher levels of genetic relatedness than males ($r_{females}$ = 0.1 – 0.27; r_{males} = 0.04 – 0.14). Wimmer and Kappeler (2002)

observed a similar pattern in populations of *Eulemur fulvus rufus* (now *E*. *rufifrons*) in the central west ($r_{females}$ = 0.26 – 0.70; r_{males} = -0.10 – 0.27).

Nuclear (usually microsatellites) and sex-specific markers (mtDNA and Y-linked genes), together with spatial patterns of diversity and relatedness, can provide complementary information on sex-biased dispersal and kinship composition. For example, Wimmer and Kappeler (2002) found in *E*. *rufifrons* a few instances of high average relatedness among adult males of a social group, while observing high mtDNA diversity, suggesting that those individuals were relatives but that they were not maternally related. Using highly variable mtDNA *D-loop* sequences Wimmer et al. (2002) assessed the matrilineal structure of a *M*. *murinus* population in a restricted area (~4.5 ha) of the Kirindy (CNFEREF) forest. They found that 80% of the female mouse lemurs were sharing only one haplotype, whereas males exhibited most of the mtDNA diversity of the population. Given that males inherit their mtDNA from their mothers and cannot transmit it, this suggests notably higher dispersal rates in males, specifically recent migration from other local populations or social groups. Moreover, females with distinct haplotypes formed distinct spatial clusters with little overlap, suggesting a hierarchical social organization in which females form groups with their kin (social unit) and matrilineal structure sharing the same mtDNA haplotype above the social unit (matrilines). Overall, these findings were thus a strong evidence that females were philopatric and that males were the dispersing sex, as is usually observed in mammals (Greenwood 1980; Trochet et al. 2016). In a later study, Fredsted et al. (2004) examined the same population but at a larger spatial scale (~1200 ha). They confirmed that the matrilineal structure of *M*. *murinus* and male-biased dispersal were consistent even at a much larger geographical scale (~250 times larger) (Figure 1). In the sympatrically occurring *M*. *berthae*, Dammhahn and Kappeler (2005) found that females were more related than males ($r_{females}$ = 0.049; r_{males} = −0.007), confirming male-biased dispersal also in this mouse lemur species, although with lower values of relatedness than those mentioned above in *M*. *murinus*. A similar conclusion was drawn by Nievergelt et al. (2002) in *H*. *alaotrensis*, where most genetic diversity (nuclear and mitochondrial) was brought by reproducing males, thus suggesting male-biased dispersal.

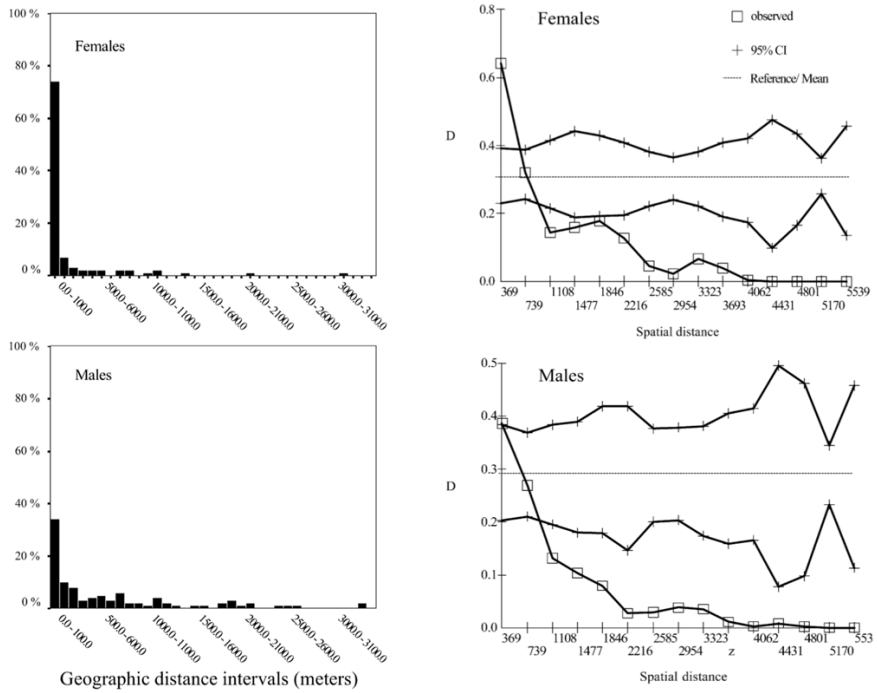


Figure 13. Sociogenetic analysis performed on *Microcebus murinus* individuals from the Kirindy CNFEREF Forest, central western Madagascar. **First column:** Distribution of the geographic distances to nearest individual with the same haplotype (%), for males and females (upper and lower histogram, respectively). **Second column**: Spatial "autocorrelation" analysis performed on measures of genetic similarity (mean number of haplotypes in common of individuals belonging to a given distance class), separately, for females and males. Figure adapted from Fredsted et al. (2004).

Other quantitative estimates of sex-specific dispersal patterns (e.g., distance) can be obtained by comparing genetic relatedness and geographical distance between individuals of the same sex. For instance, Radespiel et al. (2003) computed the geographic distances between closely related female or male dyads (r = 0.5) in *M*. *murinus* populations from the northwest. They found that inter-individual distances were on average smaller between closely related females than between males (100 and 400 m, respectively). This, together with the results on co-sleeping individuals mentioned above, suggested that in *M*. *murinus* males move away from their natal group, while females establish a hierarchical social organization together with their kin. Kappeler et al. (2002) studied the social

structure of *Mirza coquereli*, a solitary lemur species where individuals forage alone and rarely share nests with other adults (see Kappeler, p. xx, for details on this genus). Since no social groups are formed in this species, these authors investigated sex-biased dispersal by quantifying the mean distance between centers of activity of either male or female dyads sharing or not a mtDNA haplotype. They found that maternally related females were significantly clustered in space ($dist_{same\text{-}hap}$ = 90-250 m; $dist_{different\text{-}hap}$ = 350-380 m), whereas males did not show this pattern ($dist_{same\text{-}hap}$ = 240-440 m; $dist_{different\text{-}hap}$ = 320-340 m), thus suggesting that also in *M. coquereli* males are the dispersing sex.

Other approaches that explicitly model dispersal can be used to estimate more precisely dispersal distances in males and females, such as spatial autocorrelation analysis (see Box 4). However, accurate estimates can be obtained only if certain criteria are met. For instance, Vekemans and Hardy (2004) noted that the sampling area should neither be too large nor too small in relation to the species' mean dispersal distance. Namely, they suggested a sampling area on the order of 10 to 50 times σ^2, where σ^2 is proportional to the area where dispersal takes place (see Box 4). Otherwise, relative rather than absolute values of dispersal parameters may be estimated. For instance, Fredsted et al. (2005) used 10 microsatellite loci from 167 grey *Microcebus murinus* (83 males, 83 females and one individual not sexed) across an area ~12 km^2 in the Kirindy CNFEREF forest. They computed the mean number of shared alleles between individual pairs within the same distance interval for males and females, separately. The results showed no significant departure in the number of shared alleles across all distance classes for males, even if a higher value was observed for the first distance class. For females, however, a significant and positive value was observed for the first distance class (up to ~250 m); hence, suggesting longer dispersal distances in males than in females. Similar results were obtained on the same population using mitochondrial *D-loop* sequences (Fredsted et al. 2004; see Figure 1).

Genetic parentage analyses have been used to quantify reproductive skew (i.e. variance in the reproductive success), and extra-group reproduction (i.e. the frequency of offspring sired by individuals outside the social group), components of social organization that are not necessarily easy to study without genetic data.

Evidence of male reproductive skew were found by Wimmer and Kappeler (2002) in *E. rufifrons*. They found that within four social groups there were only one or two reproductively successful males siring offspring, even though each group contained several adult males. The two social groups for which there was additional behavioral data showed that three out of four (75%) and three out of five (60%) offspring, respectively, were sired by one single male, suggesting an average male reproductive skew of 67% (six out nine offspring). Gachot-Neveu et al. (1999) used RAPD markers (see Box 3) to determine paternity in two captive social groups of *E. fulvus mayottensis* (now *E. fulvus*) during a four year study. In one group, the socially dominant male sired 100% of seven infants, whereas in the second group two out of five infants (40%) were sired by a subordinate male. Pronounced male reproductive skew was also observed in a population of *P. verreauxi* at Kirindy (CNFEREF) forest by Kappeler and Schäffler (2008). During the mating seasons in which social groups contained multiple males, they found that 31 out of 34 offspring (~90%) were sired by the dominant male of the different groups.

Both these studies suggest that, in these species, the resident males and females were generally the genetic parents of the offspring. However, they also showed cases where other males had managed to reproduce in the group. Genetic studies on several pair-living and group-living lemur species have shown that a non-negligible proportion of offspring can be indeed sired by a male from outside the social unit, typically defined as extra-group paternity (EGP), which includes also extra-pair paternity in the case of pair-living species. For instance, in a recent study, Bonadonna et al. (2019) investigated genetic monogamy in *Indri indri*, a lemur species known to be socially monogamous, even though individuals engaging in extra-pair copulation have been observed (Bonadonna et al. 2014). Using genetic parentage analysis, Bonadonna et al. (2019) demonstrated that EGPs occur at a rate of ~8% in *I. indri*. Similar rates were reported for *H. alaotrensis* (8.5%; Nievergelt et al. 2002), *E. rubriventer* (~5%; Jacobs et al. 2018), whereas much higher rates of EGP were found in other lemur species such as *C. medius* (40%; Fietz et al. 2000), *Lemur catta* (3/14 offspring = 21% and 3/9 offspring = 30%; Parga et al. 2016), and *Phaner furcifer* (75%, n = 4; Schülke et al. 2004). Also, studies on the same species but carried out in different locations may generate different results. For instance,

Lawler (2007) reported that 29 out of 52 males (56%) sired offspring outside their resident group in *P. verreauxi* a number that is very different from the 10% observed by Kappeler and Schäffer (2008).

Other genetic studies have reported cases of mixed paternity, where some of the offspring of a social unit appeared to be sired by males that entered temporarily in the group. For instance, Kappeler et al. (2002) found in a population of *Mirza coquereli* that no resident male could be assigned as the genetic father of four out of 13 offspring from known mothers. They interpreted these results as evidence for mixed paternity, suggesting that those four offspring were sired by males that entered in the study area only for a short period of time. A genetic study on two social groups of *L. catta* living in natural habitat enclosures showed that two immigrant males, introduced in the forest enclosure, did sire offspring with resident females. Interestingly, even in presence of resident males, all estrous females were engaging in sexual encounters with immigrant males, and the females exclusively attracted to the immigrants reproduced with them (Pereira and Weiss 1991).

We found only two studies on the sociogenetics of non-primate Malagasy mammals, both in endemic euplerid Carnivora. One was measuring relatedness of 22 adult male *Cryptoprocta ferox* (Lührs et al. 2013) captured over 9 km^2 in the Kirindy (CNFEREF) Forest, showing that four out of six dyads (individuals sharing the same home range) were composed of litter mates ($r \geq 0.40$). The other study, Schneider et al. (2016) is the first population genetic and sociogenetic study in *Mungotictis decemlineata*, characterizing its social organization and mating system. In this species, males and females form separate social units with either one to three females or two to four males (Schneider and Kappeler 2016). Schneider et al. (2016) showed that mtDNA diversity was significantly higher in males than females ($h_{males} = 0.818$, $h_{females} = 0.714$) and that average pairwise relatedness was higher in female social units than in most male associations ($r_{female} = 0.25 – 0.54$, $r_{male} = 0 – 0.32$), suggesting that 1) female units may be formed by mothers and their offspring, and 2) that male *M. decemlineata* disperse further away from their natal unit than females. Parentage analyses pointed to a mildly promiscuous mating system and low levels of reproductive skew. However, the authors observed that the infants surviving the first year of life (28%) were all born from the dominant female of the

social unit, leading to the hypothesis that postpartum mechanisms might be involved on regulating reproductive skew in female *M. decemlineata*.

The consequences of social structure and mating systems on patterns of genetic or genotypic diversity are still not fully understood, despite early theoretical work not directly related to Madagascar by pioneers such as Prout (1981) and Chesser (1991). Population geneticists have tended to ignore social structure, and social groups were either modelled as small populations or simply ignored. In other words, it was assumed that social groups were unlikely to have major effects on global genetic patterns, and that such effects, if they existed, would be ephemeral with limited consequences on allelic frequencies and thus negligible from an evolutionary perspective (e.g., Storz 1999). Recent theoretical work suggests that this may not be that simple (Parreira and Chikhi 2015). First, genotypic changes may be important for genes under selection affecting the overall fitness of individuals and populations. Secondly, social structure may be crucial if we want to understand data such as patterns of inbreeding at the genomic scale. For example, behavioral ecologists have tended to interpret heterozygosity excess within social groups as an indication of inbreeding avoidance. This interpretation has been criticized, as an increasing number of field studies have demonstrated that inbreeding tolerance and preference exist in wild populations (Szulkin et al. 2013). Parreira and Chikhi (2015) explicitly simulated social groups and found that social structure may naturally generate excesses in heterozygosity (i.e. $H_{obs} > H_{exp}$, see Box 1), without the need to incorporate inbreeding-avoidance behaviors. These and other results suggest that in order to understand patterns of genotypic diversity at the genomic level, it may be necessary to consider social structure, mating systems, and other details of social systems. Thus, instead of interpreting “outbreeding” as a consequence of inbreeding avoidance, one should test whether the measured “outbreeding” could be the result of how social systems are organized and connected, and whether an appropriate neutral model that incorporates social structure and connectivity, could be used to predict the observed genotypic diversity patterns (Di Fiore and Valencia 2014; Parreira and Chikhi 2015).

Parreira and Chikhi (2015) also found that when social groups are ignored and pooled for genetic analyses the heterozygote excess decreases and

eventually disappears as the number of pooled samples increases. This gives the misleading impression that the (pooled) sample is at Hardy-Weinberg equilibrium, when actually mating strongly departs from random mating (see Box 1 and Box 3). Now, returning to Madagascar, there are still too few empirical studies on inbreeding. Lawler et al. (2003) measured within social group inbreeding in a study on *Propithecus verreauxi* revealing an excess of heterozygosity in offspring, but not so much in reproductive adults. They attributed this result to female philopatry and male-biased dispersal. Quéméré et al. (2010b) measured heterozygosity (H_{obs} and H_{exp}) within social groups of *P*. *tattersalli* and found extremely high heterozygosity excesses in some social groups, but did not interpret them. Recently, Parreira et al. (2020) applied the model of Parreira and Chikhi (2015) to genetic and genotypic data from *P*. *tattersalli*, finding that the results published by Quéméré et al (2010) could also be a simple consequence of the social structure, mating system, and sampling scheme. Altogether, this stresses the importance of both field observations and models to improve our understanding of genetic and genotypic data. More work is needed on Malagasy mammals to integrate sociogenetic work and population genetic studies and modelling.

Population studies of diversity and differentiation: From early to recent studies: As noted above, population genetic studies of Malagasy mammals took several years, after those of Mayor et al. (2002) and Louis et al. (2005), to appear in the literature. Many of these early studies were descriptive and looked at genetic diversity and differentiation within single species. This was the case of the two companion articles of Ramanamahefa et al. (2009a, 2009b) on *Eulemur coronatus* and *E*. *sanfordi* living sympatrically in the same forests of northern Madagascar, that of Ranaivoarisoa et al. (2010) on *E*. *collaris* in the southeast, that of Razakamaharavo et al. (2010) on *Varecia rubra* in the northeast, and Quéméré et al. (2010a, 2010b) on *Propithecus tattersalli*.

Interestingly, the first population genetic studies performed on another group of Malagasy mammals focused on the introduced *Rattus rattus* (Hingston et al. 2005; Gilabert et al. 2007; Loiseau et al. 2008; Tollenaere et al. 2010). Beyond rats, most genetic studies of Malagasy mammals, including recent ones (e.g.,

Rakotoarivelo et al. 2019 on *Macronycteris commersoni*), are based on mtDNA analyses. There are exceptions such as Andrianaivoarivelo et al. (2009) who identified and typed 22 microsatellite loci in one southern population of the endemic bat *Rousettus madagascariensis*, and Goodman et al. (2010) who analyzed the genetic structure of the same species using mtDNA sequence data and six microsatellites developed for another *Rousettus* species, *R*. *leschenaultii*, from southeastern Asia (Hua et al. 2006). Goodman et al. (2010) found very low levels of genetic differentiation (Fst ~0.004) between Malagasy populations distributed across the island, and they also studied diversity of the closely related species *R*. *obliviosus* from the nearby Comoros Islands.

Using mtDNA sequence data, several authors have looked at the genetic structure of other bat species. Ratrimomanarivo et al. (2009) observed limited genetic differentiation at the level of the island for the endemic *Mormopterus jugularis* and Chan et al. (2011) reported low genetic differentiation for two *Pteropus* species, one occurring within Madagascar (*P*. *rufus*) and the other in the Malagasy Region (*P*. *seychellensis comorensis*), showing parallel results to those of Goodman et al. (2010) on *Rousettus*. The limited genetic differences described by Chan et al. (2011) at the island level were not always small. For instance, they reported large Phi_{ST} values (0.18 or 0.25, Phi_{ST} is similar to F_{ST}, see Box 1, but for DNA sequences) between Miandrivazo (central west Madagascar) and two other Malagasy populations (northwest and west), which suggests that genetic differentiation is not limited and needs further investigation in *P*. *rufus*. Similarly high and even higher Phi_{ST} values were found in the endemic *Myotis goudoti* (Weyeneth et al. 2011) suggesting ancient and regional diversification in some bat species. Finally, the recent study of Rakotoarivelo et al. (2019) suggests that there may be rather high levels and possibly ancient genetic differentiation in *Macronycteris commersoni* populations, an endemic to Madagascar. We wish to stress that results suggesting limited differentiation at a wide geographic scale, as in *Mormopterus jugularis* or *R*. *madagascariensis* do not mean that it will be impossible to detect the effect of local genetic differentiation when comparing populations of variable size between geographically close localities. Given the high level of allelic diversity observed in *Rousettus* microsatellites (Andrianaivoarivelo et al. 2009) one could potentially carry

out very fine-scale studies of local genetic differentiation. There are currently no such studies in Madagascar for bats. Clearly the bats of Madagascar, which represent nearly 30 independent colonization events of the island (Goodman et al., p. xx), remain understudied and appear to exhibit a range of species-specific patterns that need to be studied in more detail and whose important ecological role is increasingly recognized (see Kemp et al. 2019 and Box 5).

The early population genetic studies reported so far for Madagascar looked at population structure from the point of view of the sampled locations: groups of individuals sampled in different locations, called populations, were compared as we have mentioned above. However, an important shift happened with the study of Pritchard et al. (2000), not related to Madagascar mammals, who turned the question upside down. Their idea was to ignore the sampling locations of the groups of individuals (the "populations") and use only the genotypic information to identify "hidden genetic structure". They used a Bayesian approach and implemented it in the STRUCTURE software that has become one of the most widely used programs in the field for the last 20 years. Now turning back to Madagascar, these methods have been used across different taxa to ask a range of questions. Brenneman et al. (2016) used STRUCTURE and showed that there was no relation between genetic structure and the phenotypes that were associated with putative subspecies of *Indri indri*. Pastorini et al. (2009) used the method to identify putative hybrids between *E*. *fulvus* and *E*. *mongoz*. Markolf et al. (2013) employed this method, as well as others, to study the *E*. *fulvus* or the brown lemur complex, which includes seven taxa regarded as either subspecies or species by various authors, and concluded that species status could be confirmed. In most genetic studies, however, the authors were interested in identifying the potential effects of habitat fragmentation within species and in identifying putative populations that could possibly need to be protected (Olivieri et al. 2008; Ramanamahefa et al. 2009a, 2009b; Quéméré et al. 2010a).

The first of such studies is Olivieri et al. (2008) where the method was applied to three mouse lemurs, *Microcebus ravelobensis*, *M*. *bongolavensis*, and *M*. *danfossi* from northwestern Madagascar. In *M*. *ravelobensis*, they obtained samples from eight locations, four of which located in the Ankarafantsika National Park (about

1000 km^2), and the others from four forest fragments (areas: 4-37 km^2). They identified two main clusters. The first mainly consisted of individuals from Ankarafantsika and the two forest fragments closest (but not connected) to the protected area today, one (Tsiaramaso) at a distance of ca. 20 km from Ankarafantsika and the other, Sainte Marie, at a distance of 1 km that was probably connected in the recent past. The other genetic cluster corresponded to the two remaining lowland forests located around 80 km from the protected area. In a further analysis, the authors also found that the first cluster could be further subdivided into three sub-clusters corresponding each to Ankarafantsika, Sainte Marie, and Tsiaramaso. Finally, they also identified an isolation by distance pattern across all sampled populations, including those located in forest fragments, hence suggesting a recent fragmentation of forest habitats in this region. *Microcebus bongolavensis* was sampled in three forest fragments, with one location >50 km away from the other two. Despite the large geographic distance, the absence of a continuous forest and the lack of connectivity among locations, only one genetic cluster could be identified. In *M. danfossi* individuals were sampled at seven sites corresponding to forest fragments (1.5 to 37 km^2). The authors found higher levels of genetic differentiation in this species than in *M. ravelobensis* but no correlation between genetic and geographic distances. Despite the similarly large geographic distances between some *M. danfossi* samples compared to *M. ravelobensis* and the lack of a large continuous forest habitat, *M. danfossi* individuals grouped mainly in one genetic cluster with the exception of one forest fragment that formed a second cluster.

The first non-lemur studies using STRUCTURE are those of Brouat et al. (2013, 2014) on introduced *Rattus rattus* (see Ramasindrazana et al., p. xx) with broad sampling across the island. Their results showed that these animals belong to two genetic clusters with distinct distributions. One set of genetic samples cluster around Antsiranana (Diego-Suarez) in the very north of Madagascar, whereas nearly all other samples belonged to a second cluster. They also found a few locations where individuals were belonging to the two clusters or possibly admixed. In a very recent study on two shrew tenrec species (see Jenkins et al., p. xx), *Microgale fotsifoty* and *M. soricoides*, Everson et al. (2020a) used the *fineStructure*

method (Lawson et al. 2012), similar to STRUCTURE but adapted to manage genomic data. They obtained ~2500 highly conserved loci for 18 individuals and identified two genetic clusters within *M*. *fotsifotsy* and a possible admixture between *M*. *soricoides* and the southernmost *M*. *fotsifotsy* cluster.

Whereas most of the articles mentioned above mainly looked at genetic differentiation and tried to identify homogeneous genetic clusters, Quéméré et al. (2010b) went further and tried to quantify for *Propithecus tattersalli* the relative importance of various landscape features in the Loky Manambato region, such as the presence of open habitats, and the potential barrier role of the Manankolana River and the national road crossing the region. These authors first identified two main genetic clusters corresponding to individuals sampled on the two sides of the Manankolana River, hence suggesting that the river was the main landscape feature in this region shaping genetic diversity in this species (see Figure 2). The national road did not appear as a significant barrier. In a later study based on 15 microsatellites from *Microcebus tavaratra* sampled in the same region, Aleixo-Pais et al. (2019) found that the Manankolana River was not the main factor correlated to genetic differentiation. They rather identified the two most humid and southern forests as clustering apart from all other samples. Three other studies based on mtDNA have been carried out in the Loky Manambato region, the first two on Nesomyinae rodents, *Eliurus carletoni* and *E*. *tanala* (now *E*. *ellermani*; Jansa et al. 2019) (Rakotoarisoa et al. 2013a, 2013b), and the second on *M*. *tavaratra* (Sgarlata et al. 2018). Both studies found that the Manakolana River did not have such a strong effect on the genetic structure of these species. Hence, the results of landscape genetic studies carried out in the Loky Manambato region on *Microcebus*, *Eliurus*, and *Propithecus* identified different effects of landscape configuration on genetic differentiation and diversity and they underline that studying only one species in a given region is not sufficient to draw general conclusions on the effect of landscape features on genetic variation.

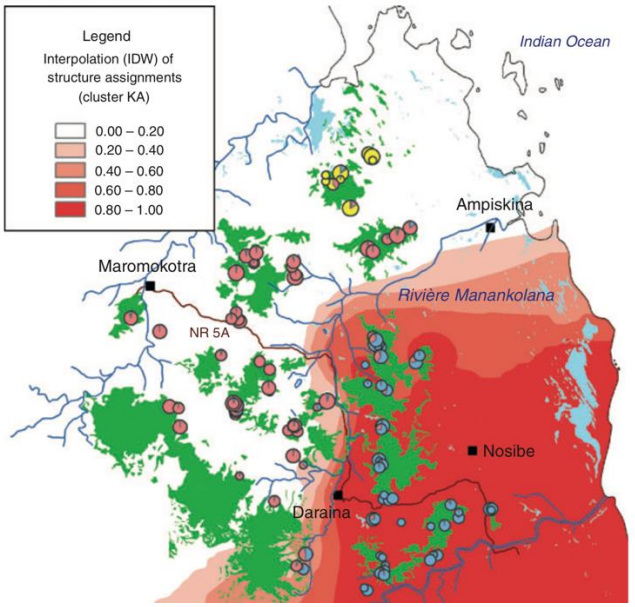


Figure 14. Population genetic structure in *Propithecus tattersalli* across its fragmented distribution range (Loky Manambato region in the north of Madagascar). Pie charts represent social groups of *P. tattersalli*. Their size is proportional to the number of sampled individuals. The colors indicate the average membership coefficients to each of the three genetic clusters (blue, red, and yellow) inferred using the STRUCTURE software. The red gradient illustrates the interpolated membership coefficients of the blue cluster across the study area, showing that the main genetic discontinuity overlaps with the location of the Manankolana River. Figure adapted from Quéméré et al. (2010b).

Genetic results suggest that even congeneric species within the same or neighboring regions can exhibit very different population genetic patterns, as presented by Olivieri et al. (2008) for *Microcebus* spp. This pattern can also be seen in the companion studies carried out in the north by Ramanamahefa et al. (2009a, 2009b) on *Eulemur coronatus* and *E. sanfordi*. These studies sampled individuals at the same locations over a large region between Montagne d'Ambre in the extreme north and the Binara Forest in the southern part of the Loky Manambato region. The authors detected significant genetic differentiation between populations (using a classical sample-based approach) in both species. However, they identified one genetic cluster in *E. coronatus* and two in *E. sanfordi* despite the rather large geographic distances and the absence of continuous habitat today between samples. As a comparison, several clusters were observed just in the Loky

Manambato region in *P. tattersalli* or *M. tavaratra*. Differences between the two *Eulemur* species could be due to different colonization histories or to behavioral or ecological differences between the species.

These examples show that multi-species studies carried out in the same regions or across regions are urgently needed on the island, to understand how and whether natural or anthropogenic landscape features such as rivers, agricultural lands, or open habitats affect different species in similar or different ways. Generation time, dispersal ability and population size, social structure, and many other factors, including the level of forest dependency and anthropogenic pressure, will likely be crucial in that respect. The more we are able to study multiple species from the same or different regions using similar and comparable genetic markers, the less our conclusion will be taxon-biased (see Avise 2000 for general phylogeographic principles; Everson et al. 2020b for a comparative analysis of genetic data across mammal and reptile species). This underlines the importance of studies developing or testing markers across species such as those of Edward Louis and colleagues at the Omaha Zoo. Among these we can identify, for example, the study of Rakotoarisoa et al. (2006) on *P. coquereli*, Tokiniaina et al. (2009) on *Eulemur cinereiceps* and Bailey et al. (2016) on *Propithecus* spp., Andrianaivoarivelo et al. (2009) on bats, and Vogler et al. (2009) on *Cryptoprocta ferox*. Genomic data will also contribute to such efforts if similar markers are used across species. For instance, Poelstra et al. (2020) used a standardized approach on *Microcebus* spp., hence making their data comparable. Among the early population genetics studies, Olivieri et al. (2008) was probably the first comparative population genetic study on Madagascar, quantifying and dating population size changes in the three species of *Microcebus* occurring in the northwest.

What these genetic studies tell us about ancient habitat or population connectivity is not always clear, because they identify ***patterns*** based on genetic structure (or lack of) but do not necessarily aim at identifying the ***processes*** responsible for the observed patterns. It is tempting and probably justified to suggest that the forested habitats currently occupied by *E. sanfordi* and *E. coronatus* in the north, or *M. ravelobensis* in the northwest, were continuous in the recent past, but without proper spatial and temporal modelling, as well as paleoecological data, it is

difficult to clarify what is meant by “recent”. A number of studies have tried to link patterns and processes by inferring the demographic history of Malagasy species for which spatial patterns had been identified and linked to habitat change. For example, the work of Olivieri et al. (2008) on the three species of *Microcebus*, Craul et al. (2009) on *Lepilemur edwardsi* from the same region, and Quéméré et al. (2012) in the Loky Manambato region. As we will present below, many of the methods developed to infer demographic history (historical processes) tend to conversely ignore space (population structure), and thus the associated patterns. But before moving to methods developed to infer the demographic history of species, we need to examine the interplay between landscape and spatial genetics approaches and how some methods have been applied to Malagasy mammals in an attempt to link spatial features and processes.

Investigating Spatial Processes in Malagasy Mammals: Towards the Estimation of Present-day Dispersal Distances Based on Isolation-by-distance (IBD) Patterns

In most species, dispersal is limited by geographical distance and most individuals will reproduce at a relatively short distance from where they were born. This produces a pattern known as ‘isolation-by-distance’ (IBD), in which individuals geographically close are genetically more similar than individuals far apart (Wright 1943). This pattern is often visualized by plotting pairwise genetic distances (between individuals or populations) against the corresponding geographical distance. Theoretical work has shown that the correlation between these geographic and genetic distances is informative about parameters such as species-specific dispersal distances (see Box 4). Similarly, such demographic parameters can also be estimated if one computes the average relatedness (e.g., kinship) between pairs of individuals separated by geographic distances that fall within a set of distance classes (Vekemans and Hardy 2004). For instance, Fredsted et al. (2005) genotyped 83 males and 83 females of *Microcebus murinus* across an area of ~12 km^2. They then computed pairwise geographic distances between all males and all females, separately. They created distance classes that had similar and sufficiently large numbers of pairs and computed the number of shared alleles across

microsatellite loci as a measure of relatedness between pairs of individuals for each distance class. They plotted the average number of shared alleles for the increasing distance classes and found that this number was higher and significant for females from the first distance class (<250 m). For males, even if relatedness was higher in the first than in other distance classes, it was not significantly different from the average number of shared alleles across the sampled area. Thus, although they did not estimate absolute values of effective dispersal distance, their results suggest relatively longer dispersal distances in males than in females (for example, Figure 1).

Theoretical studies suggest that one could go further than these semi-quantitative analyses. For instance, data from studies such as that of Fredsted et al. (2005) could in principle be used together with estimates of population density to derive an estimate of the effective dispersal distance (σ) (see Box 4). This approach has been implemented in the software SPAGeDI (Hardy and Vekemans 2002) and has been applied to microsatellite data across various taxa outside Madagascar and could thus be used to improve our understanding of dispersal patterns in Malagasy mammals. Information on mammal dispersal can also shed light on broader aspects of the ecology of a given species. Brouat et al. (2013) studied the spatial genetic structure of introduced *Rattus rattus* in four regions (area ~150-200 km^2 with 2-6 subpopulations ~2 km apart) in which human plague epidemics are regularly reported. They tested whether plague dissemination correlated with the effective dispersal of *R*. *rattus*, the fleas of which are the main reservoir of the disease on the island. These authors genotyped 910 individuals at 14 nuclear microsatellite loci and correlated genetic differentiation (F_{ST}) and differences in plague seroprevalence levels between pairs of populations within two of the four surveyed regions. They found significant relationship between these two variables only in one of the two sites, suggesting that *R*. *rattus* dispersal is important but not sufficient to explain plague dissemination in all regions of the island.

Studies Identifying Recent and Ancient Environmental Features that Influence Species Dispersal

Environmental features (biotic and abiotic) can influence species dispersal, representing either barriers to gene flow or corridors at the scale of the landscape (Hendry 2004; Nosil et al. 2008). Several approaches allow nowadays to test hypotheses on the factors limiting or favoring species dispersal (see Box 4). Quéméré et al. (2010b) have used Causal Modelling (CM), in addition to other methods, to study the effect of geographical distance, roads, river, and forest canopy on the dispersal of *Propithecus tattersalli*. This technique identified the Manankolana River crossing the region as the main barrier to dispersal in this species, in agreement with other landscape genetic approaches. In addition, a significant effect of geographical distance was detected, implicitly suggesting that the state of habitat fragmentation of the region might not disrupt gene-flow significantly in this species. In the moist evergreen forests of the east, Baden et al. (2019) have used the isolation-by-resistance (IBR) approach, to identify areas that facilitate or restrict gene-flow among sampled *Varecia variegata* individuals, and to investigate the impact of landscape heterogeneity on this species' dispersal. They tested the effect of rivers, topography, roads, habitat cover, and proximity to human activity on genetic differentiation (measured by F_{ST}) to identify the landscape variables that significantly affect *V*. *variegata* dispersal in the fragmented habitat of eastern Madagascar. The study used 10 microsatellite loci for 200 adults sampled from 18 localities over an area of ~64,000 km^2. The authors identified a significant effect of geographic distance on gene flow, hence suggesting an IBD pattern (explaining 28.2% of the genetic structure). However, they also found other more important effects in their model comparisons, in particular, habitat cover and distance to nearest village had the highest explanatory inference (~52%), as compared to habitat cover only (~25%), distance to nearest village only (~22%), and IBD only (~1%). Considering the high rate of deforestation in eastern Madagascar during the last 60 years or so and the large effect of anthropogenic-related variables on the genetic differentiation of *V*. *variegata*, they proposed that rapid land cover modification by humans have driven changes in genetic structure of this species.

The CM and IBR approaches rely on *a priori* knowledge of the landscape features that may influence dispersal, which can be difficult to obtain. An interesting method that does not require prior knowledge on landscape structure was recently

proposed by Tang et al. (2020) and applied to the *P. tattersalli* genetic dataset from Quéméré et al. (2010a). They computed resistance maps, which represent how easy or difficult it is for genes to cross a particular area of the sampled region. For that they quantified deviations of pairwise genetic differentiation between all sampled locations from expected differentiation under a simple IBD model (i.e. under the assumption of a continuous and homogenous landscape). The deviations are presented on the resistance map together with the sampled individuals and thus help identify areas that are more difficult or easier to cross. In this approach, habitat features such as rivers or roads are not explicitly tested and their contribution is not explicitly quantified. However, a qualitative comparison between the mapped resistance and landscape structure can help identify putative barriers to dispersal. For instance, Tang et al. (2020) identified several sections of the Manankolana River where resistance to gene flow was high and others where it was low, suggesting that the river is not *per se* a barrier in *P. tattersalli*. Similarly, some areas of currently non-forested habitat appeared to represent a high resistance to dispersal whereas others did not. In Madagascar where habitats are increasingly fragmented this kind of approach could help visualize fine-scale resistance maps for different species across regions. This type of analysis helps identify whether it is useful to consider a whole river or "non-forested habitats" as barriers, and then test with other approaches the relative contribution of such landscape features on species genetic structure. Indeed some methods mentioned above tend to use categorical approaches where features such as "open habitat" or "river" are considered and identified or not as potential barriers to gene flow. A finer approach might be more useful, and needs to be tested and implemented across more Malagasy mammals and regions.

How we interpret genetic pattern from natural populations can depend significantly on the assumptions of the model used to analyze the data. For example, one important limitation of the two statistical approaches mentioned above (CM and IBR) is that the relative contribution of geographical and environmental distance cannot be separated. Some authors working outside of Madagascar have thus proposed the concept of 'isolation-by-environment or by-ecology' (IBE), allowing to assess the relative contribution of an environmental variable compared to

geographical distance in limiting species gene flow (Wang and Summers 2010; Bradburd et al. 2013). Madagascar with its diversity of habitats and environmental conditions should provide a wealth of study cases to which one could apply these methods. This is exactly what Scheel et al. (2015) did with *Microcebus griseorufus* populations occurring in three neighboring vegetation types within a continuous landscape. They applied the IBE model to test whether distinct genetic characteristics attributable to habitat type preference would be observed and would thus favor the hypothesis of ecological adaptation and incipient speciation. Specifically, they genotyped 76 *M*. *griseorufus* individuals at nuclear loci (AFLP, 793 fragments of 50-400 bp) and the mitochondrial *D-loop* (~690 bp) over an area of ~3 km^2. The sampled individuals were spread along two parallel 2.5 km long transects, each with the same three vegetation formations. They quantified the relative contribution of environmental distance (IBE) on the observed genetic structure of *M*. *griseorufus*, controlling for the effect of geographical distance (IBD). The mtDNA *D-loop* data did not show a significant effect of habitat ecology when controlling for geographical distance, while geography could explain ~11% of genetic variance. The nuclear loci results were consistent, suggesting that environmental differences did not reduce gene-flow in this species, at least at the geographical scale of the study. Altogether, they found that habitat types did not have a significant effect on the genetic structure of the studied *M*. *griseorufus* populations, thus rejecting the hypothesis of incipient ecological adaption.

Whereas the studies mentioned above looked at the landscape features that affect present species dispersal, it is becoming increasingly important, on Madagascar and elsewhere, to also understand past habitat connectivity. For instance, Yoder et al. (2016) studied five *Microcebus* spp. occurring in several habitat types across the island, namely, *M*. *mittermeieri* and *M*. *rufus* (moist evergreen forests of the east), *M*. *myoxinus* and *M*. *berthae* (dry deciduous forests of the west), and *M*. *lehilahytsara* (moist evergreen forests of the Central Highlands). They genotyped 30 individuals at ~7000 SNPs using ddRAD-sequencing (see Box 2), and analyzed mitochondrial data (*cox2*, *cytb*) of 332 individuals from 26 different described *Microcebus* species. The aim of the study was to use spatial genomic patterns to test three hypotheses on past landscape changes in the Central

Highlands: 1) “forest hypothesis”, 90% of the island was covered by primary forest before human arrival; 2) “grassland hypothesis”, grasslands in Central Highlands were present long before human arrival; 3) “mosaic hypothesis”, the Central Highlands were characterized by a mixture of wooded-grassland and closed-canopy forest. They built a ‘geogenetic’ map for each studied species, by drawing arrows pointing in the direction of hypothesized genetic history and admixture using the method of Bradburd et al. (2016) (see Box 4). They found a stronger geogenetic relationship between *M*. *rufus* and *M*. *berthae*, occurring respectively in the moist evergreen forests of the east and dry deciduous forests of the west, than between species of the same ecoregion (*M*. *myoxinus* and *M*. *berthae*). The higher genetic distance between *M*. *myoxinus* and *M*. *berthae* was interpreted as evidence for the presence of a biogeographic barrier (Tsiribihina River). The surprising geogenetic proximity and the estimated divergence time of *M*. *rufus* and *M*. *berthae* suggested that forest-dwelling species such as *Microcebus* could disperse across the Central Highlands throughout the Pleistocene until ~55 Ky ago. In addition, the geogenetic map of *M*. *lehilahytsara* showed that subpopulations of this species occupied two distinct geo-genetic clusters, interpreted as an indication that the Central Highlands might have experienced cyclic events of fragmentation and reconnection. Overall, their genomic data pointed towards a “mosaic hypothesis” of pre-human arrival vegetation in which Central Highlands were characterized by a mosaic of forests and wooded grassland. This is in agreement with other studies (i.e. Quéméré et al. 2012) that have increasingly shown that the “forest hypothesis” is unlikely to have been correct and that natural forces have driven patterns of forest fragmentation even long before the arrival of humans.

Studies Investigating Spatial Processes in the Recent Past

The geogenetic study of Yoder et al. (2016) provided fascinating insights into the recent past. Another approach was followed by Schneider et al. (2010), who used spatially explicit simulations to study the demographic history of *Microcebus murinus* in northwestern Madagascar (Figure 3). They sampled 195 individuals from 15 populations (Ankarafantsika National Park and forest fragments from the same inter-river system). They sequenced the mitochondrial *D-loop* (455 bp) and used,

among various methods, mismatch distribution analysis to infer the demographic history (see section on demographic history and Box 1 and Box 2). The results of the demographic analyses supported a complex model with two rather than one spatial expansion. Identifying this complex model was made possible with the SPLATCHE software (Currat et al. 2004, 2019) allowing to carry out spatially-explicit simulations, the results of which suggested that two recent expansions took place in *M*. *murinus*, one before and one after the last glacial maximum (LGM, ~25,000 years ago). Besides evidence of a recent range expansion, the authors have also shown that IBD pattern computed using populations separated by open habitat (fragmented populations) were significantly different from those obtained using only populations sampled in continuous habitat, such as in the Ankarafantsika protected area. In particular, they found that, for the same geographical distance, genetic differentiation was significantly increased across fragmented populations compared to continuous populations. They postulated that these fragments might have been connected in the recent past, hence, explaining the IBD pattern, but that recent genetic drift had increased genetic differentiation, hence, explaining the IBD differences between connected and fragmented populations. They thus noted that 1) genetic data might maintain memory of the relatively recent pre-fragmentation past; and that 2) differences in IBD might hold information on how and when fragmentation occurred.

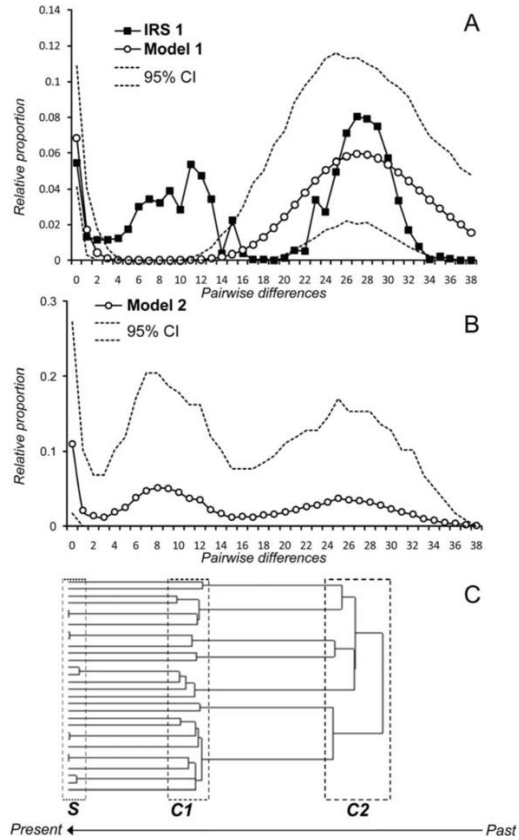


Figure 15. Inferred spatial expansion in *Microcebus murinus* from northwestern Madagascar. **A**) Comparison of the mismatch distributions obtained from the real data (mtDNA sequences sampled in the IRS1 region) against data simulated under a scenario with a single expansion (Model 1). **B**) Mismatch distribution obtained by simulating genetic data under a model with two independent spatial expansions (Model 2). **C**) One typical gene genealogy obtained under Model 2, highlighting the three main phases of coalescence (S, C1, and C2, corresponding to the so-called scattering and collecting phases, Wakeley 1999), which explain the three modes in the mismatch distribution. Figure adapted from Schneider et al. (2010).

Besides *M*. *murinus*, several genetic studies have suggested that range expansions might have taken place for many species in Madagascar, such as *M*. *tavaratra* (Sgarlata et al. 2018) or different endemic bats (*Rousettus madagascariensis*; Goodman et al. 2010; *Mormopterus jugularis*, Ratrimomanarivo et al. 2009). A well-known example of recent range expansion on the island is the invasive and often human commensal rodent *Rattus rattus*. The distribution of genetic variation in this species across the island has been investigated in part to

reconstruct the colonization history of this species and that of humans (Hingston et al. 2005; Tollenaere et al. 2010). In the most recent publication, Brouat et al. (2014) have used genetic data from 672 *R*. *rattus* individuals sampled from 35 locations widely distributed across the island. The genetic data consisted of 17 nuclear microsatellite and two mitochondrial (partial *D-loop* and *cytb*) loci. Population structure clustering method and phylogenetic analyses identified two major genetic groups (north versus rest of Madagascar, see above). Since range expansion of any species in a relatively homogeneous habitat should result in a decrease of genetic diversity as we move away from the origin of expansion, Brouat et al. (2014) tested for the rats several putative origins of expansion in Madagascar by computing, for each of them, a linear regression between genetic diversity (H_e, see Box 1) and geographic distance from the candidate sites. Two sampled sites in the southeast (Sainte Luce and Mandena) showed the strongest negative relationship between H_e and geographic distance (r ~ 12%), suggesting that southern Madagascar is the most likely region of introduction of *R*. *rattus*. The estimated area of first colonization was further corroborated by its close proximity to areas in which archaeological data suggest well-established human settlements around 1000-1500 years ago.

Altogether these studies indicate how spatial patterns can be better understood by integrating time and simple processes representing the demographic history of species. Reconstructing history is often difficult and requires specific methods. The application of such methods to Malagasy mammals is the center of interest of the next section.

Reconstructing the Demographic History of Species from Genetic Data

Detecting Changes in Population Size in Malagasy Mammals, and First Dating Attempts

Besides the study of population structure, genetic data have been increasingly used since the mid-1980s to detect, date, and quantify past demographic events such as population expansions, bottlenecks, changes in connectivity or admixture (Goldstein and Chikhi 2002; Lawler 2018). This became possible because population geneticists realized that some demographic events could potentially leave specific signatures in the genome of present-day populations (Ewens 1972; Nei et al. 1975).

For instance, Nei et al. (1975) showed that when a stationary population goes through a bottleneck, the number of alleles (n_A) and the expected heterozygosity (H_e) will exhibit different temporal trajectories during and long after the bottleneck. In other words, after a bottleneck, a population will deviate from mutation-drift equilibrium (see Box 1 and Box 2). It is for this type of signature that Brenneman et al. (2011) looked for in *Eulemur cinereiceps* in southeastern Madagascar. They used 26 microsatellite loci and tested for departures from mutation-drift equilibrium using the "Bottleneck" software (Cornuet and Luikart 1996; Piry et al. 1999). They found a significant departure in only one of the three sampled populations, which was more fragmented and impacted by human activities and cyclones. They were however unable to firmly conclude the existence of a recent or ancient bottleneck in this species of lemur.

In a similar study carried out with eight microsatellite loci on *Lemur catta* (see Sauther and Cuozzo, p. xx, for details on this species) in the southwest of the island, Parga et al. (2012) investigated departures from equilibrium using an additional method developed by Garza and Williamson (2001). They also found contrasting results depending on the population tested with significant departures for some samples under certain mutation models. Parga et al. (2012) concluded that the bottleneck signal could potentially be due to real changes in population size but they also noted that the results might also be confounded by population structure as suggested by several theoretical studies (see below and Box 1 and Box 2). Using the same approaches, Razakamaharavo et al. (2010) genotyped 15 microsatellite loci in two populations of *Varecia rubra* from Masoala Peninsula in the northeast and found significant departures from mutation-drift equilibrium in one of the two populations. Altogether, these and other studies did not provide clear signals of population bottlenecks in several lemur species. One likely reason for this is that these early approaches were not sufficiently powerful and required either a very strong bottleneck or a large number of variable loci. In other words, the statistics used were not particularly informative. Despite these limitations, there have also been cases where the results were strong and significant such as in Quéméré et al. (2012) who detected clear departures from equilibrium in *Propithecus tattersalli* using the same methods.

While this general approach was an important step in population genetics, it mainly allowed researchers to detect departures from stationarity. It did not easily allow to quantify or date the putative changes in population size. The quantification and dating of population size change events with genetic data took place thanks to theoretical work carried out in the 1990s and implemented in various approaches. These include mismatch distributions (Slatkin and 1991; Rogers and Harpending 1992) and the Bayesian skyline plots (BSP; Drummond et al. 2005), which have been extensively used on Malagasy taxa. Other approaches were based on the development of likelihood-based and Bayesian demographic inference methods (Felsenstein 1992; Beaumont 1999; Box 1 and Box 2) that have been less used in studies of Malagasy mammals.

Ratrimomanarivo et al. (2009) computed the mismatch distribution for the endemic bat *Mormopterus jugularis* using mtDNA *D-loop* sequences. They found a unimodal distribution compatible with a population expansion, estimated to have taken place between 120,000 and 230,000 years ago. Goodman et al. (2010) carried out a similar analysis based on cytochrome b sequence data for the two western Indian Ocean *Rousettus* species (*R. madagascariensis* endemic to Madagascar and *R. obliviosus* endemic to the Comoros). They found a clear unimodal distribution for *R. madagascariensis* and inferred a demographic expansion, which was poorly dated between 115,000 and 580,000 years ago. For *R. obliviosus* the mismatch distribution was multimodal and it was not clear whether the species had been stationary or subject to a more complex demographic history. The authors also tried to estimate the date of a putative expansion; the results had an even larger uncertainty.

Weyeneth et al. (2010) also used mtDNA sequence data (*D-loop* and *cytb*) from the endemic bat *Myotis goudoti*, applying spatial and demographic analyses. They identified three genetic clusters spatially distributed from the north to the south. They performed mismatch distribution analysis on the three genetic clusters, each exhibiting unimodal distributions associated with demographic expansions of decreasing age from 123,000 years ago in the north-central cluster to 66,000 and 14,000 in the southern and most southwestern population, respectively. Finally, in a more sophisticated study, Schneider et al. (2010) used several approaches to

reconstruct the demographic history of *Microcebus murinus* from two inter-river systems in northwestern Madagascar (Figure 3). They computed two statistics which are known to be sensitive to departures from mutation-drift equilibrium (namely Tajima's D and Fu's F_s; Tajima 1989a; Fu 1997) and they also plotted the mismatch distribution. What is particularly worth stressing is that the authors used spatially-explicit simulations to fit the observed mismatch distribution. This is important because most studies mainly try to fit a sudden or exponential increase under the assumption of total panmixia (Rogers and Harpending 1992). To do that Schneider et al. (2010) used the SPLATCHE software (Currat et al. 2004, 2019) and simulated scenarios of expansion of mouse lemurs over tens of thousands of years, and found that they needed to simulate two spatial expansions to explain the complex bimodal mismatch distributions found in *M*. *murinus*.

Bayesian skyline plots (BSP; Drummond et al. 2005) have also been used on several Malagasy mammals, from shrew tenrecs (*Microgale* spp., Olson et al. 2009) to mole tenrecs (*Oryzorictes* spp., Everson et al. 2018), bats (*Chaerephon* spp., Lamb et al. 2012; *Macronycteris* spp., Rakotoarivelo et al. 2019), endemic rodents (*Eliurus* spp.; Rakotoarisoa et al. 2013a, 2013b; see Carleton et al., p. xx, for details on this genus), and lemurs (*Propithecus coronatus*, Sgarlata et al. 2016; *Microcebus* spp., Blair et al. 2014; Sgarlata et al. 2018). Like mismatch distributions, BSP plots have been mainly obtained for non-primate species, probably because BSP is typically used with sequence data and was thus applied to species for which mtDNA data were the main source of genetic information. BSP plots are attractive because they visually represent the demographic history as a sequence of population size increases or decreases that is not fixed *a priori*. In other words, the method does not try to fit the model with only one or two pre-defined size changes, but to find the best number of changes that explain the data. This is different from the Beaumont methods discussed below that are able to extract more information from the data but assume models with a single monotonic size change. One of the first Malagasy studies to publish BSP plots was that of Olson et al. (2009) on two sister species of shrew tenrecs, *Microgale brevicaudata* from the northern and central western areas of the island and *M*. *grandidieri* from the central west. They found that the two species had different demographic history. *Microgale grandidieri*

exhibited a continuous but slow decrease in population size, which accelerated in the recent past and *M. brevicaudata* showed a very stationary population that exhibited a sharp increase in the recent past.

Lamb et al. (2012) also used mtDNA sequences (*cytb* and *D-loop*) to infer the history of population size changes in two bat species, *Chaerephon atsinanana* and *C. leucogaster*. They used both mismatch distributions and BSP plots. Whereas BSP plots suggested a recent population decrease in both species, the mismatch distributions were very different. For *C. atsinanana* they found a ragged distribution that could either indicate a stationary population or a complex history, possibly with population structure. For *C. leucogaster* they identified a recent population increase. There was thus disagreement with the BSP plots. One possible interpretation is that BSP and mismatch distribution may lead to different results because the different methods use different parts of the genetic information (see Box 1). Rakotoarisoa et al. (2013a) used BSP on mtDNA (*cytb* and *D-loop*) sequences from *Eliurus carletoni*, distributed in the low-altitude forests of the Loky Manambato region. They found a clear signal for an ancient stationary population that was subjected to a recent decline, possibly as a consequence of recent climate change and habitat fragmentation. These results were confirmed in another study (Rakotarisoa et al. 2013b) where the authors also studied *E. tanala* (now *E. ellermani*; Jansa et al. 2019), distributed in the same region but in forest of higher elevations, and found that *E. ellermani* exhibited a different history with no clear signal of population size change and perhaps a slight continuous increase.

These results thus appeared to confirm the idea that long-term habitat change during the Holocene led to an expansion of the distribution of *E. ellermani* and a reduction and fragmentation of the distribution of *E. carletoni*. The authors also used mismatch distributions which confirmed a possible ancient expansion for *E. ellermani*. However, for *E. carletoni* the distributions were ragged as in the bat *C. atsinanana* (Lamb et al. 2012); hence, making it difficult to make strong conclusions, despite the connection between inferred population size changes and recent Holocene habitat changes. Also, the results were in agreement with the results suggesting that *P. tattersalli* living in the same region, or the neighboring species, *P. perrieri*, had experienced recent changes in connectivity (as detailed below).

Genetic data can be difficult to interpret and compare across species. Sgarlata et al. (2018) obtained a bimodal mismatch distribution for *Microcebus tavaratra* also sampled in the Loky Manambato region. This bimodal distribution is consistent with one or several past expansions, as was found by Schneider et al. (2010) for *M. murinus* from northwestern portions of the island (Figure 3). However, the BSP plots of *M. tavaratra* were flat, indicating no signal whatsoever of population size change. Similarly, flat BSP plots were found by Sgarlata et al. (2016) in *P. coronatus* in the northwest. Blair et al. (2014), who used three nuclear genes and two mtDNA loci on *M. griseorufus* in the southern and southwestern areas, obtained BSP plots suggesting a recent population increase for *M. murinus* sampled in the southwest. This is in partial agreement with the work of Schneider et al. (2010) (see Figure 3) on this same species, who had recovered the signal for two expansions rather than one in the northwest. Finally, Everson et al. (2018) used an extension of the BSP to multi-locus data called the EBSP (Extended BSP; Heled and Drummond 2008). They identified a signal of long-term increase until around 250,000 years ago followed by a decrease since then, in the mole tenrec *Oryzorictes hova*. This species is distributed in the moist evergreen forests of Madagascar and the authors identified three genetically distinct clades, noting that they could possibly represent divergent populations in the process of speciation. The same demographic trend was found in the three clades but the cause of the decrease could not be identified, due to the large uncertainty on the EBSP plots and their inability to detect changes in the recent (Holocene) past, as the authors pointed out.

Despite their limitations, the BSP and mismatch distribution approaches have provided important insights into the demographic history of many taxa for which mtDNA sequences were the main genetic data available. Consistent expansion signals have been detected independently in *M. murinus* in the south and northwest, but also in *M. tavaratra* in the far north. In other cases, the signals were contradictory for the same species depending on the method used. In some cases, differences observed between species living in the same region might indicate different ecological requirements. This might be the case between *E. ellermani* and *E. carletoni*, who appear to prefer different forest types in different elevational zones, and these differences might be related to different demographic histories, if we

compare these rodents to *M. tavaratra* that may have recently arrived in the Loky Manambato region. There is a need to bring all these studies together and perhaps construct a comparative framework, even if some of the data are preliminary at this stage. This should become increasingly possible since certain regions of the island, including Ankarafantsika, Kirindy CNFEREF, and the Loky Manambato region, are starting to have sufficient data thanks to long-term studies. In the next section we bring together studies that have used more advanced methodologies for demographic inference.

Dating Past Population Size Changes in Malagasy Mammals in Relation to Natural Habitat Change and Anthropogenic Activities

Demographic inference significantly changed after the article of Felsenstein (1992), who argued that by using summary statistics calculated from genetic data, population geneticists were throwing away most of the information available in the genealogy of genes. He noted that if information could be extracted from the data using likelihood-based methods it would significantly increase our ability to infer the evolutionary history of species (Box 1 and Box 2). Olivieri et al. (2008) were the first to apply likelihood-based methods to lemurs, specifically the methods of Beaumont (1999) and Storz and Beaumont (2002) (hereafter referred to as the Beaumont methods), which were used to detect and date population size changes in orangutans (Goossens et al. 2006). Olivieri et al. (2008) identified, dated, and quantified a major bottleneck in three *Microcebus* species from northwestern Madagascar, which suggested that the three species had a much larger population in the recent past, possibly as a result of significant habitat loss and fragmentation. The best supported estimate suggested a major demographic event in the last 500 or 900 years. The three species showed similar results despite their non-overlapping distributions and differences in terms of population structure. In a similar study carried out in the same region, Craul et al. (2009) also found a strong and recent bottleneck signal for *Lepilemur edwardsi*, which most likely started during the last 1000 years. The generation time assumed for mouse lemurs by Olivieri et al. (2008) was one year, whereas recent studies (Radespiel et al. 2019) suggest that a more realistic value might be between two and three years (see also Yoder et al. 2016).

Despite the complexities of determining a consistent measure of generation time – a problem no doubt related to poorly known life-history strategies – the dating of the demographic event detected in these *Microcebus* species could thus be older than dates suggested originally by Olivieri et al. (2008). Nonetheless, the age would still be reasonably recent, likely within the last 1000 to 3000 years.

Altogether these two studies strongly suggested that the signal for a bottleneck found in four species from the northwest was possibly related to major habitat changes in the recent past, possibly as a consequence of human activities (see Godfrey and Douglass, p. xx, for further details on habitat change in this area of Madagascar). This appears to be the first clear genetic indication that humans might have been responsible for habitat loss generating a massive reduction in lemur population sizes. These results should also be integrated with the results of Schneider et al. (2010) suggesting two major population expansions of *M. murinus* in the same region but in a much more ancient period and on the basis of mtDNA data.

In northern Madagascar, genetic studies seem to indicate a rather different scenario, with evidence of more ancient habitat loss and fragmentation compared to the northwest. In a study carried out in Loky Manambato region, Quéméré et al. (2012) used the same likelihood-based methods on genetic data of *Propithecus tattersalli*. They also detected a bottleneck signal but with a strong support for an ancient and likely pre-human arrival habitat change, possibly between 4000 and 10,000 years ago. Since it has been shown that population genetic structure can confound demographic inference (Chikhi et al. 2010; see Box 2), Quéméré et al. (2012) also tested the effect of population structure on *P. tattersalli* results, concluding that population structure was unlikely to have generated the detected bottleneck signal, even if it might have contributed to it. Additionally, by analyzing satellite pictures, they found that forested habitats had decreased by only 2% in the previous 50 to 60 years in the Loky Manambato region. Therefore, they concluded that the fragmented habitat of the Loky Manambato, which are topographically and geologically complex (Goodman et al. 2018) was likely natural and ancient, making this region rather different from other areas of the island, where deforestation has clearly been very high in the last decades (Vieilledent et al. 2018). There is

increasing evidence that humans arrived in Madagascar much earlier than was thought when these studies were performed (see Wright and Rakotoarisoa, p. xx; Godfrey and Douglass, p. xx).

In a more recent study, Salmona et al. (2017) extended the work of Quéméré et al. (2012) by applying several methods of demographic inference to *P. tattersalli* and to *P. perrieri*, the latter lives in the neighboring Andrafiamena region. The distribution of *P. perrieri* is currently notably smaller and more fragmented than that of *P. tattersalli*. The results of this study are too complex to summarize here due to the large number of methods used, but they are generally consistent among each other, although not always in agreement regarding the dating of events. Again, this is likely due to the fact that the methods used make different assumptions and summarize genetic information in different ways. Salmona et al. (2017), following the same methods as Quéméré et al. (2012), found a major signal for a bottleneck that started at least 3000 years ago and possibly even before 10,000 years ago, depending on the generation time assumed. The results were thus consistent with Quéméré et al. (2012), strengthening the conclusions of the previous study. However, Salmona et al. (2017) were uncertain whether the timing was the same in the two species given the uncertainties of the generated dates. What makes this study unique is its use of ABC (approximate Bayesian computation, Box 2) framework to compare a large number of alternative models that incorporated both population size change and fragmentation. Altogether these authors compared 14 demographic models with and without population size change, with and without ancient population structure, and with and without recent changes in connectivity. They found that the best model explaining the genetic data for *P. perrieri* was one in which populations were structured in the past with an increase in habitat loss and fragmentation in more recent geological time. In the case of *P. tattersalli*, the same model best explained the recovered patterns, but could not distinguish it from a model in which habitat had been more continuous in the past. They could thus identify an increase in fragmentation in both species, but for *P. tattersalli* the habitat might have been more continuous than for *P. perrieri*. Due to the imprecise dating and that the date of human arrival on the island has been pushed deeper into time, they were unable to exclude human activities as a possible cause of the recent

increase in habitat fragmentation. Altogether, Salmona et al. (2017) provided the first study where complex demographic histories, incorporating population size changes, population structure, and changes in habitat connectivity, could be compared.

The population genetic studies discussed above are important because they provide insights on the relatively recent evolutionary history of species, which in turn have important implications for conservation purposes. They are based on the implicit or explicit idea that one could in principle study past changes in habitat by inferring the demographic history of the species living today in these habitats. Currently, the genetic studies suggest that the northwest of the island, specifically the Ankarafantsika and neighboring areas regions have likely had different histories of habitat fragmentation from those of the northeast, namely the Loky Manambato and Andrafiamena regions. In the first case, genetic data suggest a recent and major loss of habitats, whereas in the second case, habitats may have been fragmented over much longer periods and humans may have only contributed in a limited way to that process (excluding the last decades). This second scenario is in agreement with the geogenetic study of Yoder et al (2016) discussed above.

Interpreting genetic data must be done with caution. Linking population sizes estimated by some methods to changes in the dynamics of habitat connectivity is not simple. For instance, most methods that infer population size changes explicitly assume that population structure can be ignored (see Box 2). An implicit assumption is that if we detect, say a 10-fold size change in population size, it will seem unlikely that population structure could be responsible for a change of this magnitude. There are two problems with that assumption. First, there is an increasing recognition that population structure can actually generate major signals of change in population size even without any real change in population size. In other words, our intuition is simply not good enough and most methods aiming at detecting population size changes, while ignoring population structure (Beaumont 1999; Drummond et al. 2005; Li and Durbin 2011), may actually be misleading (Beaumont 2004; Chikhi et al. 2010; Mazet et al. 2016). Second, if we wish to understand how human activities or environmental factors, such as natural climate change, have influenced ancient patterns of habitat fragmentation in Madagascar, we need also to understand how

changes in population connectivity influence genetic diversity, differentiation, and demographic inference. Thus, ignoring population structure is unlikely to be correct for many species.

This does not mean that the results presented above are incorrect and that no population size changes took place in the presented case examples. It only means that the interpretation of genetic data is more challenging than often assumed, and that the same data may be explained by rather different scenarios. It is difficult to imagine a method that would account for all the complexities of demographic histories, including population and social structure, population size changes, changes in gene flow, and possible population splitting or extinction. It seems thus important to be modest and rather base our conclusions on the comparison of alternative scenarios relying on meaningful biological, ecological, paleoclimatological, and archaeological data. This is where ABC methods (see Box 1 and Box 2; Beaumont et al. 2002; Beaumont 2010) provide one possible solution, as we have seen with the Brouat et al. (2014) and Salmona et al. (2017) studies. The great advantage of ABC is that it can be applied to nearly any demographic model of interest. As a consequence, ABC can be used for model choice, although it can be computationally demanding with detailed genomic data and very complex models.

Finally, another word of caution is necessary when interpreting genetic data. An important but too often neglected issue is that of sampling. Many population genetics methods assume that the samples analyzed are taken from an isolated random mating population. However, obtaining samples from rare and threatened species is not a simple matter. Moreover, as we have seen for many Malagasy mammals, genetic samples can be obtained in different portions of the island, based on a limited number of individuals from non-adjacent locations. This means that some of the mismatch distributions or BSP plots discussed above were generated by pooling samples from distant locations. It is increasingly recognized that sampling methodologies can generate significant signals of demographic changes in genetic data (Wakeley 1999; Chikhi et al. 2010, 2018; Heller et al. 2013; Mazet et al. 2016), or in the analysis of genetic diversity -- see ascertainment bias in Box 1.

Studies Using Genomic Data to Detect and Date Ancient Evolutionary Events

The Pairwise Sequentially Markovian Coalescent (PSMC) method of Li and Durbin (2011) is probably the most popular method both in and outside Madagascar. It allows to reconstruct a history of population size change from the genome of a single individual and it is thus well adapted to Malagasy and other non-model species for which full genomes can only be obtained for a limited number of individuals. The method has already been applied to several lemur species such as *Eulemur macaco* and *E*. *albifrons* (Meyer et al. 2015), *Prolemur simus*, *Microcebus murinus*, *Propithecus verreauxi* (Hawkins et al. 2018), *Cheirogaleus* spp. (Williams et al. 2020), and *Microcebus* spp. (Poelstra et al. 2020).

In Figure 4 we show the PSMC plots obtained for *Cheirogaleus* spp., which are known to hibernate, and analyzed by Williams et al. (2020). The figure also shows patterns of diversity as measured by heterozygosity along the genome in the four individuals representing the different species. The four PSMC curves are different from each other, suggesting that the demographic histories of each taxon are also different from one another. For instance, Williams et al. (2020) interpreted the curves as representing changes in effective population size through time and noted that whereas three curves suggest large population sizes in the last 200,000 years, this was not the case for the *C*. *medius* (blue) PSMC plot. This can also be seen by the fact that the heterozygosity distribution shows a strong bias towards low values. *Cheirogaleus major* (red) and *C*. *medius* (blue) have very large geographical distributions, whereas the other two species are only known from a few locations. A further complication is that the genotyped *C*. *medius* individual was sampled outside the species' previously known range in a small forest within the distribution of *C*. *major*. Finally, these curves could also be interpreted in a population structure framework (Mazet et al. 2016), suggesting perhaps an increase in fragmentation rather than major changes in population size across the complete distribution of these taxa. While much work is needed to analyze more samples from other regions, we stress that this is the first study of its kind as it also tried to also detect potential admixture events between the sampled species. The results suggested that there may have been an admixture event between *C*. *medius* and both *C*. *major* and *C*. *sibreei* around four Mya. The authors also suggested that genes involved in

hibernation might have been introduced between these species during the admixture events, but the proportion of such genes among the introgressed genes was low.

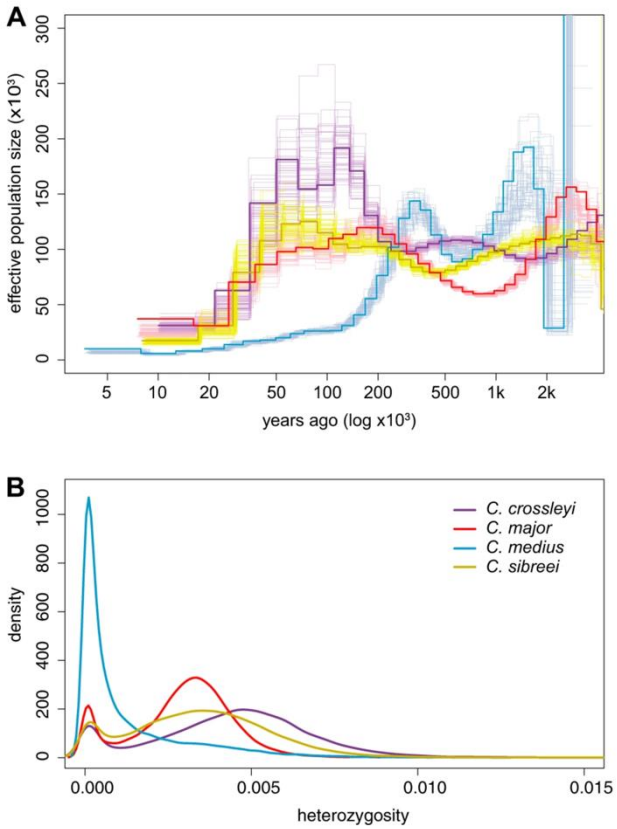


Figure 16. Demographic inference and genome-wide heterozygosity distribution of individuals of *Cheirogaleus* spp. **A**) Inferred population size changes (N_e) using single diploid genome (PSMC method, see main text) of each studied species (colors indicated in panel B). For each species, bold lines indicate the inferred population size, whereas thinner lines correspond to bootstrapped PSMC curves. **B**) Genome-wide distribution of heterozygosity, computed across the genome over 100-kb windows. Figure adapted from Williams et al. (2020).

We are still in a period in which such methods are experimental and results are preliminary. The important aspect is that the Williams et al. (2020) study was the first one that tried to use a rather complex set of inference methods. A subsequent

publication by Poelstra et al. (2020) went further in the sense that these authors not only described a new species (*Microcebus jonahi*) but also used a variety of methods to recover the demographic history of *Microcebus* spp. from eastern Madagascar. Some of the methods used were based on RAD-seq data and used the AFS (see Box 1) but the authors also used the PSMC method. They found an apparent decrease in population size using both the PSMC and AFS-based method (Gronau et al. 2011).

The PSMC method assumes that the individual was sampled in a non-connected panmictic population and that population structure is negligible. However, Mazet et al. (2016) have shown that the PSMC is profoundly influenced by population structure and that even if population size is constant, population structure will unavoidably generate a major signal of population decrease in the recent past. In addition, they showed that if migration rates (connectivity) changed in the past this should necessarily generate humps similar to those found in real PSMC curves such as those represented in Figure 4 and those observed in the other lemurs (Meyer et al. 2015; Hawkins et al. 2018). Given that habitats have been subjected to complex cycles of connectivity and disconnection, and that these have also been accompanied by changes in population size, it is difficult to determine the extent to which the genomic patterns uncovered with the PSMC method should be interpreted as indicating changes in connectivity. Chikhi et al. (2018) proposed a general framework allowing researchers to construct alternative models of great complexity with both population size changes and changes in connectivity but this framework does not allow to automatically infer the models. There is still no simple manner to identify the best model.

While these comments may seem technical, they are sufficiently general to be mentioned. Data on a genomic scale are just starting to be applied towards a better understanding of Madagascar's biodiversity and the methods adapted to the corresponding large data sets are being developed as we write this contribution. The advantage of genomic data is that they provide large amounts of information and thus allow researchers to infer demographic parameters with statistical confidence. However, this also comes at a cost. First, the analysis of genetic data is computationally expensive and technically challenging, making it difficult for

people outside the field to contribute critically to the interpretation of the data. Second, this selects for new types of students who will be less and less able to physically conduct fieldwork as they will become increasingly tied to the computers required to manipulate the large text files needed to analyze. Third, the complexity of the many program scripts required to analyze and filter the data is such that it is very difficult to test and validate properly the results obtained by increasingly specialized computational biologists. Fourth, using large amounts of data under the wrong model may generate very precise estimates of potentially misleading parameters. A common saying in statistics is that “all models are wrong but some are useful”. The reason for this is that models are simplified versions of reality stressing one or several aspects of that reality. They are constructed for a hopefully clear goal and should only be used beyond that if it can be proven that they are indeed applicable. Unfortunately, scientists working in these and other domains tend to believe very strongly the methods they use and the underlying models are correct. They often forget the assumptions that were required for the cherished conclusions to hold. Genetic and genomic data, together with models and inference methods are fundamental to help make sense of a complex world and to make the best use of the data. However, models and data need to be interpreted with an open mind and in an increasingly comparative framework. There are still few published studies and Ph.D. theses using genomic data from Malagasy mammals and aiming at reconstructing the demographic history of species.

Taxonomy, Mammal Species Diversity, and Hybridization Studies

Taxonomy and Mammal Species Diversity

Genetic data have significantly changed our understanding of Malagasy mammals not just at the local, landscape or species scale, as we have seen in the previous sections. These data have been increasingly used, often with morphological and fossil data, to reconstruct evolutionary relationships among Malagasy mammals, clarifying the phylogenetic position of diverse mammals at different taxonomic levels, such as the family allocation of *Daubentonia madagascariensis* (see Sterling et al., p. xx); identifying the Malagasy Carnivora as unique radiation to the island and showing remarkable morphological convergence with other members of the

order found elsewhere in the world, determining the monophyly of all Malagasy land mammal orders, and dating speciation events or periods of diversification. These issues are discussed in other contributions herein and will thus be only briefly presented here.

While molecular data have generally helped answer important evolutionary questions, they have also generated controversies, particularly regarding the number of lemur species that have been identified in the last two decades. For instance, Tattersall (2007) questioned the validity of many of these taxa, noting that the number of lemur species had increased from 36 to 83 since 1982, even though few new populations of lemurs had been discovered during that period. A few years later, Markolf et al. (2011) made a similar statement, noting that the number of newly described species had been on the order of 4.5 per year for the period 2006-2011 and pointed out that in too many instances the new species had been described on the basis of a few individuals. These authors suggested that genetic differences between individuals from different species were in some cases smaller than between individuals within a previously defined species. For instance, they noted that Louis et al. (2008) estimated a 3.7% divergence between *D-loop* sequences of individuals of *M*. *margotmarshae* and *M*. *mamiratra*, while Fredsted et al. (2004) had identified much higher divergence (~8.2%) between the most divergent sequences within the *M*. *murinus* clade. It is important to note that *M*. *murinus* has the widest distribution of any member of this genus and Hotaling et al. (2016) later described two new *Microcebus* species from individuals previously identified as *M*. *murinus* (*M*. *ganzhorni* and *M***.** *manitatra*) (see Radespiel et al., p. xx, on this genus).

This tendency has continued in the last decade and has been criticized by Tattersall and Cuozzo (2018) who considered that about half of the currently identified species should possibly be merged with previously described species. The number of recognized *Microcebus* species has significantly increased from only two species in the 1990s (*M*. *rufus* and *M*. *murinus*) to ~25 today (Petter 1962; Poelstra et al. 2020). Similarly, there are currently 26 recognized *Lepilemur* species (Louis et al. 2006; Lei et al. 2017), when only one species with six subspecies were accepted by Tattersall (1982). Other genera, such as *Propithecus*, have not seen a proliferation of new species, but rather former recognized subspecies being

elevated to full species (Mittermeier et al. 2006). Markolf et al. (2013) have defended the idea that the subspecies from the brown lemur complex, *E*. *fulvus*, should be considered as different species, justifying this decision by the significant morphological and genetic differences identified between them.

The regular increase in the number of lemur species is probably the result of different factors, among which genetic data played a major role. First, a certain number of new species are actually the result of populations or subspecies already known and being redescribed as new species (see Tattersall 2007). Second, the increased use of molecular data (mainly mtDNA) and the application of the "phylogenetic species concept" (sensu Cracraft 1989) has led to the identification of monophyletic groups, which were often not diagnosable based only on morphological data. These monophyletic groups that could not be allocated to recognized taxa were considered as new species. Third, from a conservation point of view, the IUCN currently recognizes the species as the basic unit of conservation. It is possible, though difficult to determine with certainty, that a form of conservation politics (and perhaps a cautionary principle) might have led to the "upgrade" of divergent populations, further stressing the critical need to preserve Malagasy forests. It is certainly better to upgrade now a subspecies or divergent population and merge taxa later than realize, once it is too late, that the taxon or population in question indeed represents a now extinct species.

In some cases, genetic information have not always led to an increase in the number of recognized species. Brennemann et al. (2016) used genetic data and concluded that the *Indri indri* should not be divided into subspecies as suggested on the basis of pelage differences. Recently, Poelstra et al. (2020) used genomic data together with complex demographic analyses, as well as morphological data (Schüßler et al. 2020), to identify a putative new *Microcebus* species in eastern Madagascar, named *M*. *jonahi*. Specifically, Poelstra et al. (2020) used criteria based on the multispecies coalescent to test whether the newly discovered lineage was a species or a divergent population. While the criteria allowed them to describe *M*. *jonahi* as a new species, the same criteria also suggested that two currently identified species (*M*. *mittermeieri* and *M*. *lehilahytsara*) might actually be divergent populations of the same species. Applying this general multispecies coalescent

framework to all mouse lemur and sportive lemur species described in the last two decades might be a useful test to confirm their taxonomic status or their synonymy as suggested by Tattersall and Cuozzo (2018). There is a serious need to develop methods that can help determine more objectively whether divergent lineages should or not be identified as new species and this procedure has important consequences for conservation priorities.

During the last two decades, field surveys and museum-based research have greatly contributed to a better understanding of the endemic rodents (subfamily Nesomyinae) biodiversity on the island with nine genera being recognized (see Goodman et al., p. xx). Genetic data have now well established that Malagasy rodents came from Africa during a single colonization event (see Jansa and Carleton, p. xx) and these data have also been used to describe new species or clarify taxonomy (e.g., Jansa et al. 1999, 2019).

The Malagasy Carnivora, belonging to the endemic family Eupleridae, currently include seven genera with eight species (see Veron et al., p. xx). There have been debates on whether these animals were monophyletic and the result of a single colonization event, due to their morphological resemblance with other Carnivora elsewhere in the world. In the first comprehensive genetic study of these animals, Yoder et al. (2003) investigated their molecular phylogeny using four genetic loci (two mtDNA and two nuclear loci), and provided clear evidence on their monophyly. However, the taxonomy within Malagasy Carnivora is still debated, mostly due to the limited sampling of certain putative forms (Veron et al. 2017, 2018). For example, Durbin et al. (2010) described a new species, *Salanoia durrelli*, sister-species of *S. concolor* (see Farris and Murphy, p. xx), on the basis of morphological divergence, despite a low level of *cytb* genetic divergence (0.8%). In a later study, Veron et al. (2017) increased the number of *S. concolor* sampled individuals, while only having access to one sample of *S. durrelli*, and used a higher number of genetic markers (two mtDNA loci and one nuclear locus). They confirmed the generally low levels of genetic divergence between *S. durrelli* and *S. concolor* (0.2% - 1.2 %) and concluded that the two taxa should be merged. In the same study, two species of the genus *Galidictis* (see Farris and Goodman, p. xx) were also merged as the authors found similarly low levels of *cytb* divergence (1.1%-

1.2%). Veron et al. (2018) also noted considerably low levels of genetic divergence (0%-1.4%) between specimens allocated to *Eupleres majori* and *E*. *goudotii* (see Rivera et al., p. xx), two previously described species differentiated based on morphology (Goodman and Helgen 2010). Overall, these examples show how more work, sampling, and morphological and genetic data are still needed to robustly validate the species diversity within the endemic family Eupleridae.

Until about two decades ago, members of the family Tenrecidae, generally referred to as tenrecs, were originally associated to shrews, moles, and other insectivores (order Lipotyphla) and previously to the order Insectivora due to their apparent morphological similarities. Molecular studies have played a major role in clarifying their evolutionary history and taxonomy, showing that tenrecs are genetically closer to other African mammals belonging to the superorder Afrotheria than the two orders to which they have been previously placed (Springer et al. 1997; Stanhope et al. 1998; Mouchaty et al. 2000; Meredith et al. 2011). Everson et al. (2016) in the most recent and the largest molecular phylogenetic study of Tenrecidae, with samples from all species, used 11 nuclear genes (~9600 bp). They suggested to place the *Limnogale* within *Microgale* and place two species (*M*. *dobsoni* and *M*. *talazaci*) in Nesogale. With 31 Malagasy species recognized, tenrecs thus exhibit a species discovery rate of about one new species every two years since 1992, with 20 species having been described during this period.

Hybridization studies

Most of the articles mentioned in the previous sections assume that species are clearly identified and represent taxonomic units that are reproductively isolated from other related taxa. As a consequence, most studies tend to measure genetic differentiation between populations from the same and supposedly clearly identified species, or computed genetic distances and phylogenetic relationships between well separated species. However, several studies have also identified areas of putative hybridization within several Malagasy genera, such as the genus *Eulemur* (Rabarivola et al. 1991; Lehman and Wright 2000; Wyner et al. 2002; Pastorini et al. 2009). At this stage and given the major changes that have happened in the last 20 years in the taxonomy of lemurs and the fact that genetic studies are still

taxonomically and geographically limited, we can fairly say that hybridization analyses are very limited.

Wyner et al. (2002) worked at Andringitra in the south central portion of the island, where most individuals from one *Eulemur* population were identified as hybrids between *E*. *albocollaris* (now *E*. *cinereiceps*) and *E*. *fulvus rufus* (now *E*. *rufus*). Gligor et al. (2009) and Hapke et al. (2011) identified and studied a hybrid zone in the south between *Microcebus murinus* and *M*. *griseorufus*. Pastorini et al. (2009) identified hybridization between *E*. *fulvus* and *E*. *mongoz* using microsatellite and mtDNA data in the northwest and they identified six *E*. *mongoz* individuals with putative *E*. *fulvus* alleles introgressed in their genotype and one *E*. *mongoz* individual with the mtDNA haplotype of *E*. *fulvus* (Figure 5). More recently, Williams et al. (2020) used full genome sequences from dwarf lemurs (*Cheirogaleus medius*, *C*. *major*, *C*. *crossleyi*, *and C*. *sibreei*) and identified putative events of ancient admixture, about 4 Mya, between *C*. *medius* and both *C*. *major* and *C*. *sibreei*. They interpreted these admixture events as being related to adaptation to hibernation. However, they also noted that hibernation was ancestral in *Cheirogaleus* (who diverged from a common ancestor ~18 Mya) and that there was no particular high representation of hibernation genes in the genomic regions that appeared to be introgressed. This would definitely need further research. Another recent study by Everson et al. (2020a) also claimed to have identified ancient admixture events between two shrew tenrec species (*Microgale soricoides* and *M*. *fotsifotsy*).

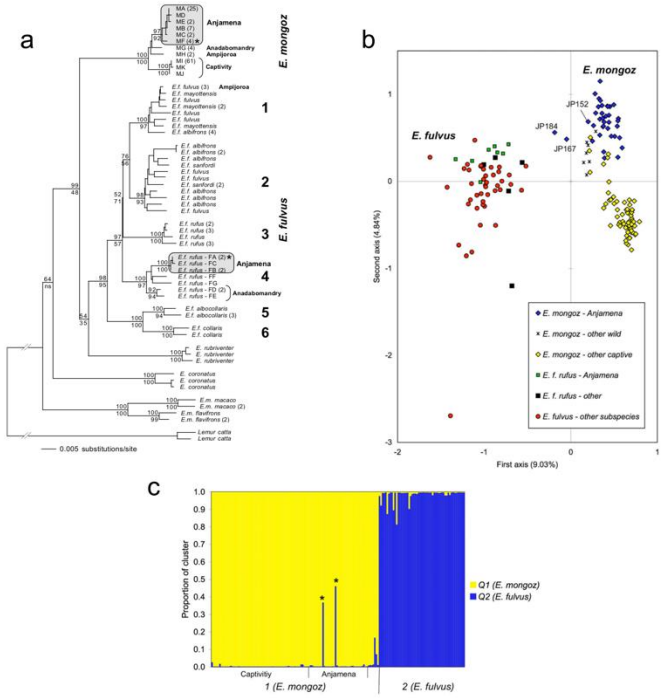


Figure 17. Hybridization analysis on *Eulemur mongoz* and *E. fulvus* sampled at Anjamena, western Madagascar. **A**) Phylogenetic tree inferred from D-loop and ND4 loci of the mtDNA. The asterisks (*) show the two mtDNA haplotypes of *E. mongoz* and *E. fulvus* that were also found in other species. **B**) Principal component analysis of individual *E. mongoz* and *E. fulvus* genotypes highlighting the two hybrids (JP167 and JP184) between *E. mongoz* and *E. fulvus*. **C**) STRUCTURE analysis of sampled individuals of *E. mongoz* and *E. fulvus*. The two hybrids (JP167 and JP184) are marked with * and show mixed genetic ancestry with each of the two *Eulemur* species. Figure adapted from Pastorini et al. (2009).

In addition, putative hybridization has been described for other species but was not necessarily supported by genetic data. Hybrids between *Eulemur f. fulvus* and *E. f. rufus* were documented by Lehman and Wright (2000). *Propithecus coronatus* and *P. deckenii* are thought to hybridize in the central western portions of the island (Thalmann et al. 2002; Rakotonirina et al. 2014), and *E. macaco* and *E. flavifrons* are also thought to hybridize in some forest fragments in the northwest (Rabarivola et al. 1991). We may also mention the study of Djelati et al. (1997) who carried out a meiotic study on seven male hybrids of different *Eulemur* species, three of which were infertile.

Hybridization studies of Malagasy mammals are notably limited, and this line of research needs a more solid theoretical basis, as well as more field work if we want to understand the distribution of hybridization in Malagasy mammals. Hybridization events, whether recent or ancient, are by nature dynamic, and require questions to be posed about species identity, taxonomy, and phylogeny. Clarifying the evolutionary processes taking place in contact zones between closely related or diverging taxa would be very important for evolutionary biologists but also to conservation programs. The very rapid habitat changes taking place in the forests of the island will likely change the distribution of many species and in some cases reduce inter-species secondary contact (and thus hybridization), but may also generate new zones of contact between animals of closely related species. For instance in a theoretical sense, if conspecifics are rare within an isolated forest fragment, it is not necessarily far-fetched to imagine that such a situation could favor hybridization between sympatric closely related species.

Conclusion

In this contribution, we tried to summarize insights that have been gained on Malagasy mammals based on recent genetic studies. Other contributions in this chapter look in greater detail at certain groups and at different taxonomic levels. However, while this current contribution is extensive, it was not possible to address all the different developments over the past decades. For example, we have acknowledged the importance of karyological studies in early taxonomic and phylogenetic studies and in more recent research (e.g., Andriaholinirina et al. 2006). While not much has been presented herein about karyological studies, we would not be surprised if they made a comeback in the next few years as our understanding of structural variants increases for non-model species. We have also given little space to studies aiming at detecting selection or adaptation (but see Box 5), and those related to developing technical aspects of non-invasive sampling (e.g., fecal samples) for genetic studies. Nevertheless, we have tried to mention some of these studies, which represent a minority of research studies but may become more common in the future as the importance of linked selection is better understood.

Considering that there are 200-250 mammal species living on the island, we can state that most have not yet been properly studied genetically. In most other groups of Malagasy vertebrates, invertebrates, and plants, the situation is even less advanced, but research teams are actively working to rectify the situation. For most mammal species, for which genetic data are available, they are often limited to mtDNA sequences that were used to address taxonomical or phylogenetic questions, and sometimes pooled to carry out some demographic inference studies.

We also tried to convey the fact that we are in a very exciting period in which genomic data are increasingly easy to obtain. The recent work of Blanco et al. (2020) is for instance an interesting case wherein it was shown that the possibility of next-generation methods could in principle be used in the field for research on Madagascar. Similarly, Guevara et al. (2018) developed a simple and cheap (economically and ecologically) protocol to carry out genetic typing in the field and tested on *Propithecus verreauxi* fecal samples from the southwest of the island. It is too early to make strong claims about such studies and their use in a larger context of biodiversity loss but it is important to stress that there is a need to move the analyses of genetic data from outside to within Madagascar. From the perspective of conservation, we should also reflect on the fact that many research activities that rely on laboratory analyses or computing work, including genetic and genomic studies, are typically polluting with respect to required chemicals and electronic resources, that can unfortunately be important. There is thus a need for our community to perhaps balance the gains and the costs that our activities generate.

Our review also showed that the statistical methods used to analyze genetic data are becoming increasingly sophisticated, hence, allowing us to ask more complex questions. However, we have also stressed that genetic and genomic data are not always easy to use and interpret. The information brought forth by these associated techniques is often counterbalanced by their magnitude and complexity. Nowadays, genomic analyses require the users to go through various filtering and computation programs that cannot be validated “by hand”. This means that errors are likely to be published as less and less people are able to analyze, check, and test the results. Also, the interpretation of data remains a crucial issue for which

there is no simple solution. Training students to analyze the new (big) genomic data sets will be a major challenge that requires understanding the theory and the assumptions behind the models. In a recent review, Allendorf (2017) wrote that population genetics had moved from a period in which there was a lot of theory and limited data to a period in which we have loads of data and too little theoretical understanding and he added (p. 26) that “current training focuses too much on technique” and that “even the most basic aspects of population genetics theory are commonly misunderstood or overlooked”. We have tried to avoid the over interpretation of the genetic data but cannot guarantee that we did not fail. We hope that we managed to provide tools for students to look critically at such data and re-interpret themselves the results obtained by others.

Genetic data are often seen as being more “objective” than other types of data. It may be true to some extent since we do understand how genes are transmitted from generation to generation. However, there are often several competing interpretations or models that could explain the same genomic data, even when authors only discuss one of them. The danger with genomic data is that they tend to convey a misleading overconfidence on statistically significant signals. For instance, Chikhi et al. (2010) showed that simulated genetic data under a model of population structure and constant size, the detection, dating, and quantification of population size changes will increase with increasing data, even though no population size change was simulated. In other words, the more data we analyze, under a model whose assumptions are not met, the more confident (and precise) we will be on the inference of a demographic event that may never have taken place (see also Philippe et al. 2011 on similar issues in phylogenetics). The PSMC plots discussed above are an example of demographic inference based on genomic data that could either identify very clear and undisputable population size changes under one model (no population structure), or actually be generated under a model with changes in connectivity and absolutely no change in population size. These two interpretations are simplistic but the point is that they are profoundly different and difficult to separate.

Conservation genetics has contributed to the development of theoretical population genetics, and as such it has enriched the theory of evolution. For

instance, this field of research played a major role in the study of small and isolated populations (Lynch et al. 1990), where deleterious mutations are expected to accumulate as a consequence of genetic drift and inbreeding. The observation of genetic defects in various endangered species that went through serious bottlenecks (e.g., Roelke et al. 1993; Munson et al. 1996; Madsen et al. 1999), and the genetic restoration of some of these isolated populations by the re-introduction of individuals from other areas have demonstrated that the theory could be successfully used for practical short-term conservation projects (Madsen et al. 1999; Hostetler et al. 2012). Unfortunately, we are not aware of solid research in these areas carried out in Madagascar, while we may soon be in the situation where we will need people able to analyze genetic data from species with increasing genetic defects.

We however strongly defend the use of genetic data, despite their current limitations. Genetic data have played and will continue to play a central role in clarifying major questions related to Malagasy biodiversity across geographical and temporal scales. They will definitely do so for some time to come, and the next few years will, without any doubt, bring exciting new insights for an increasing number of mammals. We need these data if we are to say anything of relevance regarding Malagasy biodiversity.

Acknowledgements

We are grateful to the Government of Madagascar for permission to conduct research on the island and specifically to the Direction Générale du Ministère de l'Environnement et des Forêts de Madagascar (Région Sava), Comité Ad Hoc pour la Faune et la Flore et le Comité Organisationel pour la Recherche Environnementale (CAFF/CORE), the Département de Biologie Animale et Ecologie de l'Université de Mahajanga, the Département de Biologie Animale de l'Université d'Antananarivo and the Département de Biologie de l'Université d'Antsiranana, and the NGO Fanamby NGO for their support over the years. We also would like to thank Guillaume Besnard, Filipa Borges, Jörg Ganzhorn, Steve Goodman, Richard Lawler, Beatriz Mourato, Bárbara Parreira, Ute Radespiel, Jordi Salmona, Rémi Tournebize, and Anne Yoder, for their constructive, and useful comments that significantly helped us improve the manuscript. This work was supported by the Fundação para a Ciência e Tecnologia (PTDC-BIA-EVL/30815/2017 to LC; PD/BD/114343/2016 to GMS), the 2015-2016 BiodivERsA COFUND with the national funders ANR (ANR-16-EBI3-0014) and the Fundação para a Ciência e Tecnologia ref. Biodiversa/0003/2015 and PT-DLR (01LC1617A), the LABEX entitled TULIP (ANR-10-LABX-41 and ANR-11-IDEX-0002-02), the LIA BEEG-B (Laboratoire International Associé–Bioinformatics, Ecology, Evolution, Genomics and Behaviour), and the Investissement d'Avenir grant of the Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01).

References

Al-Asadi, H., Petkova, D., Stephens, M., and Novembre, J. 2019. Estimating recent migration and population-size surfaces. *PLoS Genetics* 15: e1007908.

Aleixo-Pais, I., Salmona, J., Sgarlata, G. M., Rakotonanahary, A., Sousa, A. P., et al. 2019. The genetic structure of a mouse lemur living in a fragmented habitat in northern Madagascar. *Conservation Genetics* 20: 229-243.

Allendorf, F. W. 2017. Genetics and the conservation of natural populations: Allozymes to genomes. *Molecular Ecology* 26: 420-430.

Anantharaman, R., Hall, K., Shah, V. B., and Edelman, A. 2020. Circuitscape in Julia: High performance connectivity modelling to support conservation decisions. *Proceedings of the Julia Conferences* 1: 58.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., and Hohenlohe, P. A. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* 17: 81-92.

Andriaholinirina, N., Fausser, J.-L., Roos, C., Zinner, D., Thalmann, U., et al. 2006. Molecular phylogeny and taxonomic revision of the sportive lemurs (*Lepilemur*, Primates). *BMC Evolutionary Biology* 6: 17.

Andrianaivoarivelo, A. R., Shore, G. D., McGuire, S. M., Jenkins, R. K. B., Ramilijaona, O., Louis, E. E., and Brenneman, R. A. 2009. Characterization of 22 microsatellite marker loci in the Madagascar rousette (*Rousettus madagascariensis*). *Conservation Genetics* 10: 1025-1028.

Arnaud, J., Meier, B., Dugoujon, J. M., and Rumpler, Y. 1992. Study of the variability of erythrocyte enzymes in captive and wild populations of the black lemur (*Eulemur macaco macaco*). An indispensable preliminary in captive breeding programmes. *Primates* 33: 139-146.

Avise, J. C. 2000. *Phylogeography: The history and formation of species*. Cambridge, MA: Harvard University Press.

Bachmann, L., Rumpler, Y., Ganzhorn, J. U., and Tomiuk, J. 2000. Genetic differentiation among natural populations of *Lepilemur ruficaudatus*. *International Journal of Primatology* 21: 853-864.

Baden, A. L., Mancini, A. N., Federman, S., Holmes, S. M., Johnson, S. E., Kamilar, J., Louis, E. E., and Bradley, B. J. 2019. Anthropogenic pressures drive

population genetic structuring across a Critically Endangered lemur species range. *Scientific Reports* 9: 16276.

Baden, A. L., Webster, T. H., and Bradley, B. J. 2020. Genetic relatedness cannot explain social preferences in black-and-white ruffed lemurs, *Varecia variegata*. *Animal Behaviour* 164: 73-82.

Bailey, C. A., McLain, A. T., Paquette, S. R., McGuire, S. M., Shore, G. D., et al. 2016. Evaluating the genetic diversity of three Endangered lemur species (Genus: *Propithecus*) from northern Madagascar. *Journal of Primatology* 5(1): 132.

Beaumont, M. A. 1999. Detecting population expansion and decline using microsatellites. *Genetics* 153: 2013-2029.

Beaumont, M. A. 2004. Recent developments in genetic data analysis: What can they tell us about human demographic history? *Heredity* 92: 365-379.

Beaumont, M. A. 2010. Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics* 41: 379-406.

Beaumont, M. A., and Nichols, R. A. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London, series B: Biological Sciences* 263: 1619-1626.

Beaumont, M. A., Zhang, W., and Balding, D. J. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162: 2025-2035.

Bertoncini, S., D'Ercole, J., Brisighelli, F., Ramanamanjato, J.-B., Capelli, C., Tofanelli, S., and Donati, G. 2017. Stuck in fragments: Population genetics of the Endangered collared brown lemur *Eulemur collaris* in the Malagasy littoral forest. *American Journal of Physical Anthropology* 163: 542-552.

Blair, C., Heckman, K. L., Russell, A. L., and Yoder, A. D. 2014. Multilocus coalescent analyses reveal the demographic history and speciation patterns of mouse lemur sister species. *BMC Evolutionary Biology* 14: 57.

Blanco, M. B., Elfawal, M. A., Durden, L. A., Beati, L., Xu, G., Godfrey, L. R., and Rich, S. M. 2013. Genetic diversity of ixodid ticks parasitizing eastern mouse and dwarf lemurs in Madagascar, with descriptions of the larva, nymph, and male of *Ixodes lemuris* (Acari: Ixodidae). *Journal of Parasitology* 99: 11-18.

Blanco, M. B., Greene, L. K., Rasambainarivo, F., Toomey, E., Williams, R. C., Andrianandrasana, L., Larsen, P. A., and Yoder, A. D. 2020. Next-generation technologies applied to age-old challenges in Madagascar. *Conservation Genetics* 21: 785-793.

Bolnick, D. I., and Otto, S. P. 2013. The magnitude of local adaptation under genotype-dependent dispersal. *Ecology and Evolution* 3: 4722-4735.

Bonadonna, G., Torti, V., Randrianarison, R. M., Martinet, N., Gamba, M., and Giacoma, C. 2014. Behavioral correlates of extra-pair copulation in *Indri indri*. *Primates* 55: 119-123.

Bonadonna, G., Torti, V., De Gregorio, C., Valente, D., Randrianarison, R. M., Pozzi, L., Gamba, M., and Giacoma, C. 2019. Evidence of genetic monogamy in the lemur Indri (*Indri indri*). *American Journal of Primatology* 81: e22993.

Bonner, T. I., Heinemann, R., and Todaro, G. J. 1980. Evolution of DNA sequences has been retarded in Malagasy primates. *Nature* 286: 420-423.

Bradburd, G. S., Ralph, P. L., and Coop, G. M. 2013. Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution* 67: 3258-3273.

Bradburd, G. S., Ralph, P. L., and Coop, G. M. 2016. A spatial framework for understanding population structure and admixture. *PLoS Genetics* 12: e1005703.

Brenneman, R. A., Johnson, S. E., Bailey, C. A., Ingraldi, C., Delmore, K. E., Wyman, T. M., Andriamaharoa, H. E., Ralainasolo, F. B., Ratsimbazafy, J. H., Louis, E.E. 2011. Population genetics and abundance of the endangered grey-headed lemur *Eulemur cinereiceps* in south-east Madagascar: Assessing risks for fragmented and continuous populations. *Oryx* 46: 298-307.

Brenneman, R. A., McLain, A. T., Taylor, J. M., Zaonarivelo, J. R., Lei, R., McGuire, S. M., Andriantompohavana, R., Rylands, A. B., and Louis, E. E. 2016. Genetic analysis of the Indri reveals no evidence of distinct subspecies. *International Journal of Primatology* 37: 460-477.

Brouat, C., Rahelinirina, S., Loiseau, A., Rahalison, L., Rajerison, M., Laffly, D., Handschumacher, P., and Duplantier, J.-M. 2013. Plague circulation and

population genetics of the reservoir *Rattus rattus*: The influence of topographic relief on the distribution of the disease within the Madagascan focus. *PLoS Neglected Tropical Diseases* 7: e2266.

Brouat, C., Tollenaere, C., Estoup, A., Loiseau, A., Sommer, S., et al. 2014. Invasion genetics of a human commensal rodent: The black rat *Rattus rattus* in Madagascar. *Molecular Ecology* 23: 4153-4167.

Cavalli-Sforza, L. L. 1966. Population structure and human evolution. *Proceedings of the Royal Society of London, series B, Biological Sciences* 164: 362-379.

Chan, L. M., Goodman, S. M., Nowak, M. D., Weisrock, D. W., and Yoder, A. D. 2011. Increased population sampling confirms low genetic divergence among *Pteropus* (Chiroptera: Pteropodidae) fruit bats of Madagascar and other western Indian Ocean islands. *PLoS Currents* 3: RRN1226.

Chesser, R. K. 1991. Gene diversity and female philopatry. *Genetics* 127: 437-447.

Chikhi, L., and Beaumont, M. A. 2005. Modelling human genetic history. In *The encyclopedia of genetics, genomics, proteomics & bioinformatics*, eds. M. J. Dunn, L. B. Jorde, P. F. R. Little, and S. Subramaniam. New York: John Wiley & Sons.

Chikhi, L., Sousa, V. C., Luisi, P., Goossens, B., and Beaumont, M. A. 2010. The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics* 186: 983-995.

Chikhi, L., Rodríguez, W., Grusea, S., Santos, P., Boitard, S., and Mazet, O. 2018. The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: Insights into demographic inference and model choice. *Heredity* 120: 13-24.

Comeron, J. M. 2017. Background selection as null hypothesis in population genomics: insights and challenges from *Drosophila* studies. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372: 20160471.

Cornelis, G., Vernochet, C., Malicorne, S., Souquere, S., Tzika, A. C., et al. 2014. Retroviral envelope syncytin capture in an ancestrally diverged mammalian clade for placentation in the primitive Afrotherian tenrecs. *Proceedings of the National Academy of Sciences of the USA* 111: E4332-E4341.

Cornuet, J. M., and Luikart, G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144: 2001-2014.

Cracraft, J. 1989. Speciation and its ontology: The empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In *Speciation and its consequences*, eds. D. Otte and J. A. Endler, pp. 27-59. Sunderland: Sinauer

Craul, M., Chikhi, L., Sousa, V., Olivieri, G. L., Rabesandratana, A., Zimmermann, E., and Radespiel, U. 2009. Influence of forest fragmentation on an endangered large-bodied lemur in northwestern Madagascar. *Biological Conservation* 142: 2862-2871.

Crovella, S., and Rumpler, Y. 1992. Confirmation of the specific status of *Hapalemur aureus* (Primates, Strepsirhini) by restriction genomic DNA banding patterns. *Human Evolution* 7: 63-67.

Crovella, S., Montagnon, D., Rakotosamimanana, B., and Rumpler, Y. 1994. Molecular biology and systematics of an extinct lemur: *Pachylemur insignis*. *Primates* 35: 519-522.

Currat, M., Ray, N., and Excoffier, L. 2004. Splatche: A program to simulate genetic diversity taking into account environmental heterogeneity. *Molecular Ecology Notes* 4: 139-142.

Currat, M., Arenas, M., Quilodràn, C. S., Excoffier, L., and Ray, N. 2019. SPLATCHE3: Simulation of serial genetic data under spatially explicit evolutionary scenarios including long-distance dispersal. *Bioinformatics* 35: 4480-4483.

Cushman, S. A. 2006. Effects of habitat loss and fragmentation on amphibians: A review and prospectus. *Biological Conservation* 128: 231-240.

Dammhahn, M., Kappeler, P. M. 2005. Social system of *Microcebus berthae*, the world's smallest primate. *International Journal of Primatology* 26: 407-435.

Davis, J. M., and Stamps, J. A. 2004. The effect of natal experience on habitat preferences. *Trends in Ecology & Evolution* 19: 411-416.

de Winter, I. I., Qurkhuli, T., de Groot, N., de Vos-Rouweler, A. J. M., van Hooft, P., Heitkönig, I. M. A., Prins, H. H. T., Bontrop, R. E., and Doxiadis, G. G. M.

2019. Determining Mhc-DRB profiles in wild populations of three congeneric true lemur species by noninvasive methods. *Immunogenetics* 71: 97-107.

Del Pero, M., Crovella, S., Cervella, P., Ardito, G., and Rumpler, Y. 1995. Phylogenetic relationships among Malagasy lemurs as revealed by mitochondrial DNA sequence analysis. *Primates* 36: 431-440.

Dene, H., Goodman, M., Prychodko, W., and Moore, G. W. 1976. Immunodiffusion systematics of the primates. III. The strepsirhini. *Folia Primatologica* 25: 35-61.

Dene, H., Goodman, M., and Prychodko, W. 1980. Immunodiffusion systematics of the Primates. IV: Lemuriformes. *Mammalia* 44: 211-224.

Di Fiore, A. 2003. Molecular genetic approaches to the study of primate behavior, social organization, and reproduction. *American Journal of Physical Anthropology*. 122: 62-99.

Di Fiore, A., and Valencia, L. M. 2014. The interplay of landscape features and social system on the genetic structure of a primate population: An agent-based simulation study using “tamarins.” *International Journal of Primatology* 35: 226-257.

Djelati, R., Brun, B., and Rumpler Y., 1997. Meiotic study of hybrids in the genus *Eulemur* and taxonomic considerations. *American Journal of Primatology* 42: 235-245.

Douady, C. J., Catzeflis, F., Kao, D. J., Springer, M. S., and Stanhope, M. J. 2002. Molecular evidence for the monophyly of Tenrecidae (Mammalia) and the timing of the colonization of Madagascar by Malagasy tenrecs. *Molecular Phylogenetics and Evolution* 22: 357-363.

Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* 22: 1185-1192.

Durbin, J., Funk, S. M., Hawkins, F., Hills, D. M., Jenkins, P. D., Moncrieff, C. B., and Ralainasolo, F. B. 2010. Investigations into the status of a new taxon of *Salanoia* (Mammalia: Carnivora: Eupleridae) from the marshes of Lac Alaotra, Madagascar. *Systematics and Biodiversity* 8: 341-355.

Dutrillaux, B. 1979. Chromosomal evolution in Primates: Tentative phylogeny from *Microcebus murinus* (Prosimian) to man. *Human Genetics* 48: 251-314.

Eisfeldt, J., Pettersson, M., Vezzi, F., Wincent, J., Käller, M., Gruselius, J., Nilsson, D., Syk Lundberg, E., Carvalho, C. M. B., and Lindstrand, A. 2019. Comprehensive structural variation genome map of individuals carrying complex chromosomal rearrangements. *PLoS Genetics* 15.

Everson, K. M., Soarimalala, V., Goodman, S. M., and Olson, L. E. 2016. Multiple loci and complete taxonomic sampling resolve the phylogeny and biogeographic history of tenrecs (Mammalia: Tenrecidae) and reveal higher speciation rates in Madagascar's humid forests. *Systematic Biology* 65: 890-909.

Everson, K. M., Hildebrandt, K. B. P., Goodman, S. M., and Olson, L. E. 2018. Caught in the act: Incipient speciation across a latitudinal gradient in a semifossorial mammal from Madagascar, the mole tenrec *Oryzorictes hova* (Tenrecidae). *Molecular Phylogenetics and Evolution* 126: 74-84.

Everson, K. M., Goodman, S. M., and Olson, L. E. 2020a. Speciation and gene flow in two sympatric small mammals from Madagascar, *Microgale fotsifotsy* and *M*. *soricoides* (Mammalia: Tenrecidae). *Molecular Ecology* 29: 1717-1729.

Everson, K. M., Jansa, S. A., Goodman, S. M., and Olson, L. E. 2020b. Montane regions shape patterns of diversification in small mammals and reptiles from Madagascar's moist evergreen forest. *Journal of Biogeography* 47: 2059-2072.

Ewens, W. J. 1972. The sampling theory of selectively neutral alleles. *Theoretical Population Biology* 3: 87-112.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., and Foll, M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genetics* 9: e1003905.

Faherty, S. L., Villanueva-Cañas, J. L., Klopfer, P. H., Albà, M. M., and Yoder, A. D. 2016. Gene expression profiling in the hibernating primate, *Cheirogaleus medius*. *Genome Biology and Evolution* 8: 2413-2426.

Faherty, S. L., Villanueva-Cañas, J. L., Blanco, M. B., Albà, M. M., and Yoder, A. D. 2018. Transcriptomics in the wild: Hibernation physiology in free-ranging dwarf lemurs. *Molecular Ecology* 27: 709-722.

Fausser, J. L., Rabarivola, C., Meier, B., Hahn, T., Rumpler, Y. 2000. Genetic comparison between different populations of Eulemur macaco flavifrons in northwest Madagascar using RAPD markers. *American Journal of Primatology* 51: 249-255.

Felsenstein, J. 1992. Estimating effective population size from samples of sequences: Inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genetics Research* 59: 139-147.

Fietz, J., Zischler, H., Schwiegk, C., Tomiuk, J., Dausmann, K. H., and Ganzhorn, J. U. 2000. High rates of extra-pair young in the pair-living fat-tailed dwarf lemur, *Cheirogaleus medius*. *Behavioral Ecology and Sociobiology* 49: 8-17.

Fredsted, T., Pertoldi, C., Olesen, J. M., Eberle, M., and Kappeler, P. M. 2004. Microgeographic heterogeneity in spatial distribution and mtDNA variability of gray mouse lemurs (*Microcebus murinus*, Primates: Cheirogaleidae). *Behavioral Ecology and Sociobiology* 56: 393-403.

Fredsted, T., Pertoldi, C., Schierup, M. H., and Kappeler, P. M. 2005. Microsatellite analyses reveal fine-scale genetic structure in grey mouse lemurs (*Microcebus murinus*). *Molecular Ecology* 14: 2363-2372.

Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915-925.

Gachot-Neveu, H., Petit, M., and Roeder, J. J. 1999. Paternity determination in two groups of *Eulemur fulvus mayottensis*: Implications for understanding mating strategies. *International Journal of Primatology* 20: 107-119.

Garza, J. C., and Williamson, E. G. 2001. Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* 10: 305-318.

Gilabert, A., Loiseau, A., Duplantier, J.-M., Rahelinirina, S., Rahalison, L., Chanteau, S., and Brouat, C. 2007. Genetic structure of black rat populations in a rural plague focus in Madagascar. *Canadian Journal of Zoology* 85: 965-972.

Glenn, T. C. 2011. Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11: 759-769.

Gligor, M., Ganzhorn, J. U., Rakotondravony, D., Ramilijaona, O. R., Razafimahatratra, E., Zischler, H., and Hapke, A. 2009. Hybridization between mouse lemurs in an ecological transition zone in southern Madagascar. *Molecular Ecology* 18: 520-533.

Goldstein, D. B., and Chikhi, L. 2002. Human migrations and population structure: What we know and why it matters. *Annual Review of Genomics and Human Genetics* 3: 129-152.

Gomard, Y., Dietrich, M., Wieseke, N., Ramasindrazana, B., Lagadec, E., Goodman, S. M., Dellagi, K., and Tortosa, P. 2016. Malagasy bats shelter a considerable genetic diversity of pathogenic *Leptospira* suggesting notable host-specificity patterns. *FEMS Microbiology Ecology* 92.

Goodman, M., Baileym, W. J., Hayasaka, K., Stanhope, M. J., Slightom, J., and Czelusniak, J. 1994. Molecular evidence on primate phylogeny from DNA sequences. *American Journal of Physical Anthropology* 94: 3-24.

Goodman, S. M., and Helgen, K. M. 2010. Species limits and distribution of the Malagasy carnivoran genus *Eupleres* (Family Eupleridae). *Mammalia* 74: 177-185.

Goodman, S. M., Chan, L. M., Nowak, M. D., and Yoder, A. D. 2010. Phylogeny and biogeography of western Indian Ocean *Rousettus* (Chiroptera: Pteropodidae). *Journal of Mammalogy* 91: 593-606.

Goossens, B., Chikhi, L., Utami, S., de Ruiter, J., and Bruford, M. 2000. A multi-samples, multi-extracts approach for microsatellite analysis of faecal samples in an arboreal ape. *Conservation Genetics* 1: 157-162.

Goossens, B., Chikhi, L., Ancrenaz, M., Lackman-Ancrenaz, I., Andau, P., Bruford, M. W. 2006. Genetic signature of anthropogenic population collapse in orangutans. *PLoS Biology* 4: e25.

Goossens, B., Anthony, N., Jeffery, K., Johnson-Bawe, M., and Bruford, M. W. 2011. Collection, storage and analysis of non-invasive genetic material in primate biology. In Field and laboratory methods in primatology: A practical guide,

eds. J. M. Setchell and D. J. Curtis, pp. 371-386. Cambridge, UK: Cambridge University Press.

Greene, L. K., Clayton, J. B., Rothman, R. S., Semel, B. P., Semel, M. A., Gillespie, T. R., Wright, P. C., and Drea, C. M. 2019. Local habitat, not phylogenetic relatedness, predicts gut microbiota better within folivorous than frugivorous lemur lineages. *Biology Letters* 15: 20190028.

Greenwood, P. J. 1980. Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour* 28: 1140-1162.

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., and Siepel, A. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics* 43: 1031-1034.

Guevara, E. E., Frankel, D. C., Ranaivonasy, J., Richard, A. F., Ratsirarson, J., Lawler, R. R., and Bradley, B. J. 2018. A simple, economical protocol for DNA extraction and amplification where there is no lab. *Conservation Genetics Resources* 10: 119-125.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., and Bustamante, C. D. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics* 5: e1000695.

Hafen, T., Neveu, H., Rumpler, Y., Wilden, I., and Zimmermann, E. 1998. Acoustically dimorphic advertisement calls separate morphologically and genetically homogenous populations of the grey mouse lemur (*Microcebus murinus*). *Folia Primatologica* 69(Suppl 1): 342-356.

Han, G.-Z., and Worobey, M. 2012. An endogenous foamy virus in the Aye-Aye (*Daubentonia madagascariensis*). *Journal of Virology* 86: 7696-7698.

Hapke, A., Gligor, M., Rakotondranary, S. J., Rosenkranz, D., and Zupke, O. 2011. Hybridization of mouse lemurs: different patterns under different ecological conditions. *BMC Evolutionary Biology* 11: 297.

Hardy, O. J., and Vekemans, X. 2002. Spagedi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618-620.

Hasegawa, M., Kishino, H., Hayasaka, K., and Horai, S. 1990. Mitochondrial DNA evolution in primates: Transition rate has been extremely low in the lemur. *Journal of Molecular Evolution* 31: 113-121.

Hawkins, M. T. R., Culligan, R. R., Frasier, C. L., Dikow, R. B., Hagenson, R., Lei, R., and Louis, E. E. 2018. Genome sequence and population declines in the critically endangered greater bamboo lemur (*Prolemur simus*) and implications for conservation. *BMC Genomics* 19: 445.

Heled, J., and Drummond, A. J. 2008. Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology* 8: 289.

Heller, R., Chikhi, L., and Siegismund, H. R. 2013. The confounding effect of population structure on Bayesian Skyline Plot inferences of demographic history. *PLoS ONE* 8: e62992.

Hendry, A. P. 2004. Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. *Evolutionary Ecology Research* 6: 1219-1236.

Herbots, H. M. J. D. 1994. Stochastic models in population genetics: Genealogy and genetic differentiation in structured populations. Ph.D. thesis, University of London.

Hingston, M., Goodman, S. M., Ganzhorn, J. U., and Sommer, S. 2005. Reconstruction of the colonization of southern Madagascar by introduced *Rattus rattus*. *Journal of Biogeography* 32: 1549-1559.

Hohenbrink, P., Mundy, N. I., and Radespiel, U. 2017. Population genetics of mouse lemur vomeronasal receptors: Current versus past selection and demographic inference. *BMC Evolutionary Biology* 17: 28.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., and Cresko, W. A. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* 6(2): e1000862.

Hostetler, J. A., Kneip, E., Van Vuren, D. H., and Oli, M. K. 2012. Stochastic population dynamics of a montane ground-dwelling squirrel. *PLoS One* 7: e34379.

Hotaling, S., Foley, M. E., Lawrence, N. M., Bocanegra, J., Blanco, M. B., Rasoloarison, R., Kappeler, P. M., Barrett, M. A., Yoder, A. D., and

Weisrock, D. W. 2016. Species discovery and validation in a cryptic radiation of endangered primates: Coalescent-based species delimitation in Madagascar's mouse lemurs. *Molecular Ecology* 25: 2029-2045.

Hua, P. Y., Chen, J. P., Sun, M., Liang, B., Zhang, S. Y., and Wu, D. H. 2006. Characterization of microsatellite loci in fulvous fruit bat *Rousettus leschenaulti*. *Molecular Ecology Notes* 6: 939-941.

Hubby, J. L., and Lewontin, R. C. 1966. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* 54: 577-594.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. In *Oxford surveys in evolutionary biology*, eds. D. Futuyma and J. Antonovics, pp. 1-44. Oxford: Oxford University Press

Hudson, R. R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18: 337-338

Hunnicutt, K. E., Tiley, G. P., Williams, R. C., Larsen, P. A., Blanco, M. B., et al. 2020. Comparative genomic analysis of the pheromone receptor class 1 family (V1R) reveals extreme complexity in mouse lemurs (Genus, *Microcebus*) and a chromosomal hotspot across mammals. *Genome Biology and Evolution* 12: 3562-3579.

Jacobs, R. L., Frankel, D. C., Rice, R. J., Kiefer, V. J., and Bradley, B. J. 2018. Parentage complexity in socially monogamous lemurs (*Eulemur rubriventer*): Integrating genetic and observational data. *American Journal of Primatology* 80: e22738.

Jansa, S. A., Goodman, S. M., and Tucker, P. K. 1999. Molecular phylogeny and biogeography of the native rodents of Madagascar (Muridae: Nesomyinae): A test of the single-origin hypothesis. *Cladistics* 15: 253-270.

Jansa, S. A., Carleton, M. D., Soarimalala, V., Rakotomalala, Z., and Goodman, S. M. 2019. A review of the *Eliurus tanala* complex (Rodentia, Muroidea, Nesomyidae), with description of a new species from dry forests of western Madagascar. *Bulletin of the American Museum of Natural History* 430: 1-67.

Jung, K. Y., Crovella, S., and Rumpler, Y. 1992. Phylogenetic relationships among lemuriform species determined from restriction genomic DNA banding patterns. *Folia Primatologica* 58: 224-229.

Kappeler, P. M., and Schäffler, L. 2008. The lemur syndrome unresolved: Extreme male reproductive skew in sifakas (*Propithecus verreauxi*), a sexually monomorphic primate with female dominance. *Behavioral Ecology and Sociobiology* 62: 1007-1015.

Kappeler, P. M., Wimmer, B., Zinner, D., and Tautz, D. 2002. The hidden matrilineal structure of a solitary lemur: Implications for primate social evolution. *Proceedings of the Royal Society of London, series B: Biological Sciences* 269: 1755-1763.

Karanth, K. P., Delefosse, T., Rakotosamimanana, B., Parsons, T. J., and Yoder, A. D. 2005. Ancient DNA from giant extinct lemurs confirms single origin of Malagasy primates. *Proceedings of the National Academy of Sciences of the USA* 102: 5090-5095.

Katzourakis, A., Aiewsakun, P., Jia, H., Wolfe, N. D., LeBreton, M., Yoder, A. D., and Switzer, W. M. 2014. Discovery of prosimian and afrotherian foamy viruses and potential cross species transmissions amidst stable and ancient mammalian co-evolution. *Retrovirology* 11: 61.

Kemp, J., López-Baucells, A., Rocha, R., Wangensteen, O. S., Andriatafika, Z., Nair, A., and Cabeza, M. 2019. Bats as potential suppressors of multiple agricultural pests: A case study from Madagascar. *Agriculture, Ecosystems & Environment* 269: 88-96.

Kiene, F., Andriatsitohaina, B., Ramsay, M. S., Rakotondramanana, H., Rakotondravony, R., Radespiel, U., and Strube, C. 2020. Forest edges affect ectoparasite infestation patterns of small mammalian hosts in fragmented forests in Madagascar. *International Journal for Parasitology* 50: 299-313.

Kingman, J. F. C. 1982. The coalescent. *Stochastic Processes and their Applications* 13: 235-248.

Kistler, L., Ratan, A., Godfrey, L. R., Crowley, B. E., Hughes, C. E., et al. 2015. Comparative and population mitogenomic analyses of Madagascar's extinct,

giant ‘subfossil’ lemurs. *Journal of Human Evolution* 79(Special Issue): 45-54.

Koop, B. F., Tagle, D. A., Goodman, M., and Slightom, J. L. 1989. A molecular view of primate phylogeny and important systematic and evolutionary questions. *Molecular Biology and Evolution* 6: 580-612.

Lamb, J. M., Naidoo, T., Taylor, P. J., Napier, M., Ratrimomanarivo, F., and Goodman, S. M. 2012. Genetically and geographically isolated lineages of a tropical bat (Chiroptera: Molossidae) show demographic stability over the late Pleistocene. *Biological Journal of the Linnean Society* 106: 18-40.

Landguth, E. L., Cushman, S. A., Schwartz, M. K., McKelvey, K. S., Murphy, M., and Luikart, G. 2010. Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology* 19: 4179-4191.

Larsen, P. A., Campbell, C. R., and Yoder, A. D. 2014. Next-generation approaches to advancing eco-immunogenomic research in critically endangered primates. *Molecular Ecology Resources* 14: 1198-1209.

Lawler, R. R. 2007. Fitness and extra-group reproduction in male Verreaux’s sifaka: An analysis of reproductive success from 1989-1999. *American Journal of Physical Anthropology* 132: 267-277.

Lawler, R. R. 2018. Emerging and enduring issues in primate conservation genetics. *Annual Review of Anthropology* 47: 395-415.

Lawler, R. R., Richard, A. F., and Riley, M. A. 2003. Genetic population structure of the white sifaka (*Propithecus verreauxi verreauxi*) at Beza Mahafaly Special Reserve, southwest Madagascar (1992-2001). *Molecular Ecology* 12: 2307-2317.

Lawson, D. J., Hellenthal, G., Myers, S., and Falush, D. 2012. Inference of population structure using dense haplotype data. *PLoS Genetics* 8: e1002453.

Lehman, S. M., and Wright, P. 2000. Preliminary study of the conservation status of lemur communities in the Betsakafandrika region of eastern Madagascar. *Lemur News* 5: 23-25.

Lei, R., Frasier, C. L., Hawkins, M. T. R., Engberg, S. E., Bailey, C. A., et al. 2017. Phylogenomic reconstruction of sportive lemurs (genus *Lepilemur*)

recovered from mitogenomes with inferences for Madagascar biogeography. *Journal of Heredity* 108, 107–119.

Leipoldt, M., Schlitz, A., Ganzhorn, J. U., and Tomiuk, J. 1996. Genetic analysis of *Lepilemur mustelinus ruficaudatus* populations using RAPD markers. In Biogéographie de Madagascar - Biogeography of Madagascar, ed. W. R. Lourenço, pp. 267-274. Paris: Editions ORSTOM.

Leitwein, M., Duranton, M., Rougemont, Q., Gagnaire, P.-A., and Bernatchez, L. 2020. Using haplotype information for conservation genomics. *Trends in Ecology & Evolution* 35: 245-258.

Li, H., and Durbin, R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475: 493-496.

Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478: 476-482.

Liu, X., and Fu, Y.-X. 2015. Exploring population size changes using SNP frequency spectra. *Nature Genetics* 47: 555-559.

Loiseau, A., Rahelinirina, S., Rahalison, L., Konečný, A., Duplantier, J.-M., and Brouat, C. 2008. Isolation and characterization of microsatellites in *Rattus rattus*. *Molecular Ecology Resources* 8: 916-918.

Louis, E. E., and Lei, R. 2016. Mitogenomics of the family Cheirogaleidae and relationships to taxonomy and biogeography in Madagascar. In *The dwarf and mouse lemurs of Madagascar: Biology, behavior and conservation biogeography of the Cheirogaleidae*, eds. S. M. Lehman, U. Radespiel, and E. Zimmermann, pp. 54-93. Cambridge, UK: Cambridge University Press.

Louis, E. E., Ratsimbazafy, J. H., Razakamaharauo, V. R., Pierson, D. J., Barber, R. C., and Brenneman, R. A. 2005. Conservation genetics of black and white ruffed lemurs, *Varecia variegata*, from southeastern Madagascar. *Animal Conservation* 8(1): 105-111.

Louis, E. E., Engberg, S. E., Lei, R., Geng, H., Sommer, J. A., et al. 2006. Molecular and morphological analyses of the sportive lemurs (Family Megaladapidae: Genus *Lepilemur*) reveals 11 previously unrecognized species. *Special Publications, Museum of Texas Tech University* 49: 1-47.

Louis, E. E., Engberg, S. E., McGuire, S. M., McCormick, M. J., Randriamampionona, R., Ranaivoarisoa, J. F., Bailey, C. A., Mittermeier, R. A., and Lei, R. 2008. Revision of the mouse lemurs, *Microcebus* (Primates, Lemuriformes), of northern and northwestern Madagascar with descriptions of two new species at Montagne d'Ambre National Park and Antafondro Classified Forest. *Primate Conservation* 23: 19-38.

Lührs, M.-L., Dammhahn, M., and Kappeler, P. 2013. Strength in numbers: Males in a carnivore grow bigger when they associate and hunt cooperatively. *Behavioral Ecology* 24: 21-28.

Lynch, M., and Gabriel, W. 1990. Mutation load and the survival of small populations. *Evolution* 44: 1725-1737.

Lynch, M., Bürger, R., Butcher, D., and Gabriel, W. 1990. The mutational meltdown in asexual populations. *Journal of Heredity* 84: 339-344.

Madsen, T., Shine, R., Olsson, M., and Wittzell, H. 1999. Restoration of an inbred adder population. *Nature* 402: 34-35.

Markolf, M., and Kappeler, P. M. 2013. Phylogeographic analysis of the true lemurs (genus *Eulemur*) underlines the role of river catchments for the evolution of micro-endemism in Madagascar. *Frontiers in Zoology* 10: 70.

Markolf, M., Brameier, M., and Kappeler, P. M. 2011. On species delimitation: Yet another lemur species or just genetic variation? *BMC Evolutionary Biology* 11: 1-7.

Markolf, M., Rakotonirina, H., Fichtel, C., von Grumbkow, P., Brameier, M., and Kappeler, P. M. 2013. True lemurs…true species - species delimitation using multiple data sources in the brown lemur complex. *BMC Evolutionary Biology* 13: 233.

Mayor, M. I., Sommer, J. A., Huebinger, R. M., Barber, R. C., and Louis, E. E. 2002. Characterization of seven microsatellite marker loci in a genus of Malagasy lemurs (*Propithecus*). *Molecular Ecology Notes* 2: 385-388.

Mazet, O., Rodríguez, W., Grusea, S., Boitard, S., and Chikhi, L. 2016. On the importance of being structured: Instantaneous coalescence rates and human evolution--lessons for ancestral population size inference? *Heredity* 116: 362-371.

McRae, B. H. 2006. Isolation by resistance. *Evolution* 60: 1551-1561.

Meredith, R. W., Janečka, J. E., Gatesy, J., Ryder, O. A., Fisher, C. A., et al. 2011. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* 334: 521-524.

Merenlender, A. 1993. The effects of sociality on the demography and genetic structure of the *Lemur fulvus rufus* (polygamous) and *Lemur rubriventer* (monogamous) and the conservation implications. Ph.D. thesis, University of Rochester, New York.

Meyer, W. K., Venkat, A., Kermany, A. R., van de Geijn, B., Zhang, S., and Przeworski, M. 2015. Evolutionary history inferred from the de novo assembly of a nonmodel organism, the blue-eyed black lemur. *Molecular Ecology* 24: 4392-4405.

Meyers, D. M. 1993. The effects of resource seasonality on behavior and reproduction in the golden-crowned sifaka (*Propithecus tattersalli*, Simons, 1988) in three Malagasy forests. Ph.D. thesis, Duke University.

Mittermeier, R. A., Konstant, W., Hawkins, F., Louis, E. E., Langrand, O., et al. 2006. *Lemurs of Madagascar*, 2nd ed. Washington, D.C.: Conservation International.

Montagnon, D., Crovella, S., and Rumpler, Y. 1993. Comparison of highly repeated DNA sequences in some Lemuridae and taxonomic implications. *Cytogenetic and Genomic Research* 63: 131-134.

Mouchaty, S. K., Gullberg, A., Janke, A., and Arnason, U. 2000. Phylogenetic position of the tenrecs (Mammalia: Tenrecidae) of Madagascar based on analysis of the complete mitochondrial genome sequence of *Echinops telfairi*. *Zoologica Scripta* 29: 307-317.

Munson, L., Brown, J. L., Bush, M., Packer, C., Janssen, D., Reiziss, S. M., and Wildt, D. E. 1996. Genetic diversity affects testicular morphology in free-ranging lions (*Panthera leo*) of the Serengeti Plains and Ngorongoro Crater. *Journal of Reproduction and Fertility* 108: 11-15.

Nei, M., Maruyama, T., and Chakraborty, R. 1975. The bottleneck effect and genetic variability in populations. *Evolution* 29: 1-10.

Neveu, H., Hafen, T., Zimmermann, E., and Rumpler, Y. 1998. Comparison of the genetic diversity of wild and captive groups of *Microcebus murinus* using the random amplified polymorphic DNA method. *Folia Primatologica* 69(Suppl 1): 127-135.

Nievergelt, C. M., Mutschler, T., Feistner, A. T. C., and Woodruff, D. S. 2002. Social system of the Alaotran gentle lemur (*Hapalemur griseus alaotrensis*): Genetic characterization of group composition and mating system. *American Journal of Primatology* 57: 157-176.

Nosil, P., Egan, S. P., and Funk, D. J. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: “Isolation by adaptation” and multiple roles for divergent selection. *Evolution* 62: 316-336.

Nunziata, S. O., Wallenhorst, P., Barrett, M. A., Junge, R. E., Yoder, A. D., and Weisrock, D. W. 2016. Population and conservation genetics in an endangered lemur, *Indri indri*, across three forest reserves in Madagascar. *International Journal of Primatology* 37: 688-702.

Olivieri, G. L., Sousa, V., Chikhi, L., and Radespiel, U. 2008. From genetic diversity and structure to conservation: Genetic signature of recent population declines in three mouse lemur species (*Microcebus* spp.). *Biological Conservation* 141: 1257-1271.

Olson, L. E., Rakotomalala, Z., Hildebrandt, K. B. P., Lanier, H. C., Raxworthy, C. J., and Goodman, S. M. 2009. Phylogeography of *Microgale brevicaudata* (Tenrecidae) and description of a new species from western Madagascar. *Journal of Mammalogy* 90: 1095-1110.

Orlando, L., Calvignac, S., Schnebelen, C., Douady, C. J., Godfrey, L. R., and Hänni, C. 2008. DNA from extinct giant lemurs links archaeolemurids to extant indriids. *BMC Evolutionary Biology* 8: 121.

Parga, J. A., Sauther, M. L., Cuozzo, F. P., Jacky, I. A. Y., and Lawler, R. R. 2012. Evaluating ring-tailed lemurs (*Lemur catta*) from southwestern Madagascar for a genetic population bottleneck. *American Journal of Physical Anthropology* 147: 21-29.

Parga, J. A., Sauther, M. L., Cuozzo, F. P., Jacky, I. A. Y., Lawler, R. R., Sussman, R. W., Gould, L., Pastorini, J. 2016. Paternity in wild ring-tailed lemurs

(*Lemur catta*): Implications for male mating strategies. *American Journal of Primatology* 78: 1316-1325.

Parreira, B. R., and Chikhi, L. 2015. On some genetic consequences of social structure, mating systems, dispersal, and sampling. *Proceedings of the National Academy of Sciences of the USA* 112: E3318-E3326.

Parreira, B., Quéméré, E., Vanpé, C., Carvalho, I., and Chikhi, L. 2020. Genetic consequences of social structure in the golden-crowned sifaka. *Heredity* 125: 328-339.

Pastorini, J., Zaramody, A., Curtis, D. J., Nievergelt, C. M., and Mundy, N. I. 2009. Genetic analysis of hybridization and introgression between wild mongoose and brown lemurs. *BMC Evolutionary Biology* 9: 32.

Pereira, M. E., and Weiss, M. L. 1991. Female mate choice, male migration, and the threat of infanticide in ringtailed lemurs. *Behavioral Ecology and Sociobiology* 28: 141-152.

Perofsky, A. C., Lewis, R. J., Abondano, L. A., Di Fiore, A., and Meyers, L. A. 2017. Hierarchical social networks shape gut microbial composition in wild Verreaux's sifaka. *Proceedings of the Royal Society, series B: Biological Sciences* 284: 20172274.

Perofsky, A. C., Lewis, R. J., and Meyers, L. A. 2019. Terrestriality and bacterial transfer: A comparative study of gut microbiomes in sympatric Malagasy mammals. *The ISME Journal* 13: 50-63.

Perry, G. H., Reeves, D., Melsted, P., Ratan, A., Miller, W., Michelini, K., Louis, E. E., Pritchard, J. K., Mason, C. E., and Gilad, Y. 2012. A genome sequence resource for the aye-aye (*Daubentonia madagascariensis*), a nocturnal lemur from Madagascar. *Genome Biology and Evolution* 4: 126-135.

Petter, J.-J. 1962. Recherches sur l'écologie et l'éthologie des lémuriens malgaches. *Mémoires du Muséum national d'Histoire naturelle*, nouvelle série, série A, Zoologie, 27: 1-146.

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J, Manuel, M., Wörheide, G., and Baurain, D. 2011. Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biology* 9(3): e1000602.

Piry, S., Luikart, G., and Cornuet, J.-M. 1999. Computer note. BOTTLENECK: A computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity* 90, 502–503.

Poelstra, J., Salmona, J., Tiley, G. P., Schüßler, D., Blanco, M. B., et al. 2020. Cryptic patterns of speciation in cryptic primates: Microendemic mouse lemurs and the multispecies coalescent. *Systematic Biology* syaa053.

Poux, C., Madsen, O., Marquard, E., Vieites, D. R., de Jong, W. W., and Vences, M. 2005. Asynchronous colonization of Madagascar by the four endemic clades of primates, tenrecs, carnivores, and rodents as inferred from nuclear genes. *Systematic Biology* 54: 719-730.

Pouyet, F., Aeschbacher, S., Thiéry, A., and Excoffier, L. 2018. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *eLife* 7: e36317.

Pritchard, J. K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Prout, T. 1981. A note on the island model with sex dependent migration. *Theoretical and Applied Genetics* 59: 327-332.

Quéméré, E., Crouau-Roy, B., Rabarivola, C., Louis Jr, E. E., and Chikhi, L. 2010a. Landscape genetics of an endangered lemur (*Propithecus tattersalli*) within its entire fragmented range. *Molecular Ecology* 19: 1606-1621.

Quéméré, E., Louis, E. E., Ribéron, A., Chikhi, L., and Crouau-Roy, B. 2010b. Non-invasive conservation genetics of the critically endangered golden-crowned sifaka (*Propithecus tattersalli*): High diversity and significant genetic differentiation over a small range. *Conservation Genetics* 11: 675-687.

Quéméré, E., Amelot, X., Pierson, J., Crouau-Roy, B., and Chikhi, L. 2012. Genetic data suggest a natural prehuman origin of open habitats in northern Madagascar and question the deforestation narrative in this region. *Proceedings of the National Academy of Sciences of the USA* 109: 13028-13033.

Quéméré, E., Hibert, F., Miquel, C., Lhuillier, E., Rasolondraibe, E., et al. 2013. A DNA metabarcoding study of a primate dietary diversity and plasticity across its entire fragmented range. *PLoS One* 8: e58971.

Rabarivola, C., Meyers, D., Rumpler, Y. 1991. Distribution and morphological characters of intermediate forms between the black lemur (*Eulemur macaco macaco*) and the Sclater's lemur (*E. m. flavifrons*). *Primates* 32: 269-273.

Rabarivola, C., Meier, B., Langer, C., Scheffrahn, W., and Rumpler, Y. 1996. Population genetics of *Eulemur macaco macaco* (Primates: Lemuridae) on the islands of Nosy-Be and Nosy-Komba and the Peninsula of Ambato (Madagascar). *Primates* 37: 215-225.

Radespiel, U., Sarikaya, Z., Zimmermann, E., and Bruford, M. W. 2001. Sociogenetic structure in a free-living nocturnal primate population: Sex-specific differences in the grey mouse lemur (*Microcebus murinus*). *Behavioral Ecology and Sociobiology* 50: 493-502.

Radespiel, U., Lutermann, H., Schmelting, B., Bruford, M. W., and Zimmermann, E. 2003. Patterns and dynamics of sex-biased dispersal in a nocturnal primate, the grey mouse lemur, *Microcebus murinus*. *Animal Behaviour* 65: 709-719.

Radespiel, U., Lutermann, H., Schmelting, B., and Zimmermann, E. 2019. An empirical estimate of the generation time of mouse lemurs. *American Journal of Primatology* 81: e23062.

Raharinosy, V., Olive, M.-M., Andriamiarimanana, F. M., Andriamandimby, S. F., Ravalohery, J.-P., Andriamamonjy, S., Filippone, C., Rakoto, D. A. D., Telfer, S., and Heraud, J.-M. 2018. Geographical distribution and relative risk of Anjozorobe virus (Thailand orthohantavirus) infection in black rats (*Rattus rattus*) in Madagascar. *Virology Journal* 15: 83.

Rakotoarisoa, G., Shore, G., Mcguire, S., Engberg, S., Louis, E., and Brenneman, R. 2006. Characterization of 20 microsatellite marker loci in Coquerel's sifaka (*Propithecus coquereli*). *Molecular Ecology Notes* 6: 1119-1121.

Rakotoarisoa, J.-E., Raheriarisena, M., and Goodman, S. M. 2013a. Late Quaternary climatic vegetational shifts in an ecological transition zone of northern Madagascar: Insights from genetic analyses of two endemic rodent species. *Journal of Evolutionary Biology* 26: 1019-1034.

Rakotoarisoa, J.-E., Raheriarisena, M., and Goodman, S.M. 2013b. A phylogeographic study of the endemic rodent *Eliurus carletoni* (Rodentia:

Nesomyinae) in an ecological transition zone of northern Madagascar. *Journal of Heredity* 104: 23-35.

Rakotoarivelo, A. R., Goodman, S. M., Schoeman, M. C., and Willows-Munro, S. 2019. Phylogeography and population genetics of the endemic Malagasy bat, *Macronycteris commersoni* s.s. (Chiroptera: Hipposideridae). *PeerJ* 7: e5866.

Ramanamahefa, R., McGuire, S. M., Louis, E. E., and Brenneman, R. A. 2009a. Population genetic parameter estimates for six populations of crowned lemurs, *Eulemur coronatus* (Gray, 1842), from northern Madagascar. *Lemur News* 14: 21-26.

Ramanamahefa, R., McGuire, S. M., Louis, E. E., and Brenneman, R. A. 2009b. Population genetic parameter estimates for five populations of Sanford's lemur, *Eulemur sanfordi* (Archbold, 1932), from northern Madagascar. *Lemur News* 14: 26-31.

Rakotonirina, L. H. F., Randriantsara, F., Rakotoarisoa, A. H., Rakotondrabe, R., Razafindramanana, J., Ratsimbazafy, J., and King, T. 2014. A preliminary assessment of sifaka (*Propithecus*) distribution, chromatic variation and conservation in western central Madagascar. *Primate Conservation* 28:43-53.

Ranaivoarisoa, J. F., Brenneman, L., McGuire, S. M., Lei, R., Ravelonjanahary, S. S., et al. 2010. Population genetic study of the Red-collared Brown Lemur (*Eulemur collaris* É. Geoffroy) in southeastern Madagascar. *The Open Conservation Biology Journal* 4: 1-8.

Ratrimomanarivo, F. H., Goodman, S. M., Taylor, P. J., Melson, B., and Lamb, J. 2009. Morphological and genetic variation in *Mormopterus jugularis* (Chiroptera: Molossidae) in different bioclimatic regions of Madagascar with natural history notes. *Mammalia* 73: 110-129.

Razakamaharavo, V. R., McGuire, S. M., Vasey, N., Louis, E. E., and Brenneman, R. A. 2010. Genetic architecture of two red ruffed lemur (*Varecia rubra*) populations of Masoala National Park. *Primates* 51: 53-61.

Reynes, J.-M., Razafindralambo, N. K., Lacoste, V., Olive, M.-M., Barivelo, T. A., Soarimalala, V., Heraud, J.-M., and Lavergne, A. 2014. Anjozorobe

hantavirus, a new genetic variant of Thailand virus detected in rodents from Madagascar. *Vector-Borne and Zoonotic Diseases* 14: 212-219.

Ringbauer, H., Coop, G., and Barton, N. H. 2017. Inferring recent demography from isolation by distance of longs shared sequence blocks. *Genetics* 205: 1335-1351.

Roelke, M. E., Martenson, J. S., and O'Brien, S. J. 1993. The consequences of demographic reduction and genetic depletion in the endangered Florida panther. *Current Biology* 3: 340-350.

Rogers, A. R., and Harpending, H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9: 552-569.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145: 1219-1228.

Rumpler, Y., and Albignac, R. 1978. Chromosome studies of the lepilemur, an endemic Malagasy genus of lemurs: Contribution of the cytogenetics to their taxonomy. *Journal of Human Evolution* 7: 191-196.

Rumpler, Y., and Dutrillaux, B. 1976. Chromosomal evolution in Malagasy lemurs. *Cytogenetic and Genome Research* 17: 268-281.

Rumpler, Y., Ishak, B., Warter, S., and Dutrillaux, B. 1985. Chromosomal evolution in Malagasy lemurs. VIII. Chromosome banding studies of *Lepilemur ruficaudatus*, *L leucopus*, and *L septentrionalis*. *Cytogenetic and Genomic Research* 39: 194-199.

Russell, A. L., Ranivo, J., Palkovacs, E. P., Goodman, S. M., and Yoder, A. D. 2007. Working at the interface of phylogenetics and population genetics: A biogeographical analysis of *Triaenops* spp. (Chiroptera: Hipposideridae). *Molecular Ecology* 16: 839-851.

Salmona, J., Teixeira, H., Rasolondraibe, E., Aleixo-Pais, I., Kun-Rodrigues, C., et al. 2015. Genetic diversity, population size, and conservation of the critically Endangered Perrier's sifaka (*Propithecus perrieri*). *International Journal of Primatology* 36: 1132-1153.

Salmona, J., Heller, R., Quéméré, E., and Chikhi, L. 2017. Climate change and human colonization triggered habitat loss and fragmentation in Madagascar. *Molecular Ecology* 26: 5203-5222.

Schad, J., Ganzhorn, J. U., and Sommer, S. 2005. Parasite burden and constitution of major histocompatibility complex in the Malagasy mouse lemur, *Microcebus murinus*. *Evolution* 59: 439-450.

Scheel, B. M., Henke-von der Malsburg, J., Giertz, P., Rakotondranary, S. J., Hausdorf, B., and Ganzhorn, J. U. 2015. Testing the influence of habitat structure and geographic distance on the genetic differentiation of mouse lemurs (*Microcebus*) in Madagascar. *International Journal of Primatology* 36: 823-838.

Schneider, N., Chikhi, L., Currat, M., and Radespiel, U. 2010. Signals of recent spatial expansions in the grey mouse lemur (*Microcebus murinus*). *BMC Evolutionary Biology* 10: 105.

Schneider, T. C., and Kappeler, P. M. 2014. Social systems and life-history characteristics of mongooses. *Biological Reviews* 89: 173-198.

Schneider, T. C., and Kappeler, P. M. 2016. Gregarious sexual segregation: The unusual social organization of the Malagasy narrow-striped mongoose (*Mungotictis decemlineata*). *Behavioral Ecology and Sociobiology* 70: 913-926.

Schneider, T. C., Kappeler, P. M., and Pozzi, L. 2016. Genetic population structure and relatedness in the narrow-striped mongoose (*Mungotictis decemlineata*), a social Malagasy carnivore with sexual segregation. *Ecology and Evolution* 6: 3734-3749.

Schülke, O., Kappeler, P. M., Zischler, H. 2004. Small testes size despite high extra-pair paternity in the pair-living nocturnal primate *Phaner furcifer*. *Behavioral Ecology and Sociobiology* 55: 293-301.

Schüßler, D., Blanco, M. B., Salmona, J., Poelstra, J., Andriambeloson, J. B., et al. 2020. Ecology and morphology of mouse lemurs (*Microcebus* spp.) in a hotspot of microendemism in northeastern Madagascar, with the description of a new species. *American Journal of Primatology* 32: e23180.

Schwensow, N., Fietz, J., Dausmann, K. H., and Sommer, S. 2007. Neutral versus adaptive genetic variation in parasite resistance: Importance of major histocompatibility complex supertypes in a free-ranging primate. *Heredity* 99: 265-277.

Schwensow, N., Dausmann, K., Eberle, M., Fietz, J., and Sommer, S. 2010a. Functional associations of similar MHC alleles and shared parasite species in two sympatric lemurs. *Infection, Genetics and Evolution* 10: 662-668.

Schwensow, N., Eberle, M., and Sommer, S. 2010b. Are there ubiquitous parasite-driven major histocompatibility complex selection mechanisms in gray mouse lemurs? *International Journal of Primatology* 31: 519-537.

Sexton, J. P., Hangartner, S. B., and Hoffmann, A. A. 2014. Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution* 68: 1-15.

Sgarlata, G. M., Salmona, J., Razanaparany, T. P., Rabarivola, C., Jan, F., et al. 2016. Mitochondrial genetic diversity in the Crowned Sifaka (*Propithecus coronatus*) in a fragmented landscape. *Primate Conservation* 30: 39-57.

Sgarlata, G. M., Salmona, J., Aleixo-Pais, I., Rakotonanahary, A., Sousa, A. P., et al. 2018. Genetic differentiation and demographic history of the Northern Rufous Mouse Lemur (*Microcebus tavaratra*) across a fragmented landscape in northern Madagascar. *International Journal of Primatology* 39: 65-89.

Sgarlata, G. M., Salmona, J., Pors, B. L., Rasolondraibe, E., Jan, F., et al. 2019. Genetic and morphological diversity of mouse lemurs (*Microcebus* spp.) in northern Madagascar: The discovery of a putative new species? *American Journal of Primatology* 81: e23070.

Shafer, A. B. A., and Wolf, J. B. W. 2013. Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology. *Ecology Letters* 16: 940-950.

Slatkin, M., and Hudson, R. R. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129: 555-562.

Sommer, S., Schwab, D., and Ganzhorn, J. U. 2002. MHC diversity of endemic Malagasy rodents in relation to geographic range and social system. *Behavioral Ecology and Sociobiology* 51: 214-221.

Springer, A., Mellmann, A., Fichtel, C., and Kappeler, P. M. 2016. Social structure and *Escherichia coli* sharing in a group-living wild primate, Verreaux's sifaka. *BMC Ecology* 16: 6.

Springer, M. S., Cleven, G. C., Madsen, O., de Jong, W. W., Waddell, V. G., Amrine, H. M., and Stanhope, M. J. 1997. Endemic African mammals shake the phylogenetic tree. *Nature* 388: 61-64.

Stanhope, M. J., Waddell, V. G., Madsen, O., Jong, W. de, Hedges, S. B., Cleven, G. C., Kao, D., and Springer, M. S. 1998. Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proceedings of the National Academy of Sciences of the USA* 95: 9967-9972.

Steiner, C. C., Putnam, A. S., Hoeck, P. E. A., and Ryder, O. A. 2013. Conservation genomics of threatened animal species. *Annual Review of Animal Biosciences* 1: 261-281.

Storz, J. F. 1999. Genetic consequences of mammalian social structure. *Journal of Mammalogy* 80: 553-569.

Storz, J. F., and Beaumont, M. A. 2002. Testing for genetic evidence of population expansion and contraction: An empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution* 56: 154-166.

Szulkin, M., Stopher, K. V., Pemberton, J. M., and Reid, J. M. 2013. Inbreeding avoidance, tolerance, or preference in animals? *Trends in Ecology & Evolution* 28: 205-211.

Sugg, D. W., Chesser, R. K., Stephen Dobson, F., Hoogland, J. L. 1996. Population genetics meets behavioral ecology. *Trends in Ecology & Evolution* 11: 338-342.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L. P., and Bouvet, J. 1996. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* 24: 3189-3194.

Tajima, F. 1989a. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Tajima, F. 1989b. The effect of change in population size on DNA polymorphism. *Genetics* 123: 597-601.

Tang, Q., Fung, T., and Rheindt, F. E. 2020. ResDisMapper: An r package for fine-scale mapping of resistance to dispersal. *Molecular Ecology Resources* 20(3): 819-831.

Tattersall, I. 1982. *Primates of Madagascar*. New York: Columbia University Press.

Tattersall, I. 2007. Madagascar's lemurs: Cryptic diversity or taxonomic inflation? *Evolutionary Anthropology* 16: 12-23.

Tattersall, I., and Cuozzo, F. P. 2018. Systematics of the extant Malagasy lemurs (order Primates). In *The terrestrial protected areas of Madagascar: Their history, description, and biota*, eds. S. M. Goodman, M. J. Raherilalao, S. Wohlhauser, pp 403-424. Antananarivo: Association Vahatra.

Terhorst, J., Kamm, J. A., and Song, Y. S. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics* 49: 303-309.

Thalmann, U., Kümmerli, R., and Zaramody, A. 2002. Why *Propithecus verreauxi deckeni* and *P. v. coronatus* are valid taxa – quantitative and qualitative arguments. *Lemur News* 7: 11-16.

Tokiniaina, H., Bailey, C. A., Shore, G. D., Delmore, K. E., Johnson, S. E., Louis, E. E., and Brenneman, R. A. 2009. Characterization of 18 microsatellite marker loci in the White-Collared Lemur (*Eulemur cinereiceps*). *Conservation Genetics* 10: 1459.

Tollenaere, C., Brouat, C., Duplantier, J.-M., Rahalison, L., Rahelinirina, S., Pascal, M., Moné, H., Mouahid, G., Leirs, H., and Cosson, J.-F. 2010. Phylogeography of the introduced species *Rattus rattus* in the western Indian Ocean, with special emphasis on the colonization history of Madagascar. *Journal of Biogeography* 37: 398-410.

Tomiuk, J., Bachmann, L., Leipoldt, M., Ganzhorn, J. U., Ries, R., Weis, M., and Loeschcke, V. 1997. Genetic diversity of *Lepilemur mustelinus ruficaudatus*, a nocturnal lemur of Madagascar. *Conservation Biology* 11: 491-497.

Tomiuk, J., Bachmann, L., Leipoldt, M., Atsalis, S., Kappeler, P. M., Schmid, J., and Ganzhorn, J. U. 1998. The impact of genetics on the conservation of Malagasy lemur species. *Folia Primatologica* 69: 121-126.

Trochet, A., Courtois, E. A., Stevens, V. M., Baguette, M., Chaine, A., Schmeller, D. S., Clobert, J., and Wiens, J. J. 2016. Evolution of sex-biased dispersal. *The Quarterly Review of Biology* 91: 297-320.

Veilleux, C. C., Louis, E. E., and Bolnick, D. A. 2013. Nocturnal light environments influence color vision and signatures of selection on the OPN1SW opsin gene in nocturnal lemurs. *Molecular Biology and Evolution* 30: 1420-1437.

Vekemans, X., and Hardy, O. J. 2004. New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* 13: 921-935.

Veron, G., Dupré, D., Jennings, A.P., Gardner, C. J., Hassanin, A., and Goodman, S. M. 2017. New insights into the systematics of Malagasy mongoose-like carnivorans (Carnivora, Eupleridae, Galidiinae) based on mitochondrial and nuclear DNA sequences. *Journal of Zoological Systematics and Evolutionary Research* 55: 250-264.

Veron, G., Dupré, D., Lührs, M.-L., Kappeler, P. M., Dollar, L., Pomerantz, J., and Goodman, S. M. 2018. Genetic polymorphism and structure of wild and zoo populations of the fosa (Eupleridae, Carnivora), the largest living carnivoran of Madagascar. *Mammalian Biology* 92: 68-77.

Vieilledent, G., Grinand, C., Rakotomalala, F. A., Ranaivosoa, R., Rakotoarijaona, J.-R., Allnutt, T. F., and Achard, F. 2018. Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. *Biological Conservation* 222: 189-197.

Vogler, B. R., Bailey, C. A., Shore, G. D., McGuire, S. M., Engberg, S. E., Fickel, J., Louis, E. E., and Brenneman, R. A. 2009. Characterization of 26 microsatellite marker loci in the fossa (*Cryptoprocta ferox)*. *Conservation Genetics* 10: 1449.

Wakeley, J. 1999. Nonequilibrium migration in human history. *Genetics* 153: 1863-1871.

Wang, I. J. 2012. Environmental and topographic variables shape genetic structure and effective population sizes in the endangered Yosemite toad. *Diversity and Distributions* 18: 1033-1041.

Wang, I. J., and Bradburd, G. S. 2014. Isolation by environment. *Molecular Ecology* 23: 5649-5662.

Wang, I. J., and Summers, K. 2010. Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology* 19: 447-458.

Wang, K., Mathieson, I., O'Connell, J., and Schiffels, S. 2020. Tracking human population structure through time from whole genome sequences. *PLoS Genetics* 16(3): e1008552.

Wasimuddin, Corman, V. M., Ganzhorn, J. U., Rakotondranary, J., Ratovonamana, Y. R., Drosten, C., and Sommer, S. 2019. Adenovirus infection is associated with altered gut microbial communities in a non-human primate. *Scientific Reports* 9: 13410.

Watterson, G. A. 1978. The homozygosity test of neutrality. *Genetics* 88: 405-417.

Weyeneth, N., Goodman, S. M., and Ruedi, M. 2011. Do diversification models of Madagascar's biota explain the population structure of the endemic bat *Myotis goudoti* (Chiroptera: Vespertilionidae)? *Journal of Biogeography* 38: 44-54.

Williams, R. C., Blanco, M. B., Poelstra, J. W., Hunnicutt, K. E., Comeault, A. A., and Yoder, A. D. 2020. Conservation genomic analysis reveals ancient introgression and declining levels of genetic diversity in Madagascar's hibernating dwarf lemurs. *Heredity* 124: 236-251.

Wimmer, B., and Kappeler, P. M. 2002. The effects of sexual selection and life history on the genetic structure of redfronted lemur, *Eulemur fulvus rufus*, groups. *Animal Behaviour* 64: 557-568.

Wimmer, B., Tautz, D., and Kappeler, P. M. 2002. The genetic population structure of the gray mouse lemur (*Microcebus murinus*), a basal primate from Madagascar. *Behavioral Ecology and Sociobiology* 52: 166-175.

Wright, S. 1943. Isolation by distance. *Genetics* 28: 114-138.

Wyner, Y. M., Johnson, S. E., Stumpf, R. M., and DeSalle, R. 2002. Genetic assessment of a white-collared × red-fronted lemur hybrid zone at Andringitra, Madagascar. *American Journal of Primatology* 57: 51-66.

Yoder, A. D., Vilgalys, R., and Ruvolo, M. 1996. Molecular evolutionary dynamics of cytochrome b in strepsirrhine primates: The phylogenetic significance of third-position transversions. *Molecular Biology and Evolution* 13: 1339-1350.

Yoder, A. D. 2013. The lemur revolution starts now: The genomic coming of age for a non-model organism. *Molecular Phylogenetics and Evolution* 66: 442-452.

Yoder, A. D., Burns, M. M., Zehr, S., Delefosse, T., Veron, G., Goodman, S. M., and Flynn, J. J. 2003. Single origin of Malagasy Carnivora from an African ancestor. *Nature* 421: 734-737.

Yoder, A. D., Campbell, C. R., Blanco, M. B., Reis, M. dos, Ganzhorn, J. U., et al. 2016. Geogenetic patterns in mouse lemurs (genus Microcebus) reveal the ghosts of Madagascar's forests past. *Proceedings of the National Academy of Sciences of the USA* 113: 8049-8056.

Box 1. Measures of Genetic Diversity and Definitions

Number of alleles (n_A): The number of recognized variants observed in a sample of individuals for a specific locus (or stretch of DNA). It is the simplest measure of genetic diversity but it does not account for the relative frequencies of alleles, and is strongly influenced by the sample size.

SNPs (single nucleotide polymorphism) or segregating sites: SNPs are genetic variants observed at the level of a single nucleotide. SNPs were called "segregating sites" in classical population genetics studies but the term SNP is now dominant. In many cases SNPs are assumed to have only two alleles, one considered ancestral and the other one derived.

Observed heterozygosity (H_{obs}): The proportion of actual heterozygotes in the sample averaged across the loci that have been genotyped. H_{obs} varies between 0 and 1 and can only be computed by knowing the observed genotypes for the sampled (diploid) individuals. It is thus a measure of *individual* diversity averaged across the loci and sampled individuals. Contrary to a common belief it is not necessarily a reliable measure of *population* diversity because it is influenced by the mating system. It can thus be very low even when there are many alleles in the sample or population, and therefore even when genetic diversity is high. An extreme example would be one in which all individuals are homozygous for different alleles at the same locus: H_{obs} would be zero since no heterozygotes would be observed, but H_{exp} (see below) would be maximal.

Expected heterozygosity (H_{exp}): The proportion of heterozygotes one would observe in a sample, if the observed alleles were distributed randomly within the sampled diploid individuals. This is thus the proportion of heterozygotes that one expects under random mating (or at Hardy-Weinberg equilibrium), i.e. under panmixia. In other words, H_{obs} and H_{exp} should be equal under random mating. Differences between H_{exp} and H_{obs} are usually interpreted in terms of departures from Hardy-Weinberg equilibrium. H_{exp} is also called *gene diversity* and varies between 0 and 1. It can be computed by knowing the observed *allele* frequencies

(p_i) as $H_{exp}= 1- \Sigma p_i^2$. Genotypes are thus not necessary. This is why it can be computed for haploid markers (such as mitochondrial DNA data) and is then called *haplotype diversity*.

Expected heterozygosity under mutation drift equilibrium (H_{exp-eq}): This is the value expected for H_{exp} when a random mating population has reached mutation drift equilibrium (see below). At mutation-drift equilibrium, Ewens (1972) has shown that we can predict the expected frequencies of observed alleles (p_{i_eq}) by only knowing their number (n_A), the sample size (k) and the population mutation rate (the product of the population size, N, and per locus mutation rate, u). Ewens' results thus allow the computation of $H_{exp-eq}= 1- \Sigma(p_{i_eq})^2$ without using the actual, observed allele frequencies (p_i) which are required for H_{exp}. By comparing H_{exp} and H_{exp-eq} we can thus test whether the population is at mutation-drift equilibrium (assuming random mating) or not (see Box 2). When testing for mutation drift equilibrium (see below) H_{exp} is thus considered as the *observed* H_{exp} because its computation requires *observed* allele frequencies (p_i).

Allele frequency spectrum (AFS) also called site frequency spectrum (SFS): Assuming that we have genotyped a large number of independent bi-allelic loci for a sample of size k, and that we can identify an ancestral and a derived allele, the AFS (or SFS) is a histogram with k+1 values obtained by counting the number of loci having either 0, 1, 2, … or k copies of the derived allele in the sample. It is thus a genomic description of the frequencies of SNPs. Theoretical studies have shown how the shape of the AFS changes with the demographic history of species. It can thus be used as a summary of genomic data to detect population size changes (Excoffier et al. 2013; Liu and Fu 2015), or population structure (Excoffier et al. 2013). The 2D AFS can also be computed as a matrix for two populations and can be used to infer parameters in models with two populations such as splitting times (Gutenkunst et al. 2009; Excoffier et al. 2013).

Mismatch distribution or distribution of pairwise differences: Histogram obtained for a sample of k sequences of the same locus by computing the number

of differences (substitutions) for all k*(k-1)/2 possible pairwise comparisons. The shape of the mismatch distribution is influenced by the demographic history of species and can be used to detect departures from mutation-drift equilibrium. Under panmixia and neutrality it can be used to detect and date bottlenecks and expansions.

F_{ST}: Used as a measure of genetic differentiation, it actually measures fixation of different alleles in different populations. It varies between 0 (when the allele frequencies are identical between populations) and 1 (when populations have lost genetic diversity and are fixed for different alleles).

Effective population size (Ne): Real populations exhibit complex features such as different numbers of males and females, variation in the reproductive success of individuals from the same of from different sexes, or variation in population size through time. The effective population size is the size of an ideal population (constant size, same number of males and females, and a Poisson variation in the number of offspring) that would have the same properties as the real population from the point of view of some genetic measure such H_{exp}. The concept of N_e is central in population genetics as it measures the importance of genetic drift.

Ascertainment bias: This is a statistical bias due to the non-random discovery of genetic markers. For instance, to minimize genotyping costs, microsatellites developed in one species are typically selected as a function of their variability in the initial sample. Typing the same markers in other populations or species will lead to an overestimation of genetic diversity in the initial sample and an underestimation in the other samples, even within the same species. Ascertainment bias is not necessarily problematic to identify genetic clusters (population structure), but it is problematic for demographic inference, because it biases the representation of genomic diversity and the AFS.

Box 2. Demographic Inference and Coalescent Theory

Mutation-drift equilibrium: Mutation-drift equilibrium is reached when mutations (that create new alleles) and genetic drift (that randomly eliminates alleles) balance each other. Ewens (1972) has shown that at mutation-drift equilibrium it is possible to predict the *expected* frequencies of alleles observed in a sample (p_{i_eq}) under the Wright-Fisher model (constant size, non-overlapping generations, and neutrality). This equilibrium is very different from Hardy-Weinberg equilibrium (HWE). HWE is reached after only one generation of random mating whereas mutation-drift equilibrium is an evolutionary equilibrium that may never be reached. The results of Ewens (1972) were published in a period of controversy between neutralists and selectionists. As a consequence, a departure from mutation-drift equilibrium was interpreted as a possible sign of selection (Watterson 1978), and led to the classical Ewens-Watterson test, based on the comparison of homozygosity values (i.e. $1\text{-}H_{exp}$). However, Ewens clearly stated that his results were obtained under an evolutionary set of models that made several assumptions such as the infinite allele mutation model, neutrality, panmixia, constant size, and no population structure. Thus, departures from *any* of these assumptions, or combinations thereof could potentially explain a statistically significant result. In the 1980-1990s the population genetics community started to be interested in population size changes, and new statistics were developed to detect departures from mutation-drift equilibrium. Some were presented as detecting selection, while others were introduced as detecting population size changes (Tajima 1989a, 1989b; Fu 1997). Theoreticians were aware of these interpretation issues, and knew that they were in reality proposing improvements on the original Ewens-Watterson test. For instance, Tajima (1989a) developed the D statistic (now Tajima's D) as a way to test neutrality. He, however, also showed that the D statistic was also influenced by population size changes (Tajima 1989b).

Mutation models and mutation-drift equilibrium: Demographic inference depends on assumptions on the mutation model. The analytical results of Ewens (1972) were developed for the infinite allele model. Simulations have shown that significant departures from mutation-drift equilibrium could also be detected not because of selection or population size change, but because the wrong mutation model was assumed

Population structure and mutation-drift equilibrium: In the last 20 years there has been an increasing recognition that population structure and the sampling strategy can generate signals in the genetic data that can be interpreted as population size changes under the assumption of local panmixia and the absence of connectivity with other populations (Chikhi et al. 2010; Heller et al. 2013; Mazet et al. 2016).

Selection and the rest: We have mainly focused on neutral processes, but do not wish to minimize the likely effects of selection. There are, however, many types of selection and it is not clear how important they have been in shaping patterns of genomic diversity. One general argument used by population geneticists is that neutral forces shape the whole genome in a similar way, whereas loci under selection are expected to exhibit different signals, since they are subject to different types of selection. For several decades it was a common assumption that neutral loci represent thus the core of genetic data and loci under selection a minority that could be detected once we could predict the expected neutral patterns (Cavalli-Sforza 1966; Beaumont and Nichols 1996; Goldstein and Chikhi 2002). However, an increasing number of studies are now considering that some forms of selection, such as background selection, may directly or indirectly influence major parts of the genome (Comeron 2017; Pouyet et al. 2018) to the point that Comeron (2017) recently defended the view that the null model for genome evolution should include background selection. In addition, other molecular processes such as recombination, biased gene conversion, etc.

may play an important role in shaping patterns of diversity at the level of genomes.

Coalescent theory, strength and limitations: Theory developed in the 1980s and 1990s to study the statistical properties of gene trees (Kingman 1982). The standard coalescent theory assumes that we have a non-recombining and neutral DNA fragment (a locus) for which several copies (genes) have been sampled. It is typically assumed that the size of the population, N, is constant and that the population is panmictic and isolated from other populations. The coalescent theory is a retrospective method, as it starts with the sample (which could be present-day material of those from museum or ancient DNA specimens) and then goes backward until the MRCA (most recent common ancestor) of the sample is reached, at a time called T_{MRCA}. The coalescent theory allows population geneticists to better understand the very large variance associated with gene trees. Under the same evolutionary assumptions, two loci can exhibit different gene trees with very different T_{MRCA}. This is why it is difficult to draw strong conclusions regarding the demographic history inferred from a limited number of loci. This is particularly true for mtDNA, which is a notably large set of linked loci and thus behaves as a single locus. Coalescent theory thus differs from the theoretical work done before the 1980s, a period in which theoreticians tended to use a forward-looking approach, starting from a particular situation and predicting various possible outcomes, such as the fate of advantageous or deleterious mutations. Forward-looking theory is population-based, since whole populations need to be modelled or simulated. The coalescent theory is sample-based, as it focuses on the properties of the sample. These properties are a function of the whole population (or populations in structured models) but coalescent simulations are computationally much more efficient as they mainly depend on the size of the sample, not the size of the population. This explains why they became extremely popular and have been at the heart of population genetics for three decades (taking Hudson's 1990

review article as a possible start for its widespread use). Coalescent theory is extremely important but it also has limitations. First, it typically assumes that the sample size (k) must be much smaller than the population size (N), which is assumed to be very large (k<<N). For rare species, which are often endangered, this is rarely true. How good the approximation is depends on many factors that are little discussed. For instance, it also assumes that the probability to have several coalescent events is very small. In species with social units and a high variance in the reproductive success, this is unlikely to be true.

In the 1990s several studies have shown that the shape of gene trees is influenced by demographic events such as bottlenecks and expansions, hence allowing us to detect, quantify and date such events. Most theoretical work is however dependent on the standard coalescent theory, which assumes panmixia. This means that two important levels of population structure have often been ignored. The first is the within-population structure, since the consequences of social structure generating variance in reproductive success have not been very much studied. Second, the existence of gene flow and connectivity among populations has also been ignored. Coalescent theory allows to account for population structure (Herbots 1994; Wakeley 1999), but not all methods account for that as we discuss in the main text.

Likelihood-based and Bayesian methods: The objective of Likelihood-based approaches is to compute the probability P_M(Data | Param) of generating the observed Data under a pre-defined demographical model M, defined by a set of parameters Param = (Par_1, . . . , Par_k). The vertical bar represents a conditional probability and means “given that”. While this is a probability, it is crucial to stress that in a likelihood framework, we compare probabilities across parameter values. We thus do not compare probabilities of alternative outcomes (for one specific set of parameter values), which would then sum to one. For instance, we may want to compute the probability

that a migrant individual originated from a specific source population based on its genotype and allele frequencies in that population. Let us call this probability P_M(Genotype | Pop Origin). If there are three potential source populations with similar allele frequency, we may find that the three probabilities are similar and close to say 50%. In that case the sum would be > 100%. We might also find that the probabilities are small, say less than 1%, and in that case the sum would be <<100%. Likelihoods are thus probabilities that are seen as functions of the model parameters (Param) and therefore of the model. Importantly, the model M is assumed to be true and the Data are observed. Thus, P_M(Data | Param) is typically rewritten as either L_M(Param | Data), by inverting the order of Param and Data, or simply L(Param) (implicitly, we mean "given the observed data"). Likelihood methods are used to find the parameter values that maximize the function L(Param). These maximum likelihood estimates (MLE) are then used as the best estimates explaining the observed Data. Other likelihood-based methods may use a Bayesian perspective and estimate probability density functions for these parameters (rather than a point estimate like the MLE). To do that they require the use of *prior* distributions of the parameters P_M(Param) that specify the values that the parameters can take and with which probability. For instance, we could state *a priori* that the three source populations have the same *prior* probability or we may have some prior information suggesting that the first population has a higher probability. This allows the Bayesian analysis to compute a *posterior* distribution, which is obtained using the Bayes formula: P_M(Params | Data) = P_M(Param) * P_M(Data | Param) / P_M(Data). Note that the prior and posterior distributions are both probabilities of the parameters. One can be seen as the probability before we see the data and the other after we see the data and they are both conditional on the model M that we assumed was true.

Approximate Bayesian computation (ABC): General simulation-based approach used to infer parameter values of a demographic model when the

likelihood function is difficult or impossible to evaluate (see above). This is particularly useful to compare complex models with changes in connectivity or population size, as there is not simple way to compute the likelihood for such models. Here the idea is to approximate the likelihood function and the data. The first step requires a summary of the data by a limited number of summary statistics (such as n_A, Tajima's D or H_{exp}, see Box 1). Thus, instead of computing or estimating the full likelihood P_M(Data | Param), the objective is to estimate P_M(SummData | Param), where SummData is a vector of summary statistics. In order to estimate this probability, a very large number of simulations are carried out (usually $>10^5$-10^6) and the summary statistics used to represent the data are computed for each single simulation. The simulations producing summary statistics closest to the observed ones (for the real data) are selected (usually the 1% or 0.1% best). Then the parameter values corresponding to these "optimal" simulations are kept for inference. There are thus two levels of approximation: the data are approximated by summary statistics and the fit to the data is approximated by the closest simulations. The term ABC was introduced by Beaumont et al. (2002) but the principle had been independently used by several authors in the 1997-1999 (Beaumont et al. 2002, see Chikhi and Beaumont 2005; Beaumont 2010 for reviews). The ABC framework has been particularly useful in population genetics thanks to the coalescent theory and simulation software such as *ms* (Hudson 2002), which simulate genetic data very efficiently. On Madagascar, the ABC framework has been used in only a limited number of studies (Brouat et al. 2014; Salmona et al. 2017) but could be important to develop comparative studies across species and regions.

Box 3. Genetic Markers

Polymorphism and Sex-linked Markers: The Diversity of Genetic Markers and of their Use

One of the first steps in population genetic studies is to describe patterns of genetic diversity and differentiation at different geographical scales. This can be done with different genetic markers that can differ by their variability or by their transmission mode (see below for an overview of marker types and methods). Allozymes and (nuclear) microsatellites both inform on nuclear diversity and are both transmitted by males and females, but they differ in their respective levels of variability. Mitochondrial DNA and Y chromosome data are only transmitted by females and males, respectively, and therefore can provide unique information on female- and male-specific behavior, such as dispersal, or demographic history. However, important differences exist between mtDNA and Y chromosome data. MtDNA data can be collected from both males and females, whereas only males have Y chromosomes. Also, the molecular mechanism generating mutations and the mutation rate in the genetic markers that can be obtained from mtDNA (DNA sequences that vary in mutation rates along the mtDNA molecule) and Y chromosome (DNA sequences and linked microsatellites) data can be rather different. They also differ from autosomes in that their effective population size (N_e, Box 1) is much smaller than that of nuclear loci and they are therefore more subject to genetic drift. X chromosomes are transmitted by both males and females, but since there is only one copy in males, the effective size of X chromosome is smaller than that of autosomes.

While we are not aware of studies examining Y or X chromosome diversity in Malagasy mammals, recent genomic studies suggest that it may now become possible to extract both Y and X chromosome data from ongoing sequencing projects, which have produced full genome sequences (Meyer et al. 2015 for *Eulemur macaco* and *E. flavifrons*; Hawkins et al. 2018 for *Prolemur simus*; Williams et al. 2020 for several *Cheirogaleus* species). Finally, most genetic studies have usually focused on genetic markers thought

to behave neutrally, but a number of genetic studies have also focused on regions thought to be under selection (Sommer et al. 2002 on *Hypogeomys antimena*; Schwensow et al. 2007 on *C. medius*; Hunnicutt et al. 2020 for *Microcebus*). The genetic markers that have been used to study Malagasy mammals varied over time (see main text).

Immuno-diffusion: Immuno-diffusion consists in comparing the antigens (proteins, lipids, saccharides, etc.) of different taxa to quantify their genetic divergence. It is based on a reaction between antibodies and antigens diffused in an agarose gel. Patterns observed in the gels allow to determine whether antigens are identical, similar or dissimilar. It is then possible to build similarity matrices and species trees (e.g., Dene et al. 1976, 1980).

Cytogenetics: Cytogenetic studies quantify chromosomal differences across taxa (differences in number of chromosomes, centromere position, re-arrangements such as inversions, etc.). Metaphase chromosomes (condensed chromatin) are stained with special dyes that uncover the unique banding pattern of chromosome pairs. It has been extensively used for phylogenetic studies, taxonomy and hybridization studies before DNA sequencing became more accessible to researchers. One of the main advantages of this technique is that it allows to easily detect large chromosomal rearrangements and investigate the role of these rearrangements as reproduction barriers between taxa. Today, new techniques use fluorescently labelled probes and, more recently, genomic data to identify chromosomal rearrangements (e.g., Rumpler and Dutrillaux 1976; Eisfeldt et al. 2019).

Allozymes: General term used to identify protein variants from the same locus that can be separated on a gel. Allozymes revolutionized population genetics as they allowed population geneticists to easily identify large

amounts of single locus polymorphisms (Hubby and Lewontin 1966). They have been rarely used with Malagasy mammals (see main text).

PCR (Polymerase Chain Reaction): Molecular technique that produces a very large number of copies of a short pre-specified region of DNA. A thermo-resistant enzyme (DNA polymerase) synthetizes new DNA fragments that are complementary to the region of interest, delimited by a pair of oligonucleotides (short artificial DNA fragments called primers) flanking the region to be amplified, each being complementary to one of the two DNA strands. The primers are used to initiate DNA synthesis. The PCR technique is widely used and recognized as a revolutionary method that makes use of small amounts of tissue and DNA (e.g., fecal samples).

RAPD (Random Amplification of Polymorphic DNA): PCR-based technique used to detect polymorphism in the genome. While a PCR typically needs two primers, each binding on either side of the DNA fragment to be amplified, RAPDs only use one short primer. If this primer is short enough, it is expected that it will randomly bind in a very large number of genomic regions. When pairs of locations are close enough to each other to be effectively amplified by PCR, a band will be produced. RAPDs thus focus on the banding patterns across individuals or populations. On Madagascar they have been often used to quantify genetic diversity or differentiation between populations since the early 1990s.

RFLP (Restriction Fragment Length Polymorphism): Polymorphism based on the presence or absence of short sequences (typically between 4 and 8 nucleotides) called restriction sites. These sequences are recognized by restriction enzymes that cut the DNA, hence, generating DNA fragments of variable length. This approach has been extensively used to produce banding patterns that are visualized through gel electrophoresis. Variation in

banding patterns across individuals from different species or populations can then be used for evolutionary inference.

Sanger sequencing: DNA sequencing approach based on the incorporation of nucleotides labelled with a fluorescent dye that can be easily detected by a sequencing machine. It typically reads sequences between a few hundreds and ~1 Kbp. The Sanger method has been the dominant approach for decades and is still used for the analysis of small datasets (in number of loci) (e.g., Everson et al. 2016; Sgarlata et al. 2019). This method has also been used for sequencing mitogenomes (~16 Kbp) in several lemur species, mainly *Microcebus* spp. and *Lepilemur* spp. (Lei et al. 2017; Louis and Lei 2016). It has been particularly useful in studies that aim at characterizing target genomic regions, such as major histocompatibility complex (Schad et al. 2005; de Winter et al. 2019), vomeronasal receptors (Hohenbrink et al. 2017), opsin genes (Veilleux et al. 2013), or identifying microbe community within mammal hosts, such as viruses (Wasimuddin et al. 2019), bacteria (Gomard et al. 2016), and parasites (Kiene et al. 2020).

Microsatellites: A microsatellite locus is a DNA sequence composed of several repetitions of the same 2-8 nucleotides long motif. They became a marker of choice in conservation genetics because they are highly variable, non-coding, likely neutral, have genome-wide distribution, and could be polymorphic even in species with small population sizes. They also can be obtained from minute non-invasive samples, such as feces (e.g., Quéméré et al. 2010b; Bertoncini et al. 2017). Most conservation genetic studies, in Madagascar and elsewhere, typically use between 5 and 20 microsatellites, even if studies with more than 20 loci exist. Microsatellite loci identified in one species can often be used in other species, but this can typically generate strong biases in demographic inference (see ascertainment bias Box 1).

Next-generation sequencing (NGS): DNA sequencing technologies that provide genomic data at relatively low cost, thanks to the parallel and massive production of relatively short sequences. Depending on the NGS technique, short sequences are then assembled to build larger DNA fragments, from few hundreds to tens of millions of base-pairs. The advent of NGS methods allows to address questions to be addressed that could not be done rapidly with Sanger sequencing. For example, one of the widely used NGS machines in 2010 (Illumina HiSeq 2000) was able to sequence DNA 10,000 times faster and 0.0001 times cheaper than Sanger sequencing (Glenn 2011). This is a field in constant evolution and new NGS approaches can now obtain long sequences (~1 Mbp), which facilitate the assembly of complex genomic regions (e.g., highly repetitive) or the study of chromosomal-scale evolution. These technologies can be adapted to target specific genomic regions, acquire large-scale data on the transcription of thousands of genes or rapidly sequence whole genomes. Working with genomic data has major challenges related to 1) the computational power needed to assemble massive amount of data into reliable genomic information, 2) the capacity of storing large amount of processed data, and 3) the identification of the best parameters to be used in the genomic data processing pipelines.

For Malagasy mammals, NGS approaches have been used to sequence DNA extracted from fecal samples to study the diet of *Propithecus tattersalli* (Quéméré et al. 2013) and different bat species (Kemp et al. 2019), or to characterize the gut microbial community of several lemurs (Springer et al. 2016; Greene et al. 2019; Perofsky et al. 2019). Few studies have explored profiles of gene transcription in Malagasy mammals, but an example are the studies of Faherty et al. (2016, 2018), which have used NGS technology to investigate the gene-expression profile of adipose tissue in *Cheirogaleus crossleyi* during hibernation.

Restriction site associated DNA-sequencing (RAD-seq): A next-generation sequencing method that produces a massive amount of short DNA

sequences flanking a specific restriction site (see RFLP). The genomic information retrieved represent a non-random sample of the whole genome (typically less than 1%), since only regions around the restriction sites are sequenced. There are several variants of the RAD-seq technique, which mainly differ in the number of restriction enzymes used to digest the genome, the adaptor ligation, and the barcoding and size selection of the ideal DNA fragment length used for sequencing (Andrews et al. 2016). RAD-seq is particularly adapted for classical population genetic studies of non-model organisms, because it does not require prior knowledge on the genome of the species of interest (model organisms such as mice, humans, and *Drosophila*, have extremely good genetic and genomic resources, including reference genomes that are necessary to statistical analyses of population genomic data with limited coverage) and provides a large amount of population genomic data at a reasonable cost. The number of independent DNA fragments and their coverage (the number of times the same fragment is read, which represents a measure of genotyping reliability) depends also on the frequency of the restriction enzyme cutting site and it is typically on the order of 10^3-10^6 (locus size is 30-500 bp). The term ddRAD is used when two different restriction enzymes are used (e.g. Yoder et al. 2016 on *Microcebus*) typically leading to reduction in the total number of loci. These data can be used to study population diversity and differentiation, they can also be summarized using the AFS (see Box 1) for demographic inference (Excoffier et al. 2013) or examined for detecting loci under selection (Hohenlohe et al. 2010). So far, few Malagasy mammal species have been studied with RAD-seq data, and include Salmona et al. (2015) on *Microcebus* and *Lepilemur*, Yoder et al. (2016) and Poelstra et al. (2020) on *Microcebus*.

Whole-genome sequencing (WGS): NGS approaches have opened a new avenue for the sequencing of full genomes, even for wild non-model organisms. WGS has two main challenges: 1) the acquisition of sequence data spread across the whole genome; and 2) the assembly of these

sequences to compose chromosome-scale contiguous sequences. The fast evolution of sequencing technologies has given the possibility to sequence both short-reads (~150 bp) and long-reads (~10 Kbp), and combine the information of these two different types of approaches for the assembly of a full genome. Among many parameters that quantify the quality of a genome, two are widely used: coverage, which defines the number of reads representing a given genomic region; and N50 scaffold/contig, which is a standardized measure of sequence contiguity of the assembled genome. In lemurs, the first draft-genome assemblies were available in the early 2010s, with the release of *Microcebus murinus* (coverage: ~1.9X and N50 scaffold: 140 Kbp; Lindblad-Toh et al. 2011) and *Daubentonia madagascariensis* (coverage: ~19X and N50 scaffold: ~14 Kbp; Perry et al. 2012) draft-genomes. The cost of WGS has subsequently decreased and it is now possible to obtain sequencing data from non-model organisms with a 20-30X coverage and N50 contigs of 10–30 Kbp for 500-1500 Euros per individual.

Box 4. Spatial Genetics

Isolation-by-Distance

A pattern known as 'isolation-by-distance' (IBD) can be observed when dispersal and reproduction are limited by geographical distance (Wright 1943). Theoretical work has shown that IBD patterns can be informative on several species-specific parameters, such as the mean squared axial parent-offspring distance (σ^2) and the neighborhood size ($4\pi D_e\sigma^2$), where D_e is the effective population density. For instance, Rousset (1997) demonstrated that in a homogeneous two-dimensional habitat accurate estimates of the neighborhood size can be obtained by computing the inverse of the slope of the regression between genetic and geographic distance between pairs of populations, namely between $F_{st}/(1-F_{st})$ and ln(d), the natural logarithm of the geographical distance. Similar theoretical results have also been found if one computes the average kinship between pairs of individuals separated by geographic distances that fall within a set of distance classes (Vekemans and Hardy 2004). Using estimates of effective population density (D_e), one could in principle derive, from the estimated neighborhood size ($4\pi D_e\sigma^2$), an estimate of the area of gene dispersal (σ^2), thus an estimate of the effective dispersal distance (σ). This approach has been implemented in the software SPAGeDI (Hardy and Vekemans 2002) and has been applied to microsatellite data across various taxa outside Madagascar, but not yet on the island. One main limitation of this approach is that effective population density and effective dispersal distance cannot be inferred at the same time, which means that it is necessary to estimate one of the two in an "independent" manner.

However, Ringbauer et al. (2017) have recently developed a method that does not require an independent estimate of D_e or σ^2, and uses the length distribution of identical genome tracts shared between pairs of individuals (identity-by-descent blocks) with known geographical location to accurately infer both D_e and σ, and recent changes in D_e. This method is probably just one among many that will likely be used in the next years. It requires either dense genome-wide data or whole-genome sequences from a few dozen

individuals of a given species and a linkage map. This type of data is currently difficult to obtain for most non-model organisms, but we hope that the situation will change on Madagascar and that appropriate genomic data will become more common in the future (e.g., Hunnicutt et al. 2020; Williams et al. 2020).

From IBD to Isolation-by-Resistance and Isolation-by-Environment

Unlike the approaches described in the section above, which assume homogeneous D_e and σ across the continuous habitat of the complete study region, several other methods have been proposed to actually identify environmental features (biotic and abiotic) that influence dispersal and may thus represent either barriers to gene flow or corridors at the scale of the landscape (Hendry 2004; Nosil et al. 2008; Sexton et al. 2014). For instance, Cushman (2006) proposed "Causal Modelling" (CM) as an approach to identify the most important landscape features (rivers, roads, agricultural lands, etc.) influencing gene flow. The idea is to build a resistance map based on these features and identify the least-cost path connecting every sampled pair. Genetic differentiation and least-cost path distances are then correlated across all samples. The features showing significant relationship with genetic distance are then considered to be those having higher impact on dispersal. CM analysis, like many landscape genetic methods, is difficult to validate as it would require heavy spatial simulations that, to our knowledge, have not yet been carried out. It also requires an *a priori* list of landscape features to compute the least-cost path of dispersal. Finally, it only considers this least-cost path between sample pairs rather than multiple dispersal pathways. McRae (2006) proposed a solution to the problem of multiple dispersal pathways by developing a model called "Isolation-By-Resistance" (IBR) that incorporates all possible paths and improves predictions of genetic differentiation. It also allows to estimate a surface of connectivity among samples (called "current map"), determining the areas that facilitate or restrict gene-flow (Anantharaman et al. 2020). In these two approaches is however impossible to separate the effect of geographically-limited dispersal and

environmental variables on gene-flow. Thus, some authors have gone further and proposed the concept of “Isolation-By-Environment or By-Ecology” (IBE), allowing to assess the relative contribution of an environmental variable and geographical distance in limiting species gene-flow (Wang and Summers 2010; Wang and Bradburd 2012; Bradburd et al. 2013). If IBE is widespread it could be studied to better understand the ecology of local adaptation (e.g., Bolnick and Otto 2013), to provide evidence for incipient ecological speciation (e.g., Shafer and Wolf 2013) or to identify biased-dispersal resulting from preference for a particular environment (Davis and Stamps 2004). While potentially interesting, it is unclear whether IBE patterns are informative about ecological speciation or adaptation, but Madagascar could be an appropriate place to test such theories, as discussed in the main text.

Identify Changes in Connectivity over Time

IBD, IBR, and IBE models, in the way in which they are currently used, assume that genetic patterns have reached ‘equilibrium’, meaning that they will not change over time and that they reflect the effect of present landscape features on gene-flow. Present day genetic diversity may actually be significantly influenced by past landscape structure. This is because certain genetic measures may take some time before showing significant changes (e.g., Landguth et al. 2010). Recently, Bradburd et al. (2016) have introduced a method that implicitly infers historical rates of dispersal between locations. Compared to the previous methods, the aim is to identify changes in connectivity and possible regions of expansion over long periods of time. This method uses genome-wide data and an IBD model to build a map in which population locations reflect genetic rather than geographic proximity (‘geogenetic map’). Its objective is to distinguish the genetic signal of local and non-local dispersal. The underlying reasoning states that if populations show a level of genetic similarity that is high compared to what would be expected given their geographical distance, this could be interpreted as a signature of admixture. Similarly, if two populations have a lower level of

genetic similarity compared to their geographical distance, this would suggest the presence of a barrier to dispersal between the two populations. The method then draws arrows on the inferred geogenetic map to indicate possible sources and targets of admixture. Note that although Bradburd et al. (2016) approach is useful to represent and visualize historical migration patterns (see main text for an application with *Microcebus*), it cannot really distinguish events that have occurred at different time intervals in the past. This is because the information used from these genetic data represent a summary of many historical events over time scales that are not always easy to separate. Other methods that we will not detail here (e.g., Al-Asadi et al. 2019) have proposed new ways to retrieve fine-scale information from genomic data, identifying how dispersal rate and population size vary in space and in recent times. This suggests that we will probably see the development of new models and methods that will better tease apart the effect of space and time on genetic patterns, allowing us to improve our inference of the changes in landscape connectivity over time.

Box 5. Miscellaneous Genetic Studies on Malagasy Mammals

The questions addressed in the main text provide a review many of the genetic studies carried out on Madagascar mammals. However, there were additional studies that could not be easily integrated in the different sections but deserve being mentioned here. While their numbers are still limited, this may change in the future as the use of NGS technologies will allow the addressing of questions that before 'genomic era' were technically challenging, time-consuming, and expensive, including the detection of bacteria and viruses in wild or captive Malagasy mammals (see Chapter 6), or the analysis of their diet or microbiota.

One rather isolated example is the study of Han and Woroby (2012) who reanalyzed whole-genome shotgun sequences (~38X coverage) to identify foamy virus-like insertions within the *Daubentonia madagascariensis* genome. Using comparative phylogenetic analyses with foamy virus (family Retroviridae, genus *Spumavirus*) sequences inserted in other primates, they found a strong correlation between the foamy virus and primate phylogeny; hence, suggesting that foamy viruses might share an ancient co-evolutionary history with primates, as observed for other retroviruses (e.g., Cornelis et al. 2014; Katzourakis et al. 2014).

Several studies have used genetic data to detect and classify viruses and bacteria in wild and captive Malagasy mammals, but we will focus here on the former. Using classical Sanger sequencing Reynes et al. (2014) assessed the presence of hantavirus across several potential hosts of Malagasy mammals and found that the Thailand virus (THAIV) species, typically found in Asia and recently identified in Africa, was also present on Madagascar. They identified a new genetic variant in both invasive *Rattus rattus* (92 out of 231) and endemic Malagasy rodent *Eliurus majori* (n = 1).

More recently, et al. (2018) studied the prevalence of Hantavirus in urban and rural areas across the island. They sampled 1242 individual mammals across a larger number of taxa (introduced rodents: *R*. *rattus*, *R*. *norvegicus*, *Mus musculus*; introduced shrews: *Suncus murinus*; and endemic tenrecs: *Setifer setosus*, *Tenrec ecaudatus*, and *Hemicentetes semispinosus*). Interestingly they only found evidence of Hantavirus infection in *R*. *rattus* (12.4%) and *M*. *musculus* (1.6%).

Recent molecular studies have also examined the prevalence of ecto- and endoparasites in Malagasy mammals. For example, Blanco et al. (2013) used molecular data (COI mitochondrial gene) to identify *Ixodes* tick species in *Microcebus rufus*, *M*. *lehilahytsara*, *Cheirogaleus sibreei* and *C*. *crossleyi* sampled at two sites, one in the east (Ranomafana) and the other in the Central Highlands (Tsinjoarivo). Except for *M*. *rufus*, they found the same *Ixodes* species (*I*. *lemuris*) in the other three lemur species. Similarly, Kiene et al. (2020) have used genetic data to identify lice and tick species sampled across several lemur and rodent taxa (*M*. *murinus*, *M*. *ravelobensis*, *Eliurus myoxinus*, and *Rattus rattus*), and investigated the effect of habitat loss and fragmentation on the composition of ectoparasite communities in those small mammals. They observed differences in parasite prevalence between host species and a reduction in parasite prevalence in proximity to forest edges.

Klein et al. (2018) used molecular approaches to identify ectoparasite species in two sympatric mouse lemur species (*M*. *murinus* and *M*. *ravelobensis*), aiming at assessing the abiotic factors that best correlate with parasite prevalence. They showed a significant contribution of season on parasite infection. Interestingly, they also found differences in parasite prevalence between the two sympatric host species. They suggested that it could be related to species-specific differences in sleeping site ecology. A study by Schwensow et al. (2010a) investigated the host genetic determinants of parasite infection in wild Malagasy mammals. They

genotyped a fragment of the major histocompatibility complex class II DRB locus (MHC-DRB), for which evidence of parasite-driven selection has been reported (Schwensow et al. 2007, 2010b). They sequenced 149 *Cheirogaleus medius* and 180 *M*. *murinus*, and tested for the association between specific MHC-DRB alleles and the intensity of infection with certain intestinal parasites. In both species, they found significant positive association between a specific MHC allele and infection with helminths of the genus *Ascaris* and suggested a functional link between MHC-DRB sequence and parasite infection. A similar association study was carried out by Tollenaere et al. (2011) on *Rattus rattus* in relation to plague (*Yersinia pestis*) infection, aiming at determining the genetic markers associated with plague resistance in this species. They found 22 loci with high differentiation between resistant and susceptible populations, two of which were confirmed to be associated with survival in an experimental challenge performed injecting *Y*. *pestis* in individuals caught from resistant or susceptible populations.

Using NGS technology Springer et al. (2016) have investigated the transmission of *Escherichia coli* (a commensal gut bacteria) among 39 individuals of *Propithecus verreauxi* from Kirindy and found that *E*. *coli* sharing reflected *P*. *verreauxi* group membership, and in part intergroup interaction and sharing of space. Perofsky et al. (2017) have carried out high-throughput 16S rRNA sequencing on fecal samples collected from seven social groups of *Propithecus verreauxi* living in Kirindy (CNFEREF) Forest (group size ranging from four to nine individuals) to characterize their gut microbiome composition and assess the relative contribution of social structure, diet, habitat overlap and relatedness on gut microbiota sharing among individuals. They found that social group membership was the most significant factor explaining gut microbiota similarity, when controlling for other factors. More recently, Greene et al. (2018) observed that *P*. *coquereli* showed significant shifts in microbiota composition even with small changes in diet. Wasimuddin et al. (2019) carried out high throughput sequencing of the v4 region of the

ribosomal 16S on DNA extracted from fecal samples of wild *Microcebus griseorufus* to study how natural infection by Adenovirus (AdV) can induce significant changes in the gut microbiota. They found a decrease in microbial diversity, microbial interactions (positive and negative) and abundance of specific bacterial taxa in AdV+ individuals (N=44) compared to AdV- individuals (N=116).

Similar high throughput sequencing approaches have also been applied to study the diet of Malagasy mammals. Quéméré et al. (2013) carried out the first of such studies among Malagasy mammals, focusing on *P. tattersalli* individuals sampled across the entire range of the species (Loky Manambato region). The objective of this study was to assess the influence of habitat type on diet composition. They sequenced part of *trnL* intron (plast gene) using DNA extracted from fecal samples obtained during the dry season, and compared the sequenced DNA with a database of *trnL* sequences from plants sampled in the same region, together with public sequence databases. They found a large dietary diversity in *P. tattersalli*, with ~80 plant genera and 49 families consistent with the only known previous study based on feeding behavior observations (88 genera, Meyers 1993).

Recently, Kemp et al. (2019) used a similar DNA meta-barcoding approach to detect insect pest species present in the diet of six Malagasy bats species (*Chaerephon atsinanana*, *Mops leucostigma*, *Mormopterus jugularis*, *Myotis goudoti*, *Miniopterus manavi*, and *M. majori*) occurring in proximity of five villages near Ranomafana National Park. They extracted genomic DNA from 58 fecal samples and sequenced part of the COI mitochondrial locus. The DNA meta-barcoding analysis showed that all bat species fed on 11-13 arthropod orders, with Coleoptera being the most represented. They identified six known insect pest species and 17 potential pest taxa, among which two are known rice pests (*Spodoptera mauritia* and *Herpetogramma licarsisalis*) that were found in three out of six investigated bat species, thus

demonstrating the role of bats in pest control. Bats are also likely playing an important role in controlling insects known to transmit diseases, but this was not explicitly mentioned in this study.

Genetic data from wild Malagasy mammals have been also used to study the evolution of certain phenotypic traits, considered to be of relevance for present species diversity and ecology. For instance, Veilleux et al. (2013) sequenced the OPN1SW gene (associated to dichromatic color vision, ideal under light exposure) from 106 nocturnal lemurs, belonging to 19 different species from five genera (*Microcebus*, *Avahi*, *Lepilemur*, *Phaner*, and *Cheirogaleus*). They found that the OPN1SW gene of most species living in open canopy forests (spiny thicket and seasonally open dry deciduous forests) had stronger signature of purifying selection (less likely to have deleterious mutations in the gene) compared to species from closed canopy moist evergreen forest. In an earlier study, Larsen et al. (2014) made use of published gene-expression data from several primates, including lemurs (*Daubentonia madagascariensis*, *Varecia variegata*, *Eulemur mongoz*, and *Propithecus coquereli*), to characterize antibody diversity across wild-caught primate species. They found high similarity in the type of antibodies expressed across studied species, and more specifically between haplorrhines and strepsirrhines, suggesting a 'baseline' similarity between the adaptive immune systems of the two main clades of primates.

In a much earlier study, Sommer et al. (2002) had studied the diversity of two MHC class II loci (involved in the adaptive immune system) in three endemic Malagasy rodents, *Eliurus myoxinus*, *Macrotarsomys bastardi*, and *Hypogeomys antimena*, the first with a promiscuous mating system and the other two with a monogamous mating system. They tested the hypothesis that monogamous species may have lower diversity in MHC class II genes than promiscuous species. They found indeed that the two monogamous species, *H*. *antimena* and *M*. *bastardi*, hold significantly lower MHC class II variation (DQA: 2 and 3 alleles) than the promiscuous species *E*. *myoxinus*

(DQA: 11 alleles), suggesting a possible link between mating system and MHC class II genetic diversity.

Hunnicutt et al. (2020) have recently used genomic data of four *Microcebus* species (*M*. *griseorufus*, *M*. *ravelobensis*, *M*. *mittermeieri*, and *M*. *tavaratra*), two *Cheirogaleus* species (*C*. *sibreei* and *C*. *medius*), one *Mirza* species (*M*. *zaza*) and other primates to study the diversity of pheromone receptor class 1 family (V1R) across lemurs. The comparative genomic approach shows significantly higher diversity of V1R receptors in nocturnal compared to diurnal primates. Important differences in V1R diversity were identified even between diurnal and nocturnal lemurs, the latter showing higher variation within individuals. For instance, nocturnal lemurs such as *M*. *mittermeieri* and *M*. *murinus* had as many as 146 and 102 intact V1R genes, respectively, against the ~20 genes found in *Propithecus coquereli* or *Prolemur simus*.

7.8 Population genomic diversity of three co-occurring lemur species across northern Madagascar

Gabriele Maria Sgarlata[1], Jordi Salmona[1,2], Barbara Le Pors[1], Emmanuel Rasolondraibe[3], Tantely Ralantoharijaona[3], Ando Rakotonanahary[3], Jacquis Randriamaroson[3], Sophie Manzi[2], Solofonirina Rasoloharijaona[3], John Rigobert Zaonarivelo[4], Nicole Volasoa Andriaholinirina[3], Lounès Chikhi[1,2]

1. Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, 2780-156 Oeiras, Portugal;
2. Laboratoire Evolution & Diversité Biologique (EDB UMR 5174), CNRS, IRD, UPS, Université de Toulouse Midi-Pyrénées, Toulouse, France;
3. Département de Biologie Animale et Ecologie, Faculté des Sciences, Université de Mahajanga, Mahajanga, Madagascar;
4. Département des Sciences de la Nature et de l'Environnement, Université d'Antsiranana, 201 Antsiranana, Madagascar

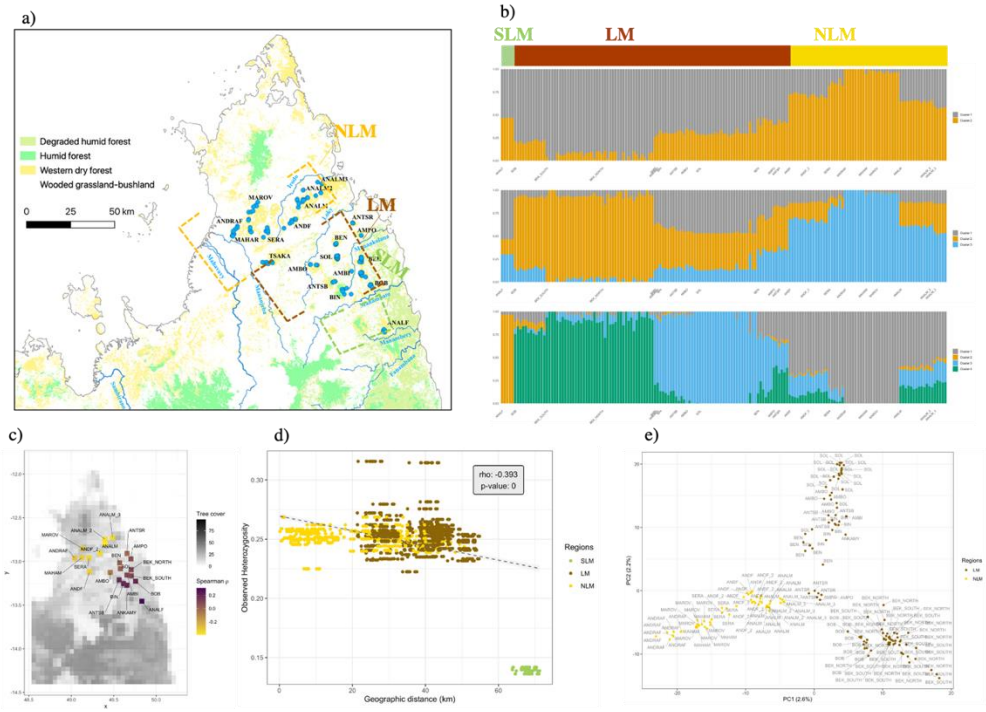


Figure 18. Population genomic diversity of the dry-associated mouse lemur (*Microcebus tavaratra*) in northern Madagascar. a) Sampling

distribution; b) ADMIXTURE results for K=2, 3 and 4; c) Squared cells indicate sampling sites, with colour code proportional to the amount of correlation (Spearman ρ) between site observed heterozygosity and geographical distance between the focal cell and the remaining sampling sites; d) Linear regression between geographic distance and individual observed heterozygosity from the site with highest correlation, namely ANALM in NLM region; e) PCA analysis.

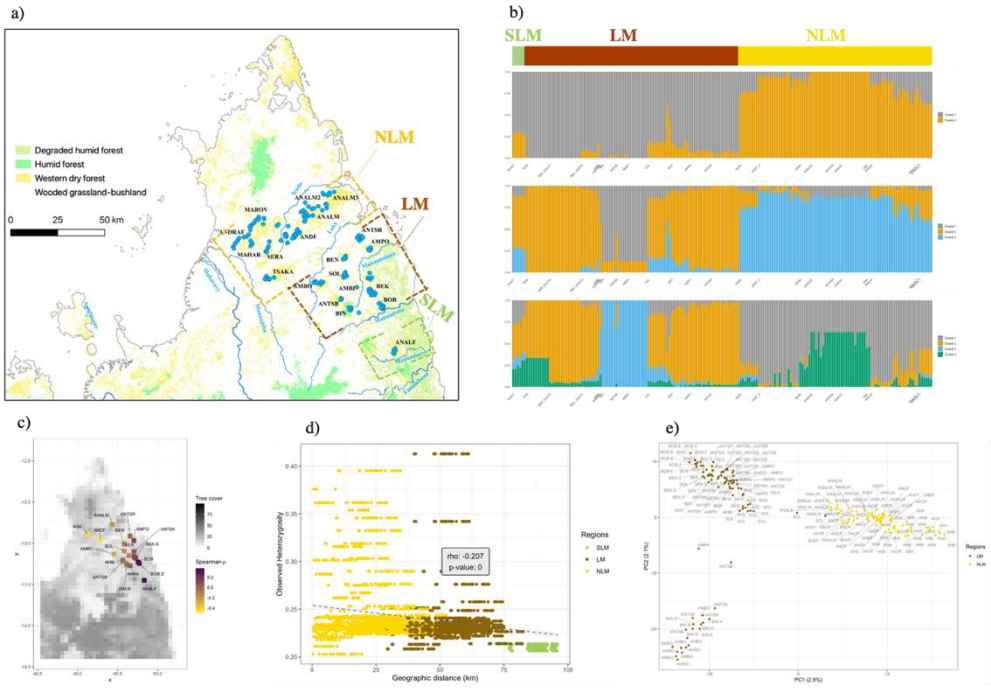


Figure 19. Population genomic diversity of the dry-associated sportive lemur (*Lepilemur cf. ankaranensis*) in northern Madagascar. a) Sampling distribution; b) ADMIXTURE results for K=2, 3 and 4; c) Squared cells indicate sampling sites, with colour code proportional to the amount of correlation (Spearman ρ) between site observed heterozygosity and geographical distance between the focal cell and the remaining sampling sites; d) Linear regression between geographic distance and individual observed heterozygosity from the site with highest correlation, namely ANK in NLM region; e) PCA analysis.

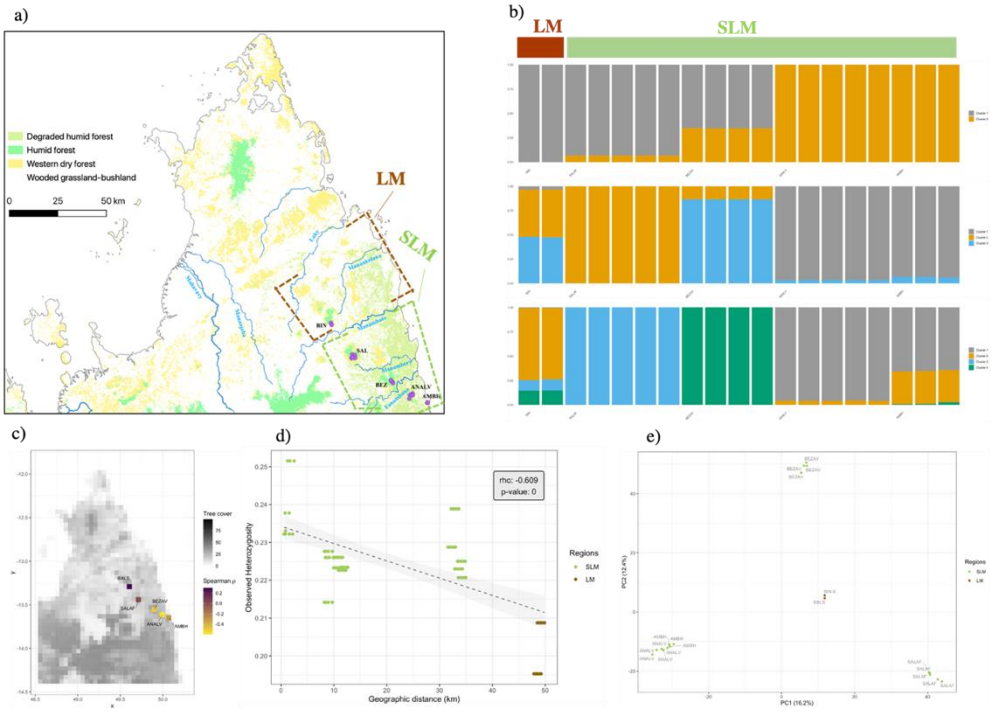


Figure 20. Population genomic diversity of the humid-associated mouse lemur (*Microcebus arnholdi*) in northern Madagascar. a) Sampling distribution; b) ADMIXTURE results for K=2, 3 and 4; c) Squared cells indicate sampling sites, with colour code proportional to the amount of correlation (Spearman ρ) between site observed heterozygosity and geographical distance between the focal cell and the remaining sampling sites; d) Linear regression between geographic distance and individual observed heterozygosity from the site with highest correlation, namely ANALV in SLM region; e) PCA analysis.





ITQB NOVA, Avenida da República, 2780-157 Oeiras, Portugal | Tel +351 214 469 100 | itqb.unl.pt