Márcia Cristina David Palma

Mestre em Microbiologia Aplicada

Elucidating the mating system of *Phaffia rhodozyma*, an astaxanthin-producing yeast with biotechnological potential

Dissertação para obtenção do Grau de Doutor em Biologia

Orientador: Doutora Paula Gonçalves, Profª. Auxiliar, FCT/UNL
Co-orientador: José Paulo Sampaio, Prof. Associado, FCT/UNL

Júri:

Presidente: Profª. Doutora Maria Adelaide de Almeida Pedro de Jesus
Arguentes: Profª. Doutora Ursula Kües
Doutora Sara Mayer Branco

Vogais: Profª. Doutora Paula Maria Theriaga Mendes Bernardo Gonçalves
Prof. Doutor Fernando José Santos Rodrigues

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Junho de 2017

Márcia Cristina David Palma

Elucidating the mating system of *Phaffia rhodozyma*, an astaxanthin-producing yeast with biotechnological potential

Acknowledgments

I want to express my gratitude to my thesis advisors, Paula Gonçalves and José Paulo Sampaio for their encouragement, guidance and trust along the years. To Diego Libking I express my gratitude for the many years of scientific collaboration and cooperation, and above all, for providing the crucial genomic information of *P. rhodozyma* that made possible my PhD project.

I also want to thank the members of the thesis committee, Fernando Rodrigues and Isabel Gordo, for their interest and suggestions throughout the development of my work. I would like to thank Gerhard Sandmann and Victor Cifuentes for providing plasmids that helped me get started and the assistance of Luís Jaime Mota for providing plasmids and strains for the two-hybrid systems used in this work.

I am indebted to all my lab colleagues (past and present), for their friendship and collaboration which provided for a great lab atmosphere. I would like to express my gratitude to Lia Godinho and Mário Ferreira for sharing with me their knowledge of molecular genetics, their friendship (and some of their restriction enzymes as well). I'm also thankful for all the help provided by our lab technician Nicole Soares.

I'm especially grateful to Carla Gonçalves that along the years has shared with me "the good, the bad and the ugly" of the scientific reality. She has been my traveling companion on the "scientific road" since the beginning and along the way became my treasured friend.

The support of my family has been essential throughout all my life. For the ones that are still beside me, I'm extremely grateful for their patience, love and support through the trying times that marked the years in which this work was developed. To the ones that have already transformed into stardust, my deepest gratitude and love. Your force will always be with me. Special thanks go to my fellow scientist, best friend and husband for never failing faith on me and for being the light in the darkest of days. Thank you.

Finally, I appreciate the financial support from Fundação para a Ciência e a Tecnologia that funded my PhD grant (SFRH/BD/81895/2011).

Resumo

Nos fungos basidiomicetes a identidade sexual é geralmente determinada por duas regiões genómicas designadas *locus PR* e *locus HD* (ou coletivamente denominadas de *loci MAT*) que habitualmente estão presentes em cromossomas distintos. O primeiro *locus* pode ser constituído por vários genes que codificam receptores de feromona (Ste3) e precursores de feromonas (Mfa) enquanto o segundo inclui, na sua versão mais simples, um par de genes, *HD1* e *HD2*, transcritos em direcções opostas, que codificam factores de transcrição com um domínio de ligação ao DNA. A maioria dos basidiomicetes apresentam um comportamento heterotálico, que requer a existência de compatibilidade entre ambos os *loci MAT*, sendo para tal necessário que cada indivíduo apresente diferentes alelos dos genes *MAT*. Existem, no entanto, fungos cujo comportamento sexual dispensa o cruzamento entre dois indivíduos geneticamente compatíveis, designados fungos homotálicos. Espécies homotálicas possuem a capacidade de completar o seu ciclo de vida em monocultura, dando origem a progenia meiótica. Enquanto os determinantes genéticos que regulam o comportamento heterotálico em basidiomicetes são já conhecidos em detalhe em várias espécies modelo, muito pouco se sabe sobre os genes e mecanismos moleculares que se encontram na base do homotalismo. O principal objectivo do trabalho apresentado nesta tese, foi preencher tanto quanto possível essa lacuna, estudando os genes e mecanismos moleculares que governam o ciclo de vida homotálico da levedura *Phaffia rhodozyma*. Foram encontrados seis genes *MAT* em *P. rhodozyma*, organizados em dois *loci* distintos, *PR* e *HD*, provavelmente localizados em cromossomas distintos. O *locus PR* encontrado apresenta dois conjuntos de genes, separados por aproximadamente 5 kb, onde cada conjunto é constituído por um gene *STE3* e um gene *MFA*. O locus *HD* é composto por um par de genes, *HD1* e *HD2*, transcritos em direcções opostas e apresentando o característico domínio de ligação ao DNA. A construção de mutantes com deleções de vários genes *MAT* permitiu propor um mecanismo molecular para o comportamento homotálico de *P. rhodozyma.* No modelo molecular proposto (i) a feromona Mfa1 activa o receptor Ste3-2 enquanto a feromona Mfa2 activa o receptor Ste3-1, existindo, portanto, redundância do sistema receptor/feromona enquanto (ii) os dois fatores de transcrição, Hd1 e Hd2, são ambos necessários para a ocorrência de reprodução sexuada. Procedeu-se igualmente à comparação das regiões *MAT* de *P. rhodozyma* com as regiões *MAT* de duas espécies novas entretanto incluídas no género *Phaffia*, (i.e. *P. novazelandica* e *P. tasmanica*) assim como com outras espécies pertencentes à ordem Cystofilobasidiales. Este trabalho permitiu pela primeira vez a elucidação dos mecanismos molecular básicos responsáveis pelo ciclo de vida homotálico de um basidiomiceta. Simultaneamente, a manipulação genética de *P. rhodozyma*, uma espécie com potencial biotecnológico para produção de astaxantina, permitiu gerar mutantes com determinantes sexuais "compatíveis" que poderão ser utilizados para melhoramento de estirpes industriais através de cruzamentos selectivos.

Palavras-chave: Basidiomicetes, Homotalismo, *Phaffia rhodozyma*, Genes *MAT*.

Abstract

In fungi belonging to the phylum Basidiomycota, sexual identity is usually determined by two genetically unlinked *MAT* loci, one named *PR* locus, which encodes one or more pheromone receptors (Ste3) and pheromone precursors (Mfa), and the other, named *HD* locus, that comprehends at least one pair of divergently transcribed genes encoding homeodomain transcription factors (*HD1* and *HD2*). The two *MAT* loci work as two distinct mating compatibility check points. Most basidiomycete species are heterothallic, meaning that sexual reproduction requires mating between two sexually compatible individuals harboring different alleles at both *MAT* loci. However, some species are known to be homothallic, one individual can complete the sexual cycle without mating with a genetically distinct partner. While the molecular underpinnings of the heterothallic life cycles of several basidiomycete model species have been dissected in detail, much less is known concerning the molecular basis for homothallism. The general aim of this research was to study the molecular mechanisms of sexual reproduction in fungi, specifically those governing the homothallic life cycle of *P. rhodozyma*.

Six *MAT* genes were found in *P. rhodozyma,* organized in two *MAT* loci, most likely located on different chromosomes. The *PR* locus was shown to be composed of two clusters, at approximately 5 kb from one another, each encoding one *STE3* gene and one *MFA* gene, while the *HD* locus encompassed two divergently transcribed homeodomain transcription factors genes, *HD1* and *HD2*. Functional genetic analysis was performed by targeted gene deletion of the *MAT* elements found in *P. rhodozyma* and the results allowed the proposal of a molecular model controlling the homothallic sexual behavior of *P. rhodozyma*. In this model (i) each pheromone interacts with the pheromone receptor of the other cluster (Mfa1 activates Ste3-2 while Mfa2 activates Ste3-1); since neither pheromone receptor is required *per se* for sporulation they seem to be functionally redundant; and (ii) both homeodomain proteins appear to work together to regulate genes required for sexual development. Comparison of the *MAT* regions of additional *Phaffia* species and of other representatives within the order Cystofilobasidiales, indicate that transitions to homothallism probably occurred several times independently. Furthermore, it revealed a particularly dynamic pattern of *MAT* gene evolution, with the generation of new receptors within each genus and exceptionally large numbers of mature pheromones encoded in the genomes of some of the species.

In conclusion, this work allowed for the first time the elucidation of the basic molecular mechanisms governing the homothallic life cycle of a basidiomycete. At the same time the genetic manipulation of the *MAT* genes of *P. rhodozyma* also allowed the generation of preferably outcrossing strains, which may be potentially useful for further improvement of this yeast as an industrial organism by way of selective breeding.

Keywords: Basidiomycetes, Homothallism, *Phaffia rhodozyma*, *MAT* genes.

Abbreviations

AIC	Akaike information criterion
bp	Base pair
D1/D2	Domains 1 and 2 of the 26S rDNA
DNA	Deoxyribonucleic acid
GDF	Gene deletion fragment
HD	Homeodomain
HMG	High-mobility group
HPLC-PAD	High-performance liquid chromatography with pulsed amperometric detector
ITS	Internal transcribed spacer
JGI	Joint Genome Institute
kb	Kilobase pair
LSU	Large subunit (ribosome)
MAP	Mitogen activated protein
MAPK	Mitogen activated protein kinase
MAT	Mating type
Mb	Megabase pair
min	Minute(s)
ML	Maximum likelihood
NCBI	National Center for Biotechnology Information
ORF	Open reading frame
rDNA	ribosomal DNA
s	Second(s)
PFGE	Pulsed-field gel electrophoresis
PCR	Polymerase chain reaction

Abbreviations of relevant culture collections

CBS	Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands
JCM	Japan Collection of Microorganisms, Riken Bioresource Center, Saitama, Japan
PYCC	Portuguese Yeast Culture Collection, CREM, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal

Abbreviated species names

C. albicans	*Candida albicans*
C. cinerea	*Coprinopsis cinerea*
C. deneoformans	*Cryptococcus deneoformans*
S. cerevisiae	*Saccharomyces cerevisiae*
S. commune	*Schizophyllum commune*
S. pombe	*Schizosaccharomyces pombe*
S. reilianum	*Sporisorium reilianum*
C. bisporidii	*Cystofilobasidium bisporidii*
C. ferigula	*Cystofilobasidium ferigula*
C. capitatum	*Cystofilobasidium capitatum*
C. macerans	*Cystofilobasidium macerans*
P. rhodozyma	*Phaffia rhodozyma*
P. novazelandica	*Phaffia novazelandica*
P. tasmanica	*Phaffia tasmanica*
K. huempii	*Krasilnikovozyma huempii*
M. blollopis	*Mrakia blollopis*
M. frigida	*Mrakia frigida*
M. aquatica	*Mrakia aquatica*
M. perniciosa	*Moniliophthora perniciosa*
T. pamirica	*Tausonia pamirica*
T. pullulans	*Tausonia pullulans*
U. megalosporus	*Udeniomyces megalosporus*
U. pyricola	*Udeniomyces pyricola*

Index

Chapter 1

General Introduction

1.1. Sexual reproduction in fungi 2

1.1.1. Sexual and asexual reproduction 2
1.1.2. Sexual identity in fungi 3
1.1.3. Mating-type (*MAT*) genes 3
1.1.3.1. *MAT* genes in Ascomycota 3
1.1.3.2. *MAT* genes in Basidiomycota 4
1.1.4. Mating behavior in fungi 6
1.1.4.1. Heterothallic mating 7
1.1.4.2. Homothallic mating 7
1.1.4.2.1. Primary homothallism 8
1.1.4.2.2. Secondary homothallism - Pseudohomothallism 9
1.1.4.2.3. Secondary homothallism - Mating-type switching 9
1.1.4.2.4. Unisexual mating 10
1.1.5. Homothallic basidiomycetes 11
1.1.5.1. *Moniliophthora permiciosa* 11
1.1.5.2. *Filobasidiela depauperata* 12
1.1.6. Molecular mechanisms enabling transitions between mating systems and mating behaviors 13

1.2. *Phaffia rhodozyma* 14

1.2.1. Discovery and taxonomic classification 14
1.2.2. Biogeography and ecological associations of *Phaffia rhodozyma* 15
1.2.3. Life cycle of *Phaffia rhodozyma* 16
1.2.4. Biotechnological relevance of *Phaffia rhodozyma* 18

1.3. Objectives and outline 19

Chapter 2

Identification of *MAT* and meiosis genes in *P. rhodozyma* genomes

2.1. Introduction **22**

2.2. Materials and methods **23**

2.2.1. Search for putative *MAT* and meiosis genes in *P. rhodozyma* genomes and their characterization.. 23
2.2.2. Comparison of *MAT* gene sequences between the three available *P. rhodozyma* genomes 24
2.2.3. Phylogenetic analysis of the putative *MAT* genes in *P. rhodozyma* strains representing distinct populations 24
2.2.4. Phylogenetic analysis of the pheromone receptor proteins of *P. rhodozyma* and other Tremellomycetes 25

2.3. Results **26**

2.3.1. *MAT* and mating related genes in *P. rhodozyma* 26
2.3.2. Meiosis genes in *P. rhodozyma* 28
2.3.3. *MAT* gene variability within *P. rhodozyma* 28
2.3.4. Phylogenies of *MAT* genes within *P. rhodozyma* 29
2.3.5. Phylogenetic relationship of pheromone receptors from *P. rhodozyma* with other basidiomycetes... 30

2.4. Discussion **32**

2.4.1. *MAT* genes are present in *P. rhodozyma* 32
2.4.2. Mating related genes and meiotic genes are present in *P. rhodozyma* 33

2.5. Conclusion **34**

Chapter 3

Genetic dissection of the homothallic sexual reproduction of *P. rhodozyma*

3.1. Introduction **36**

3.2. Materials and methods **37**

3.2.1. Strains and culture conditions 37
3.2.2. Construction of gene deletion fragments and complementation plasmids 37
3.2.3. Transformation of *P. rhodozyma* 38
3.2.4. Sporulation assays 38
3.2.5. Crosses between deletion mutant strains 39
3.2.6. Bacterial two-hybrid assays 39
3.2.7. Yeast two-hybrid assays 39

3.2.8. Search for potential alternative homeodomain transcription factor proteins 40
3.3. Results 41
3.3.1. Pheromone receptors are required for sexual reproduction 41
3.3.2. Evidence for recipocal compatibility between receptors and pheromones encoded by the two *PR* gene clusters 43
3.3.3. Genes *HD1* and *HD2* are required for sexual development in *P. rhodozyma* 43
3.3.4. Genetic evidence for the involvement of meiosis in the sexual cycle of *P. rhodozyma* 45
3.3.5. Construction of heterothallic strains 46
3.4. Discussion 46

Chapter 4

Two new species in the genus *Phaffia* - *Phaffia novazelandica* sp. nov. and *Phaffia tasmanica* sp. nov.

4.1. Introduction 52
4.2. Materials and methods 54
4.2.1. Phenotypical characterization 54
4.2.2. Genome sequencing 54
4.2.3. Genome assembly, prediction of coding regions and search for relevant genes 54
4.2.4. Phylogenetic analysis 56
4.3. Results 57
4.3.1. Morphological and physiological characterization of *P. novazelandica* and *P. tasmanica* 57
4.3.2. The divergent Australasian lineages belong to the genus *Phaffia* 60
4.3.3. Astaxanthin production by *P. tasmanica* and *P. novazelandica* 61
4.3.4. Organization and comparative analysis of *Phaffia MAT* loci 62
4.4. Discussion 64
4.4.1. Phylogenetic assignment of *P. tasmanica* and *P. novazelandica* in the genus *Phaffia* and its placement within the Cystofilobasidiales 64
4.4.2. Phenotypic and *MAT* genomic traits of *Phaffia* 64
4.4.3. Association of *P. tasmanica* and *P. novazelandica* with *Nothofagus* 65
4.4.4. Production of astaxanthin and the biotechnological potencial of *Phaffia* diversity 65

Chapter 5

MAT loci evolution in *Phaffia* and in other Cystofilobasidiales

5.1. Introduction 68

5.2. Materials and methods 69

5.2.1. Genome sequencing, identification of *MAT* genes and synteny analysis 69
5.2.2. *MAT* gene content and comparison across Cystofilobasidiales 70
5.2.3. Phylogenetic analyses 70

5.3. Results 71

5.3.1. Identification of *MAT* genes in Cystofilobasidiales 71
5.3.2. Pheromone receptors in Cystofilobasidiales 72
5.3.3. Pheromones in Cystofilobasidiales 74
5.3.4. Homeodomain transcription factors in homothallic and heterothallic Cystofilobasidiales 77
5.3.5. The *PR* locus in Cystofilobasidiales 79
5.3.6. The *HD* locus in Cystofilobasidiales 84

5.4. Discussion 87

5.4.1. *MAT* loci in *Phaffia* 87
5.4.2. *MAT* loci in *Mrakia* 88
5.4.3. *MAT* loci in *Cystofilobasidium* 90
5.4.4. *MAT* loci in other members of the Cystofilobasidiales 91

5.5. Conclusion 92

Chapter 6

Concluding remarks and future perspectives

6.1. Concluding remarks and future perspectives 96

References 102

Appendix I

Additional information pertaining to Chapter 2 114

Appendix II

Additional information pertaining to Chapter 3 122

Appendix III

Additional information pertaining to Chapter 4 136

III.1. Physiological characteristics of *P. novazelandica* and *P. tasmanica* 136
III.2. Additional information pertaining to the species phylogeny and to the *MAT* genes found in *P. novazelandica* and *P. tasmanica* 137
III.3. Additional information pertaining to the *MAT* and astaxanthin related genes found in *P. novazelandica* and *P. tasmanica* 138

Appendix IV

Additional information pertaining to Chapter 5 146

IV.1. Quality statistics for the draft genome assemblies 146
IV.2. Additional information pertaining to the *MAT* genes found in Cystofilobasidiales 146
IV.3. Taxonomic re-evaluation of strain Nwmf-AP1 146

Figure Index

Chapter 1

Figure 1.1. Summary of the modes of sexual reproduction in fungi 6

Figure 1.2. Distinc habitats associated with *P. rhodozyma* 15

Figure 1.3. Distribution, phylogeny and host tree associations of *P. rhodozyma* 16

Figure 1.4. Life cycle of *P. rhodozyma* 17

Figure 1.5. Sexual state of *P. rhodozyma* 18

Chapter 2

Figure 2.1. Gene content and organization around the putative mating type genes of *P. rhodozyma* 26

Figure 2.2. Features of Hd1 and Hd2 proteins of *P. rhodozyma* 27

Figure 2.3. Alignments of the predicted proteins for Hd1 and Hd2 of distinct *P. rhodozyma* strains 28

Figure 2.4. Phylogeny of *MAT* genes 29

Figure 2.5. Maximum likelihood phylogenetic tree of the pheromone receptors (Ste3) from Tremellomycetes 31

Chapter 3

Figure 3.1. *MAT* loci in *P. rhodozyma* 36

Figure 3.2. Deletion mutants in the *PR* locus 42

Figure 3.3. Deletion mutants in the *HD* locus 44

Figure 3.4. *SPO11* deletion mutant 45

Figure 3.5. Model describing the molecular interactions required for sexual reproduction 49

Chapter 4

Figure 4.1. ITS phylogeny inferred with maximum likelihood for *Phaffia* and *Cystofilobasidium* species 53

Figure 4.2. Detail of the Tremellomycetes phylogenetic tree based on a seven gene dataset 53

Figure 4.3. Australasia and the collection sites of *Phaffia* spp. isolates 57

Figure 4.4. Vegetative cells of *P. novazelandica* and *P. tasmanica* 58

Figure 4.5. Sexual structures of *P. novazelandica* and *P. tasmanica* 59

Figure 4.6. Phylogeny of the species of the order Cystofilobasidiales 60

Figure 4.7. *Phaffia* and astaxanthin production 61

Figure 4.8. Scaffolds encompassing *PR* and *HD* genes in the different *Phaffia* species 62

Figure 4.9. Phylogeny of *Phaffia* pheromone receptors 63

Figure 4.10. Pheromone precursor proteins from all *Phaffia* species 63

Chapter 5

Figure 5.1. Phylogeny of pheromone receptors of the Cystofilobasidiales 72

Figure 5.2. Protein sequence alignment of the pheromone precursors of members of the Cystofilobasidiales 75

Figure 5.3. Genomic regions of strains of *M. blollopis* depicting the *MFA* genes found and their similarity 76

Figure 5.4. Alignments of Hd1 and Hd2 predicted proteins 78

Figure 5.5. *PR* regions in members of the Cystofilobasidiales 81

Figure 5.6. Repeat regions present in the *PR* scaffolds of *M. blollopis* and *M. frigida* 83

Figure 5.7. *HD* regions in members of the Cystofilobasidiales 85

Appendix I

Figure I.1. *MAT* protein alignments 116

Figure I.2. Predicted pheromone receptor proteins Ste3-1 and Ste3-2 of strain CBS 7918^T 119

Figure I.3. Predicted Hd1 and Hd2 proteins of strain CBS 7918^T 119

Appendix II

Figure II.1. Plasmids used and constructed in this work 129

Figure II.2. Southern blot analysis of selected deletion mutants 130

Figure II.3. Synthetic coding sequences of the *HD1* and *HD2* genes 131

Figure II.4. Bacterial two-hybrid assays 132

Figure II.5. Yeast two-hybrid assays 133

Appendix III

Figure III.1. Alignment of the proteins from the biosynthetic route of astaxanthin of the three *Phaffia* species 138

Figure III.2. Predicted proteins of the pheromone receptor genes of *P. novazelandica* and *P. tasmanica* 143

Figure III.3. Predicted proteins of the homeodomain transcription factor genes of *P. novazelandica* and *P. tasmanica* 144

Appendix IV

Figure IV.1. Phylogenetic inferences illustrating the likelihood of strain Nwmf-AP1 belonging to *Mrakia blollopis* species instead of *Mrakia frigida* 147

Figure IV.2. Pheromone receptor phylogeny 148

Figure IV.3. *HD* regions of *M. blollopis* strains 148

Table Index

Chapter 2

Table 2.1. List of strains used in the phylogenetic study 24

Table 2.2. List of primers and PCR conditions used to amplify *MAT* genes in *P. rhodozyma* strains 25

Table 2.3. List of *MAT* genes found in *P. rhodozyma* genomes 26

Chapter 3

Table 3.1. List of *P. rhodozyma* strains used and corresponding genetically manipulated derivatives generated in this study 41

Chapter 4

Table 4.1. List of strains selected for whole genome sequencing 54

Table 4.2. List of draft genomes from species of Cystofilobasidiales 55

Table 4.3. List of strains used in phenotypic tests 58

Table 4.4. Maximum growth temperatures of *P. novazelandica* and *P. tasmanica* 58

Table 4.5. Location of genes of the biosynthetic route of astaxanthin in the draft genomes of *P. novazelandica* and *P. tasmanica* 61

Chapter 5

Table 5.1. List of all draft genome assemblies from distinct Cystofilobasidiales species used in this chapter 69

Table 5.2. List of *MAT* genes found in the analyzed Cystofilobasidiales genomes 71

Table 5.3. Percentage of amino acid identity between different Ste3 proteins in several species of Cystofilobasidiales 73

Table 5.4. *MFA* genes from Cystofilobasidiales, their pheromone precursor proteins and putative mature pheromones 74

Appendix I

Table I.1. Complete list of *MAT* genes and *MAT* related genes searched in the draft genome assembly of CBS 7918^T 114

Table I.2. Complete list of meiotic genes searched in the draft genome assembly of CBS 7918^T 114

Table I.3. Amino acid identity (%) between different alleles of pheromone receptors from selected basidiomycetes 114

Table I.4. Accession numbers and information for all sequences obtained and used in chapter 2 114

Appendix II

Table II.1. List of strains used in bacterial and yeast two-hybrid assays 122

Table II.2. Primers and plasmids used for the construction of all gene deletion fragments except for gene *STE3-2*..... 123

Table II.3. Primers and plasmid used for the construction of the *STE3-2* deletion fragment by overlap extension PCR 124

Table II.4. Primers and plasmid used for the construction of the zeocin resistance cassette 125

Table II.5. Primers and plasmid used for the construction of the complementation plasmids..... 125

Table II.6. Sporulation data pertaining to plot (e) from Figure 3.2, plot (c) from Figure 3.3 and plot (a) from Figure 3.4 126

Table II.7. Sporulation data pertaining to plot (f) from Figure 3.2..... 126

Table II.8. Data from crosses of double and triple mutant strains 126

Table II.9. List of plasmids used in the Bacterial Adenylate Cyclase Two-Hybrid System 127

Table II.10. List of primers used in the Bacterial Adenylate Cyclase Two-Hybrid System 127

Table II.11. List of plasmids used in the Matchmaker Gold Yeast Two-Hybrid System..... 128

Table II.12. List of primers used in the Matchmaker Gold Yeast Two-Hybrid System..... 128

Table II.13. List of putative homeodomain proteins retrieved from CBS 6938 genome..... 128

Appendix III

Table III.1. Physiological tests for three strains of each of the new *Phaffia* species 136

Table III.2. Statistics for the genomes assembled for this work (Chapter 4) 137

Appendix IV

Table IV.1. Statistics for the genomes assembled in this work (Chapter 5) 146

CHAPTER 1

1. General introduction

1.1. Sexual reproduction in fungi

The study of sexual reproduction in fungi has demonstrated that these organisms are an exceptional model for understanding the evolution of sex determination in eukaryotes (Fraser *et al.* 2004), including in metazoans with which fungi are evolutionarily closely related (Baldauf & Palmer 1993; Wainright *et al.* 1993; Idnurm 2011a). Of the several lineages that compose the fungal kingdom, the majority of the described fungal species are part of the Dikarya clade (James *et al.* 2006), composed of the phyla Ascomycota and Basidiomycota, which have been the object of most of the studies about sexual reproduction (Idnurm 2011a).

1.1.1. Sexual and asexual reproduction

Sexual reproduction is a pervasive trait in all the major lineages of the eukaryotic tree of life, and as such has been hypothesized that sex emerged once in an ancestral eukaryote and has since been maintained (Heitman *et al.* 2013). A remarkable aspect of this ubiquitous trait in eukaryotes is that although it presents conservation of some core features that are common to most sexually reproducing organisms, like ploidy changes, the occurrence of meiosis or cell-cell fusion between mating partners or gametes, it also displays great plasticity (Ramesh *et al.* 2005; Heitman 2015). This is particularly evident in the variety of distinct ways sexual identity may be determined or how sexual reproduction may be accomplished, depending on the organism (Fraser & Heitman 2004; Heitman *et al.* 2013). Sexual reproduction presents both costs and benefits when compared to asexual reproduction. Most sexual cycles depend on the availability of two compatible partners, demanding time and energy and ultimately allowing each parental only 50% chance of transmitting its genes to the offspring (Maynard-Smith 1978; Heitman *et al.* 2013). Additionally, sexual reproduction may break apart well adapted genomic configurations (Ni *et al.* 2011; Sun & Heitman 2011). Despite these costs, sexual reproduction allows (i) the generation of progeny with a diversity of novel genotypes, (ii) the purge from the genome of deleterious mutations (Muller 1964), (iii) transposable elements and (iv) enables the organism to keep ahead in the co-evolutionary race with its environmental challenges (Hamilton *et al.* 1990; Jokela *et al.* 2009; de Vienne *et al.* 2013; Heitman *et al.* 2013). Asexual reproduction in turn, allows a parental individual to transmit 100% of its genes to the offspring, is more energy efficient and preserves well adapted genomic configurations. However, it also leads to less diverse populations that are unable to rapidly adapt to environmental changes and are more prone to accumulate deleterious mutations over time (Sun & Heitman 2011). Most fungi are capable of reproducing both sexually and asexually, which indicates that the maintenance of these two modes of reproduction may provide selective advantages (Williams 1975; Heitman 2006; Aanen & Hoekstra 2007; Heitman *et al.* 2013).

1.1.2. Sexual identity in fungi

Sexual identity in fungi is determined in the haploid stage (Billiard *et al.* 2011) by one or two specialized genomic regions named mating-type loci (*MAT*), which in addition to defining the cell mating-type, are also responsible for orchestrating the sexual cycle (Herskowitz *et al.* 1992; Casselton 2002; Stanton & Hull 2007). The fungal *MAT* loci have been shown to vary greatly, both in length and gene content throughout both ascomycetes and basidiomycetes (Casselton 2002; Stanton & Hull 2007). Depending on whether the mating type-determining genes are genetically linked in one single *MAT* locus or genetically unlinked in two distinct *MAT* loci, the mating-system of a fungal species may be respectively, bipolar or tetrapolar (Casselton & Olesnicky 1998; Fraser & Heitman 2003). In species with a bipolar mating-system, cells of two distinct (and compatible) mating-types (e.g. *a* and *α*) exist in the population, where cell identity is governed by a single *MAT* locus that encodes one of two possible idiomorphs in Ascomycota (Butler 2007) or alleles in Basidiomycota (Fraser & Heitman 2003; Fraser *et al.* 2007). In contrast, species with a tetrapolar mating-system have two different and unlinked *MAT* loci defining the mating-type of each individual. This means that, in order to successfully mate, individuals must differ from one another at both *MAT* loci (Fraser & Heitman 2003; Fraser *et al.* 2007). If each of the two *MAT* loci harbor only two alleles (biallelic), four distinct mating-types will exist in the population. However, most tetrapolar species are multiallelic in at least one of the *MAT* loci, giving rise to a large number of different mating-types (Casselton & Olesnicky 1998; Fraser & Heitman 2003; Kües 2015). A more recently described intermediate system, named pseudo-bipolar, was found in basidiomycetes, which involves a configuration in which the *PR* and *HD* loci are partially linked enabling limited recombination (Coelho *et al.* 2010; Gioti *et al.* 2013a; Wu *et al.* 2015). In the Ascomycota the bipolar mating-system is the norm while in the Basidiomycota the prevailing system is the tetrapolar one, albeit also encompassing bipolar and pseudo-bipolar species (Fraser & Heitman 2003; Fraser *et al.* 2007; Morrow & Fraser 2009; Coelho *et al.* 2010). Given that tetrapolarity is absent in the ascomycetes and has been found in all the major lineages of the basidiomycetes (Ustilaginomycotina, Pucciniomycotina and Agaricomycotina), the tetrapolar system is considered to be the ancestral mating-system of this phylum (Fraser *et al.* 2007; Kües *et al.* 2011; Maia *et al.* 2015).

1.1.3. Mating-type (*MAT*) genes

1.1.3.1. *MAT* genes in Ascomycota

In most members of the Ascomycota, mating-type genes are located at a single *MAT* locus that may encode transcription factors belonging to three distinct families: α-domain, homeodomain (HD) or high-mobility-group (HMG) (Astell *et al.* 1981; Butler 2007; Bennett & Turgeon 2016). The HD transcription factors genes *MATα2* and *MATa1* encoded at the *MAT* loci belong to two distinct classes, respectively HD1 and HD2, based on structural homologies and distinct protein sequences of the DNA-binding motifs (Kües & Casselton 1992; Stanton & Hull 2007; Kües *et al.* 2011). Different lineages of the

ascomycetes display different sets of transcription factors present at their *MAT* loci (Butler 2007; Stanton & Hull 2007; Bennett & Turgeon 2016). Nevertheless, in general the *MATα* idiomorphs (also designated as *MAT1-1)* are defined by the consistent presence of a gene encoding a protein with a α-domain motif, whereas the *MATa* idiomorphs (also designated as *MAT1-2)* are usually characterized by the presence of a gene encoding a transcription factor with a HMG domain (Butler *et al.* 2004; Butler 2007; Bennett & Turgeon 2016). In some species, however, *MATα* may harbor more than one gene, like in *Saccharomyces cerevisiae* in which *MATα* encodes a protein with an α domain (*MATα1)* and another with a homeodomain (*MATα2*) (Herskowitz *et al.* 1992; Butler 2007). In ascomycetes, alternate versions of the *MAT* locus are named idiomorphs rather than alleles (Butler 2007) because, although they typically occupy the same chromosomal location in different haploid genomes, they are completely dissimilar and encode unrelated proteins. The transcription factors encoded at the *MAT* loci regulate an array of genes essential for mate recognition, meiosis and sexual development, like pheromones (*MFa* and *MFα*) and pheromone receptors (*STE2* and *STE3*) genes belonging to the G-protein-coupled receptor family, all of which are encoded outside of the *MAT* loci (Kronstad & Staben 1997; Chen & Thorner 2007; Tsong *et al.* 2007; Bennett & Turgeon 2016).

1.1.3.2. *MAT* genes in Basidiomycota

As previously mentioned, the most frequent mating-system in the Basidiomycota is the tetrapolar system, which means that sexual identity (mating-type) is generally determined by two genetically unlinked *MAT* loci (Casselton & Olesnicky 1998). One distinct feature of the tetrapolar mating-system and consequently of mating in the Basidiomycota is that the genes encoding pheromones and pheromone receptors have acquired a mating-type defining role, in opposition to what occurs in the Ascomycetes (Fraser *et al.* 2007). The *MAT* regions of Basidiomycetes that have been characterized thus far show an enormous level of diversity, nevertheless they all appear to derive from a common component set: a locus encoding pheromones and pheromone receptor genes, usually named *PR*, and a locus encoding homeodomain transcription factor genes referred to as *HD* (Fraser *et al.* 2007; Kües 2015).

The archetype of the *PR* locus encodes pheromone receptor genes homologous to the *STE3* receptor gene form *S. cerevisiae*, a G-protein-coupled receptor with seven transmembrane domains, and pheromone precursor genes, which are homologous to the lipid-modified *MFa* pheromone gene of *S. cerevisiae* (Bölker *et al.* 1992; Casselton & Olesnicky 1998; Raudaskoski & Kothe 2010; Kües 2015). Basidiomycete pheromones are first translated into pheromone precursor proteins which subsequently undergo posttranslational processing at both N and C-terminal regions. A general feature of pheromone precursor proteins is the presence of a “CaaX” motif at the carboxyl terminus of the precursor, where “C” is cysteine, “a” represents aliphatic amino acids and “X” stands for any residue (Raudaskoski & Kothe 2010). Some exceptions to this motif have been described in basidiomycetes, namely *Sporisorium reilianum* (Schirawski *et al.* 2005) and *Rhodotorula toruloides* (Coelho *et al.* 2008), where instead of two aliphatic amino acids, there is only one accompanied by a threonine residue. The multistep maturation process of the pheromone precursors involves farnesylation, C-

terminal proteolytic cleavage of the motif “aaX”, followed by carboxylmethylation of the cysteine residue and proteolytic cleavage at the N-terminus, originating mature lipopetide pheromones ranging from 9 to 15 amino acids (Caldwell *et al.* 1995; Kües *et al.* 2011; Kües 2015; Freihorst *et al.* 2016). The site for N-terminal cleavage is conserved in the pheromone precursor proteins of some species, like *Coprinopsis cinerea* were the protease recognition site is a pair of charged amino acids such as glutamate-arginine, or one single charged residue, as in some *Schizophyllum commune* precursors (Riquelme *et al.* 2005; Freihorst *et al.* 2016). Alignments of the predicted pheromone precursor proteins in some cases allow inference of the N-terminal cleavage site due to the levels of amino acid similarity on either side of the region that composes the mature pheromone (Fowler *et al.* 2004; Riquelme *et al.* 2005; Freihorst *et al.* 2016).

Regarding the *HD* locus, its most simple configurations contains two divergently transcribed genes encoding homeodomain transcription factors of different classes, HD1 and HD2 (Kües & Casselton 1992). The Hd1 homeodomain transcription factors are considered to have an atypical homeodomain, with three additional amino acids between helices one and two of the three helices comprising the homeodomain motif. Conversely, the Hd2 homeodomain transcription factors have a canonical homeodomain, with 60 residues forming a three helical structure and a highly conserved DNA-binding region (Kües & Casselton 1992; Kües *et al.* 2011; Freihorst *et al.* 2016). In order for these transcription factors to be functional in mating, it is necessary for them to heterodimerize with each other, however this can only occur after cell fusion, because dimerization is restricted to subunits (Hd1 and Hd2) encoded by different alleles of the *HD* loci, that originate from distinct mating partners (Banham *et al.* 1995; Kämper *et al.* 1995). It was shown that the discrimination between interactions of self and non-self-subunits is due to the N-terminal regions of both Hd1 and Hd2, which confer mating-type specificity and also mediate dimerization (Yee & Kronstad 1993; Kües *et al.* 1994; Kämper *et al.* 1995; Yue *et al.* 1997; Freihorst *et al.* 2016). Because the sequences from the N-terminal regions of different alleles of HD1 and also HD2 are usually highly divergent when compared to the homeodomain and C-terminal regions, it was hypothesized that these genes have evolved under a pattern of “divergence-homogenization duality” (Badrane & May 1999; May *et al.* 1999). This evolutionary pattern is characterized by the maintenance of divergence at the N-terminal region to ensure mating-type specificity and, at the same time, conservation in the rest of the protein to preserve the transcriptional regulatory motifs necessary to keep the homeodomain-regulated pathway functional (Badrane & May 1999; May *et al.* 1999). Although this evolutionary pattern is the most usually observed, an exception is known, namely the heterothallic basidiomycetous fungus *Phanerochaete chrysosporium* that shows hyperpolymorphism at the C-terminal instead of the N-terminal region of both Hd1 and Hd2 proteins (James *et al.* 2011). Another feature of HD proteins is the presence of a nuclear localization signal at the C-terminal end of either Hd1 or Hd2 of an interacting pair, since it is essential for the active heterodimer to reach the nucleus in order to activate the homeodomain-regulated pathway (Spit *et al.* 1998; Robertson *et al.* 2002).

1.1.4. Mating behavior in Fungi

Sexual reproduction in fungi may occur by two distinct mating behaviors: heterothallism or homothallism (Figure 1.1) (Blakeslee 1904). Heterothallic species (self-sterile) require fusion of two compatible partners of different mating types in order for successful mating to occur, while in homothallic species, each individual cell has the ability to undergo sexual reproduction by itself and produce sexual progeny (Blakeslee 1904; Lin & Heitman 2007; Casselton 2008; Wilson *et al.* 2015b). Both mating behaviors conserve the key features of sexual reproduction (ploidy changes, meiosis and production of recombinant progeny), but they differ in the fact that homothallic species do not require a compatible partner (Lin & Heitman 2007; Ni *et al.* 2011). Homothallism is considered to be common in the ascomycetes and relatively rare in basidiomycetes, where only about ≈10% of the accessed species (Whitehouse 1949) present this mating behavior (Lin & Heitman 2007; Kües *et al.* 2011). Many homothallic species are not hampered from mating with genetically distinct strains, therefore being able of both inbreeding and outcrossing (Lin & Heitman 2007; Kües *et al.* 2011). Additionally, some species have the ability to reproduce via both homothallic and heterothallic behavior, and are sometimes referred to as amphithallic (Ene & Bennett 2014; Roach *et al.* 2014). Finally, some species that are mostly heterothallic may present homothallic life cycles depending on the environmental conditions (Lin & Heitman 2007). Homothallism may in fact confer an adaptive advantage by removing the need to find a compatible partner (Billiard *et al.* 2011; Billiard *et al.* 2012) hence diminishing the cost associated with sexual reproduction (Roach *et al.* 2014).

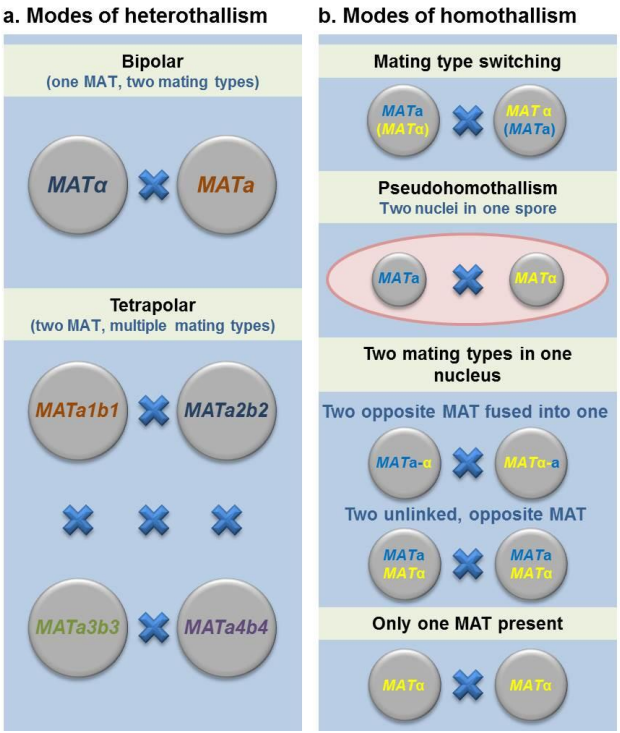


Figure 1.1. Summary of the modes of sexual reproduction in fungi. **(a)** Modes of heterothallism: (i) Bipolar: one *MAT* locus regulates sexual development. The two strains need to possess opposite *MAT* alleles/idiomorphs in order to successfully mate (ii) Tetrapolar: two *MAT* loci regulate sexual development and are often multiallelic.

Two strains need to possess opposite alleles at both *MAT* loci in order to successfully mate. **(b)** Modes of homothallism: (i) mating-type switching in which an ***α*** daughter cell mates with an ***a*** mother cell; (ii) pseudohomothallism in which two nuclei of opposite mating types are packaged into one spore; (iii) two *MAT* loci in one nucleus in which the two opposite *MAT* loci are either fused or unlinked in the genome; (iv) there is only one *MAT* idiomorph present and cells reproduce via unisexual mating. Adapted from Ni et al. (2011).

Compared to heterothallism, homothallism has been less studied and consequently less well understood within basidiomycetes. Therefore, most of the molecular determinants and mechanisms underpinning the different types of homothallism are most often illustrated in ascomycetes models (Lin & Heitman 2007; Roach *et al.* 2014; Wilson *et al.* 2015b).

1.1.4.1. Heterothallic mating

In basidiomycetes, mating between two homokaryotic hyphae (or two haploid yeast cells) usually originates a dikaryotic filamentous stage in which the two parental nuclei are replicated in a coordinated fashion. Fusion of the two nuclei (karyogamy) of distinct mating-types only takes place in specialized structures, either basidia or teliospores, which allow for the occurrence of meiosis and the formation of basidiospores which are the meiotic progeny (Hibbett *et al.* 2007). Molecularly, heterothallic mating between two compatible haploid partners, encoding different alleles of the *MAT* genes at each of their *MAT* loci, is initiated when mature pheromones are exported from the cell via a specific transporter, Ste6 (Hsueh & Shen 2005) and perceived by the other mating partner leading to cell fusion (Kües *et al.* 2011). Binding of the pheromone to its appropriate receptor causes conformational changes at the C-terminus of the receptor (Ste3) located at the plasma membrane, which in turn interacts with an heterotrimeric G protein (Hsueh *et al.* 2007; Li *et al.* 2007; Xue *et al.* 2008). The pheromone signal is subsequently transmitted via the pheromone-activated signal transduction cascade (Regenfelder *et al.* 1997; Davidson *et al.* 2003; Raudaskoski & Kothe 2010; Roach *et al.* 2014), which then is interconnected with the regulatory pathways activated by the heterodimer formed by the subunits Hd1/Hd2 encoded by the distinct mating-type alleles (Kües & Casselton 1992; Kües *et al.* 2011). This ultimately puts in place the transcriptional program that allows for the completion of the sexual cycle (Kües *et al.* 2011). Following comparable steps, heterothallic sexual reproduction in ascomycetes is orchestrated by the transcription factors encoded at their *MAT* loci that regulated the expression of genes, like those encoding pheromones and receptors. When sexually compatible cells interact successfully, cells fuse and karyogamy occurs forming a diploid cell, which in adequate environmental conditions undergoes meiosis generating haploid meiotic progeny (Stanton & Hull 2007).

1.1.4.2. Homothallic mating

By definition, homothallism is the ability of a single spore to reproduce sexually in monoculture completing the life cycle and giving rise to progeny (Blakeslee 1904). To this day, verification of this morphological trait allows the definition of the sexual state of a newly described species (Boekhout *et*

al. 2011). However, this classification does not distinguish the different molecular and cellular mechanisms employed by each organism to achieve the completion of the life cycle and the production of sexual offspring (Wilson *et al.* 2015b). The advent of molecular genetics followed by the exploration of the wealth of genomic data generated, namely by comparative genomics (Scazzocchio 2014), has unraveled profound differences between the molecular mechanisms underpinning homothallism (Lin & Heitman 2007; Wilson *et al.* 2015b). Different types of homothallism have been defined according to their morphological behavior and/or molecular bases: (i) primary homothallism, (ii) secondary homothallism encompassing pseudohomothallism and mating-type switching and finally (iii) unisexual reproduction (Lin & Heitman 2007; Wilson *et al.* 2015b; Bennett & Turgeon 2016).

1.1.4.2.1. Primary homothallism

Primary homothallism is characterized by the expression in the genome of a single individual of all the *MAT* alleles (or idiomorphs) required to trigger sexual development (Ni *et al.* 2011). In the case of an ascomycete, both *MATa* and *MATα* (or *MAT1*-1 and *MAT1*-2) idiomorphs would be present in a single genome and expressed, while in a basidiomycete this would require in principle the presence of a compatible pheromone and receptor pair, as well as an Hd1/Hd2 pair capable of forming an active heterodimer (Lin & Heitman 2007; Ni *et al.* 2011; Wilson *et al.* 2015b). The *MAT* genes may be in one of two alternative configurations: (i) both *MAT* loci are fused or closely linked as in *Cochliobolus* spp. or (ii) are unlinked and located in different regions of the genome as in *Aspergillus nidulans* (Lin & Heitman 2007; Paoletti *et al.* 2007; Ni *et al.* 2011). One classical example of primary homothallism in ascomycetes is the species *A. nidulans*, where each nucleus possesses both the *MAT1*-1 and *MAT1*-2 idiomorphs (Pontecorvo *et al.* 1953; Paoletti *et al.* 2007; Wang *et al.* 2010). This type of homothallism is considered rare among basidiomycetes and it was estimated that only 1% of species exhibit this type of breeding strategy (Lemke 1969; Koltin *et al.* 1972; Lin & Heitman 2007). One such example is *Sistotrema brinkmannii*, which consists in an aggregate of biological species displaying different patterns of sexuality, ranging from homothallism to bipolar heterothallism, and tetrapolar heterothallism (Lemke 1969). Within the *S. brinkmannii* species complex, isolates having homothallic and heterothallic bipolar mating behaviors can successfully hybridize and produce progeny, hence the homothallism and bipolar heterothallism observed were considered to be not fundamentally different (Lemke 1969; Ullrich 1973). It has been speculated that isolates exhibiting primary homothallism may arise through mutation or deletion of the entire bipolar incompatibility locus (Ullrich & Raper 1975). *Agaricus bisporus* is also a species containing different varieties with distinct mating behaviors, one of which is the *Agaricus bisporus* var. *eurotetrasporus* that displays primary homothallism, producing uninucleated haploid spores that are self-fertile (Callac *et al.* 2003). No molecular mechanisms were thus far elucidated for either primary homothallic species.

1.1.4.2.2. Secondary homothallism – Pseudohomothallism

Some fungi exhibit a particular system where post-meiotic nuclei of opposite mating types are packaged into one single spore that is then able to complete the sexual cycle on its own; hence the restitution of the diploid condition can always occur without the need to find a compatible partner (Whitehouse 1949; Billiard *et al.* 2011; Billiard *et al.* 2012). This behavior has been termed pseudohomothallism (Whitehouse 1949; Lin & Heitman 2007) and although morphologically it abides to the definition of homothallism, it actually corresponds to a heterothallic compatibility system in which automixis is strongly favored (Idnurm *et al.* 2015; Wilson *et al.* 2015b). The ascomycete *Neurospora tetrasperma* is one of the most studied species that exhibits this mating behavior. In this species, after sexual reproduction specialized structures called asci are formed, containing four ascospores, each of which possess two independent nuclei with opposite mating-types that allow for the completion of the sexual cycle (one *MAT1-2* and the other *MAT1-2*) (Dodge 1931; Raju 1992; Raju & Perkins 1994) (Wilson *et al.* 2015b). A similar example can be found in the basidiomycete species *Agaricus bisporus* var. *bisporus* (Kerrigan *et al.* 1993).

1.1.4.2.3. Secondary homothallism – Mating-type switching

Mating-type switching is characterized by a cell of a particular mating-type being able to undergo a molecular switch to the opposite mating-type (Lin & Heitman 2007). In this way a single cell can originate a colony with both compatible mating types and subsequently be able to reproduce sexually in a heterothallic manner (Lin & Heitman 2007; Wilson *et al.* 2015b). *Saccharomyces cerevisiae* has been the classical example of a species that is capable of homothallic sexual reproduction through mating-type switching. Two types of strains occur in *S. cerevisiae*, the strains Ho^{+} that can undergo mating-type switching and the strains Ho^{-} that are strictly limited to heterothallic behavior (Hicks & Herskowitz 1977; Hicks *et al.* 1977). This species has a bipolar mating-system, where mating-type is governed by one transcriptionally active *MAT* locus located in the middle of chromosome III (Rusche *et al.* 2003). In addition to the active *MAT* locus, two other mating-type-like cassettes exist, *HML* and *HMR*, which are not being expressed and encode two compatible *MAT* idiomorphs. During mitosis, an endonuclease encoded by the *HO* gene, initiates a gene conversion process that allows the replacement of the existing idiomorph at the active *MAT* locus by one of the two silent mating-type-like (*HML* and *HMR*) cassettes. Thus, one of the cells originated by that mitosis switches mating type. The above mentioned strains, Ho^{-}, possess an inactive copy of this unique endonuclease and therefore are self-sterile and only capable of heterothallic mating thus requiring a compatible partner (Mccusker 2006; Roach *et al.* 2014). Other ascomycetes like *Schizosaccharomyces pombe* display a mating-type switching mechanism that although different from the one in *S. cerevisiae*, presents general features in common, one of which is the presence of three mating-type-like loci (including two silent loci) (Butler 2007). However, more recently, alternative switching mechanisms were discovered for two other ascomycetes, *Hansenula polymorpha* and *Pichia pastoris* (Bennett & Turgeon 2016). In both species, there are two linked mating-type-like loci (*MATa* and *MATα*) but only one is actively expressed,

whereas the other is silenced or expressed only at a reduced level. The mechanism of mating-type switching occurs by recombination between inverted-repeat sequences flanking the *MAT* loci which results in expression of the opposite mating type (Hanson *et al.* 2014; Maekawa & Kaneko 2014). Because the mechanisms of mating-type switching mentioned in the previous examples are reversible processes they are called bidirectional, however there is also another form of mating-type switching that is irreversible and hence termed unidirectional (Wilson *et al.* 2015b). The self-fertile isolates of the ascomycetes *Ceratocystidaceae fimbriata and C. albifundus* encode at their *MAT* locus the genes encoding both compatible idiomorphs (*MAT1-1* and *MAT1-2*). Ascospores produced during sexual reproduction segregate into self-fertile and self-sterile isolates. Genetically, the switch from a self-fertile to a self-sterile is the result of the complete deletion of the genetic information of *MAT1-2* idiomorph and is, therefore, irreversible (Wilken *et al.* 2014; Wilson *et al.* 2015b). In the basidiomycetes there is only one species *Agrocybe aegerita* reported to undergo mating-type switching, although the molecular mechanisms remains to be explored (Labarere & Noel 1992; Nieuwenhuis & Immler 2016).

1.1.4.2.4. Unisexual mating

Unisexual mating is characterized by the ability of one cell, containing the genetic information of just one mating-type, to reproduce sexually and originate offspring (Lin *et al.* 2005; Roach *et al.* 2014). This is made possible either by endoreplication of the genome in the cell or by the fusion of two cells carrying haploid nuclei with the same mating-type allele or idiomorph. This remarkable mating behavior has been found in ascomycetes like *Neurospora africana* (Glass & Smith 1994), *Candida albicans* (Alby *et al.* 2009), *Huntiella moniliformis* (Wilson *et al.* 2015a) and also in basidiomycetes, namely *Cryptococcus deneoformans* (Lin *et al.* 2005). The fact that heterothallic reproducing species like *C. deneoformans* and *C. albicans* also show alternative homothallic sexual cycles appear to suggest (Alby *et al.* 2009; Feretzaki & Heitman 2013a) that both sexual behaviors are maintained as a way of tapping into the benefits associated with both. Detailed description of the molecular mechanisms underpinning unisexual reproduction have been undertaken for both ascomycetes and basidiomycetes, namely *Candida albicans* (Alby *et al.* 2009) and *C. deneoformans* (Lin *et al.* 2005; Feretzaki & Heitman 2013a; Ni *et al.* 2013), since this type of sexual reproduction is possibly linked to the pathogenicity of these species and thus has implications for human health (Roach *et al.* 2014; Wilson *et al.* 2015b). Unisexual mating in *C. deneoformans* occurs when nutrient depletion induces cells of one mating-type to transition from yeast to hyphae, which will subsequently grow and form basidia at its tips; after the occurrence of endoreplication, meiosis produces basidiospores that are them multiplied by several round of mitosis (Lin *et al.* 2005). Diploidization may also occur earlier during cell-cell fusion between cells of the same mating type, either being clones or genetically different individuals (Lin *et al.* 2005; Feretzaki & Heitman 2013b).

1.1.5. Homothallic basidiomycetes

Few strictly homothallic basidiomycete species have been thoroughly studied regarding the molecular determinants responsible for their homothallic life cycle. The homothallic species for which the most is known regarding their homothallic behavior and molecular determinants are *Moniliophthora perniciosa* (Griffith & Hedger 1994a) and *Filobasidiella depauperata* (Kwon-Chung 1975).

1.1.5.1. *Moniliophthora perniciosa*

Moniliophthora perniciosa is a species complex composed by three distinct biotypes (C, S and L) defined by the plant they are associated with. The C-biotype infests mainly the cacao plant, the S-biotype is associated with solanaceous hosts and the L-biotype with bignoniaceous lianas (Meinhardt *et al.* 2008). This species complex encompasses distinct mating behaviors: while the C-biotype and S-biotype are primarily homothallic, the L-biotype has a heterothallic tetrapolar mating-system with multiple alleles for both the *HD* and *PR* loci (Griffith & Hedger 1994a, b; Kues & Navarro-Gonzalez 2010). The basidiomycete *M. perniciosa* belongs to the class Agaricomycetes that generally possesses *MAT* loci with a complex structure (Casselton & Kues 2007; Kües 2015). The *HD* loci encode two classes of homeodomain transcription factors, *HD1* and *HD2*, which may be present in one or more pairs of divergently transcribed genes (Kües 2015). The *PR* loci encode pheromone precursor genes (*MFA*) and pheromone receptor genes (*STE3*) that may be arranged in groups of paralogous genes within one locus, where some of the *STE3* and *MFA* genes may be non-mating type specific (Kües 2015). As in most basidiomycetes, for successful mating to occur between two strains it is necessary that they differ in their mating-type specificities that are defined by the alleles present at both *MAT* loci (Casselton & Kues 2007). In *M. perniciosa* C-biotype, uninucleated basidiospores germinate into mycelia with clamp cells within a few days upon germination (Griffith & Hedger 1994a). The partial genome of *M. perniciosa* C-biotype was sequenced and it was estimated that the genomic information obtained represents about 69% of the total haploid genome of the organism. Inspection of the genome regarding potential *MAT* genes revealed two possible *HD1* genes, eight possible *STE3* genes and as many as 5 putative *MFA* genes (Mondego *et al.* 2008; Kues & Navarro-Gonzalez 2010). Phylogenetic analyses using partial sequences of the identified STE3 genes have shown that they group together with receptors from other Agaricales, including those from the close relative *Moniliophthora roreri*, which also infests the cacao plant and displays a heterothallic tetrapolar *MAT* loci organization (Kues & Navarro-Gonzalez 2010; Kües *et al.* 2011; Diaz-Valderrama & Aime 2016). Additionally, comparison of the predicted pheromone precursor proteins, for all the *MFA* genes found, revealed two variants (Kues & Navarro-Gonzalez 2010; Kües 2015). Although the genomic information from *M. perniciosa* C-biotype is incomplete, these studies (Mondego *et al.* 2008; Kues & Navarro-Gonzalez 2010) showed the presence of *MAT* genes in a primary homothallic basidiomycete suggesting that constitutive self-compatibility at both the *PR* and *HD* loci, may be a possible mechanism by which the primary homothallism of *M. perniciosa* C-biotype arose (Kües *et al.* 2011). However, the role of the *MAT* genes found in *M. perniciosa* C-biotype is yet to be ascertained.

1.1.5.2. *Filobasidiella depauperata*

Filobasidiella depauperata is a homothallic, non-pathogenic, basidiomycete species that grows exclusively as hyphae and is found in association with decaying insects (Rodriguez-Carres *et al.* 2010). Spores of *F. depauperata* are uninucleated and each germinated spore establishes a monokaryotic mycelium without clamp connections that produces basidia and basidiospores (Kwon-Chung 1975; Rodriguez-Carres *et al.* 2010). This species belongs to the class Tremellomycetes and is closely related to the human pathogens *Cryptococcus deneoformans* and *Cryptococcus gattii* (Findley *et al.* 2009). While most basidiomycetes have a heterothallic tetrapolar mating-system, *C. deneoformans* has a heterothallic bipolar mating-system in which a single *MAT* locus (*MATa* or *MATα*) contains the receptor and the pheromone precursor genes linked to one of the genes encoding the homeodomain proteins: *MATα* encodes the *HD1* (*SXI1*) gene and *MATa* encodes the *HD2* (*SXI2*) gene (Hull & Heitman 2002; Hull *et al.* 2005). This is also in contrast with most basidiomycetes that usually encode a pair of divergently transcribed *HD1* and HD2 genes in the *HD* locus that characterizes each of the mating-types (Kües *et al.* 2011). Alongside the *MAT* genes, the *MAT* locus of *C. deneoformans* encode some 25 genes, spanning more than 100 kb, which corresponds to about 6% of the chromosome where it is located (Lengeler *et al.* 2002; Fraser *et al.* 2004; Fraser & Heitman 2005; Fraser *et al.* 2007). As previously mentioned, *C. deneoformans* is also capable of unisexual mating that allows an isolate of a specific mating type (either ***α*** or ***a***) to complete the life cycle and produce meiotic progeny with genetic diversity in the absence of a compatible partner (Lin *et al.* 2005; Ni *et al.* 2013). With the exception of the *SXI1/SXI2* genes which are dispensable for unisexual mating, no other molecular components (e.g. like the pheromone response cascade) appear to differ between the heterothallic and the homothallic unisexual mating observed in *C. deneoformans* (Roach *et al.* 2014). The filamentation in *F. depauperata* was suggested to resemble the unisexual mating displayed by *C. deneoformans*, i.e. the hyphal growth of *F. depauperata* appears to correspond to the sexual stage of the yeast *C. deneoformans* (Rodriguez-Carres *et al.* 2010). *Filobasidiella depauperata* appears to be an obligate sexual organism leading the authors to hypothesize that this organism is constantly engaged in sexual reproduction because its slow growth adds to its ability to resist in the presence of antimicrobial compounds (Rodriguez-Carres *et al.* 2010). Comparison of the partial genomic information available for *F. depauperata* with the genome of the well-studied *C. deneoformans* revealed that synteny is conserved between the two species except for the region that composes the *MAT* locus of *C. deneoformans* (Rodriguez-Carres *et al.* 2010). No mating-type genes (*STE3*, *MFA*, *HD1* or *HD2*) were, to date, found in *F. depauperata* and the molecular mechanism underpinning its homothallic sexual cycle has, thus far, not been elucidated (Rodriguez-Carres *et al.* 2010; Kües *et al.* 2011; Roach *et al.* 2014). It has been postulated that the homothallism seen in *F. depauperata* may have evolved from an ancestral heterothallic bipolar system. Although it is not clear how, it possibly involved extensive chromosomal rearrangements as suggested by the comparison with the *MAT* loci of *C. deneoformans* (Fraser & Heitman 2004; Rodriguez-Carres *et al.* 2010; Roach

et al. 2014). It is likely that sequencing of the complete genome of *F. depauperata* will provide further insights into its sexual behavior.

1.1.6. Molecular mechanisms enabling transitions between mating systems and mating behaviors

Transitions between heterothallic and homothallic mating behaviors are common in fungi (Lin & Heitman 2007; Ni *et al.* 2011; Roach *et al.* 2014), and frequently both modes of sexual reproduction can be observed in different species of the same genus (e.g. *Neurospora*) (Wik *et al.* 2008), or even within distinct strains of the same species (e.g. *Cystofilobasidium macerans*) (Libkind *et al.* 2009). Additionally, some heterothallic fungi display the ability to undergo homothallic mating under specific environmental conditions (Lin & Heitman 2007; Ni *et al.* 2011). These transitions between heterothallism and homothallism have been viewed as a choice between outcrossing and inbreeding and depending on the environmental conditions one may be favored over the other (Ni *et al.* 2011). It is still a matter of debate whether homothallism evolved from heterothallism or if it was the other way around, as there is evidence to support both hypotheses (Ni *et al.* 2011). Some authors, studying the ascomycete genus *Neurospora* postulated that the transitions from heterothallism to homothallism occur in a single direction, suggesting that homothallism is an evolutionary dead end (Gioti *et al.* 2012). Contrastingly, others highlight the great plasticity of sexual determination and modes of reproduction in fungi and hypothesize that these transitions have occurred frequently throughout the evolution of fungal lineages and in both directions (Raper 1966; Yun *et al.* 1999; Galagan *et al.* 2005; Lin & Heitman 2007; Idnurm 2011b). In fact, evidence was found for transitions from a homothallic ancestor giving rise to a heterothallic species, namely the analysis of the genomes of *Sclerotinia sclerotiorum* (homothallic) and *Botrytis cinerea* (heterothallic) where the comparison of the *MAT* loci in both species, showed that the configuration of the *B. cinerea MAT* loci can be explained by the occurrence of two simple inversion and deletion events from a *S. sclerotiorum*-like homothallic ancestor (Amselem *et al.* 2011).

Both recombination and mutation have been indicated as possible molecular mechanisms involved in transitions from heterothallism to homothallism and also from heterothallic tetrapolar mating systems to heterothallic bipolar mating systems (Fraser *et al.* 2007; Lin & Heitman 2007; Idnurm *et al.* 2015).

Mating-type loci in fungi are genomic regions that usually present extensive rearrangements and for which recombination is largely suppressed (Fontanillas *et al.* 2015; Idnurm *et al.* 2015). Nevertheless, it has been shown that although highly repressed, interchromosomal crossovers between *MAT* regions may still occur, not only in their flaking regions but also, within the *MAT* loci itself by way of gene conversion as observed in *C. deneoformans* or *Microbotryum lychnidis-dioicae* (Sun *et al.* 2012; Fontanillas *et al.* 2015; Idnurm *et al.* 2015; Sun & Heitman 2016). These non-canonical recombination events may be potentiated by particular characteristics within the *MAT* loci, like the presence of repeats that could facilitate non-allelic homologous recombination or the presence of transposable elements (Idnurm 2011b; Sun & Heitman 2016).

As previously mentioned, in most basidiomycetes the *PR* locus and the *HD* locus are in different chromosomes, constituting a tetrapolar mating system, however, in some basidiomycetes (e.g. *C. deneoformans* or *Ustilago hordei*), these two loci have become fused giving rise to the bipolar mating system (Fraser *et al.* 2007; Idnurm *et al.* 2015). Among the possible mechanisms that lead to that genomic configuration is ectopic recombination or non-homologous end joining (Sun & Heitman 2016). These same mechanisms have been proposed for the appearance of homothallic species from heterothallic ones (Fraser *et al.* 2007; Lin & Heitman 2007; Sun & Heitman 2016). Non-homologous recombination events bringing together compatible mating type alleles in the same haploid genome can lead to the formation of a self-fertile organism (Idnurm 2011b). Additionally to recombination, mutation is also a molecular mechanism that can be at the origin of transitions between mating systems as well as mating behaviors (Fraser *et al.* 2007). Mutations at the *PR* locus giving rise to pheromone-receptor self-activation, or constitutive pheromone receptor activation are known for some basidiomycetes, like *Schizophyllum commune* and *Coprinopsis cinerea* (Olesnicky *et al.* 2000; Fowler *et al.* 2001; Kües *et al.* 2011). Similarly, mutations at the *HD* locus may lead to self-compatible *HD1/HD2* allele pairs (Kües *et al.* 1994; Kämper *et al.* 1995). It has been observed that homothallism evolved independently several times, indicating that many different evolutionary events may lead to the different types of genetic arrangement that translate into an homothallic life cycle (Reeve 2014; Roach *et al.* 2014).

1.2. *Phaffia rhodozyma*

1.2.1. Discovery and taxonomic classification

Phaffia rhodozyma is an orange pigmented yeast isolated for the first time in 1967 from exudates of deciduous trees in Japan and Alaska (Phaff *et al.* 1972; Miller *et al.* 1976). This basidiomycetous yeast belongs to the order Cystofilobasidiales, which represents a basally diverging lineage of the class Tremellomycetes, in the subphylum Agaricomycotina (Kurtzman *et al.* 2011a; Liu *et al.* 2015a; Liu *et al.* 2015b; Sharma *et al.* 2015). Recent phylogenetic studies, using partial sequence data from seven genes (three rDNA genes/regions, three housekeeping genes and a mitochondrial gene), demonstrated that species belonging to the Cystofilobasidiales order cluster into seven well-defined clades that represent the seven genera encompassed by the order (Liu *et al.* 2015a; Liu *et al.* 2015b). This is in agreement with previous phylogenetic studies using only rDNA regions (Scorzetti *et al.* 2002; Boekhout *et al.* 2011; Kurtzman *et al.* 2011a). However, while the most recent phylogenetic studies indicated that *Phaffia* and *Krasilnikovozyma* are sister genera (Liu *et al.* 2015a; Liu *et al.* 2015b), previous studies (Scorzetti *et al.* 2002; Boekhout *et al.* 2011; Kurtzman *et al.* 2011a) placed *Phaffia* as a sister group of the genus *Cystofilobasidium.* Although the studies of Liu et al. (2015a,b) used a larger dataset for phylogenetic inference, the relationships established between the seven genera within Cystofilobasidiales are poorly supported (Liu *et al.* 2015a; Liu *et al.* 2015b). To date, the *Phaffia* genus is composed by one single species, *Phaffia rhodozyma* (Liu *et al.* 2015a), although the discovery of two potential new species was recently reported (David-Palma *et al.* 2014).

1.2.2. Biogeography and ecological associations of *Phaffia rhodozyma*

Since it was first isolated in 1967 by Herman Phaff (Phaff *et al.* 1972), several other isolates were obtained from exudates (also known as spring saps) of deciduous trees, belonging to genera *Betula*, *Alnus*, *Fagus* and *Cornus* throughout the Northern Hemisphere, e.g. Germany, Finland and Russia (Golubev *et al.* 1977; Weber *et al.* 2006). Due to the geographical distribution of the trees (Chen *et al.* 1999; Chen & Li 2004; Manchester *et al.* 2007) from which *P. rhodozyma* isolates were obtained, it was assumed that the species was only circumscribed to the Northern Hemisphere (Golubev 1995; Fell & Blatt 1999). However, in the beginning of the 21th century *P. rhodozyma* strains were isolated in the Southern Hemisphere (South America, Australia and New Zealand) from leafs of *Nothofagus* trees (David-Palma *et al.* 2014), as well as in association with the fruiting bodies of *Cyttaria* (Libkind *et al.* 2007; Libkind *et al.* 2011b), a biotrophic ascomycete restricted to *Nothofagus* (Peterson *et al.* 2010). The mycelium of *Cyttaria* develops inside the tree forming in its branches persistent tumors, from which fruiting bodies rich in simple sugars are annually produced, and from which isolates of *P. rhodozyma* are consistently isolated (Libkind *et al.* 2007; Libkind *et al.* 2008; Libkind *et al.* 2011b; David-Palma *et al.* 2014).

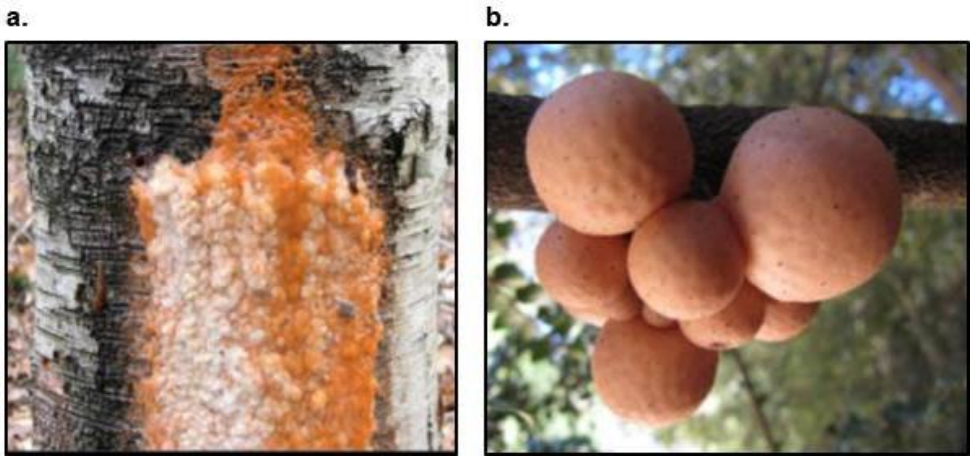


Figure 1.2. Distinct habitats associated with *P. rhodozyma*. (a) Exudate from a *Betula* tree trunk, Canada. (Photo courtesy of Beverly White); **(b)** branch of *Nothofagus cunninghamii* with *Cyttaria gunnii* fruiting bodies, Tasmania. (Photo courtesy of Diego Libkind**)**

A phylogenetic study using a multilocus dataset of seven partial gene sequences, was performed with strains encompassing isolates from all known habitats and distinct geographic locations from which *P. rhodozyma* has been isolated (David-Palma *et al.* 2014). The inferred phylogeny (Figure 1.3.) demonstrated that the species was divided into four main populations (A, B, C and D) that displayed a strong correlation with the tree genera colonized by each group of isolates, rather than with geography (David-Palma *et al.* 2014). The existence of two additional populations (Figure 3.1; populations E and F) that were considered to represent putative new species (David-Palma *et al.* 2014) was also shown.

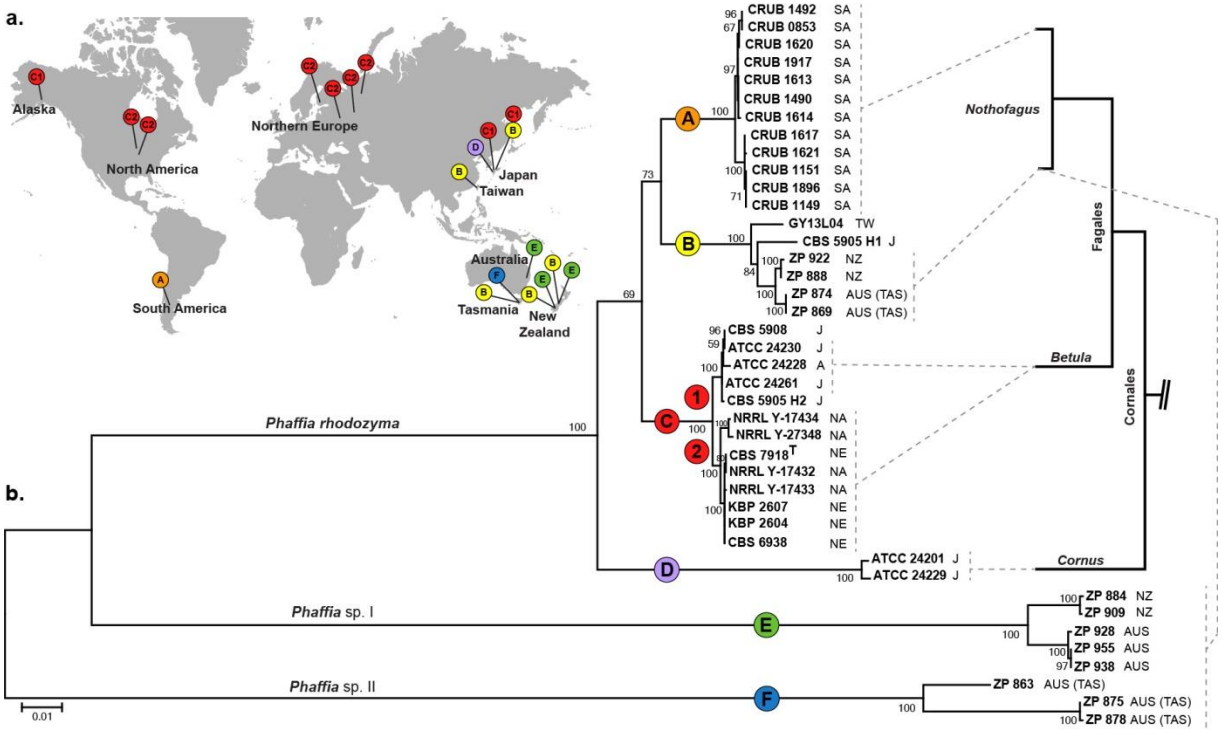


Figure 1.3. Distribution, phylogeny and host tree association of *P. rhodozyma*. (a) World map showing the distribution of populations defined by the phylogenetic inference, which are represented by capital letters and distinct colors **(b)** Unrooted maximum likelihood phylogenies of *Phaffia* (left) based on a concatenated alignment of partial sequences of seven genes, and of the host trees (right) based on internal transcribed spacer sequences. Bootstrap values (1000 replicates) higher than 50% are indicated. The geographical origin of the isolates is indicated after strain number (A, Alaska; AUS, Australia; J, Japan; NA, North America; NE, Northern Europe; NZ, New Zealand; SA, South America; TAS, Tasmania). (Adapted from David-Palma et al., 2014).

1.2.3. Life cycle of *Phaffia rhodozyma*

The sexual state of *P. rhodozyma* was observed for the first time in 1995 (Golubev 1995) after strains were inoculated in agar medium containing polyols as the sole carbon source. Under these conditions each strain, on its own, was capable of developing sexual structures. The sexual state (formerly named *Xanthophyllomyces dendrorhous*) is characterized by exhibiting three types of structures (Figure 1.4.) capable of originating the basidium in *P. rhodozyma*, the most common being conjugation between a mother-cell and its bud, also known as pedogamy (Golubev 1995). However, basidia were also observed to originate from unconjugated cells where possibly endoreplication precedes meiosis like observed in other fungi (Lin *et al.* 2009), as well as, from independent cells that conjugate prior to sexual development (Figure 1.4. and 1.5) (Golubev *et al.* 1977; Kucsera *et al.* 1998; Slaninova *et al.* 1999; Kucsera *et al.* 2000). Basidiospores develop on the apex of the slender aerial basidium on minute pegs (Figure 1.5.d), are sessile and germinate by budding, being able to undergo the complete life cycle (Golubev 1995; Kucsera *et al.* 1998; Kucsera *et al.* 2000). Several studies have provided evidence that the ploidy of the vegetative stage may vary depending on the strain, some being diploid, others haploid, while the existence of possible aneuploids has also been suggested (Wery *et al.* 1997; Kucsera *et al.* 1998; Medwid 1998; Kucsera *et al.* 2000; Hermosilla *et al.* 2003). Mating experiments between genetically marked strains followed by pulse-field gel electrophoresis of the chromosomal

DNA of individual spores resulting from the crosses, revealed evidence of karyogamy, meiosis and recombination. However, segregation of the markers was not as expected from a normal meiosis, leading the authors to suggest the possible existence of aneuploids among the studied spores (Kucsera *et al.* 1998). The sexual stage of *P. rhodozyma* (Figure 1.5) represents an unusual departure from what is commonly observed in the sexual stages of most basidiomycetes, because it does not involve a transition between a unicellular/yeast stage to a filamentous/hyphal one (Fraser *et al.* 2007).

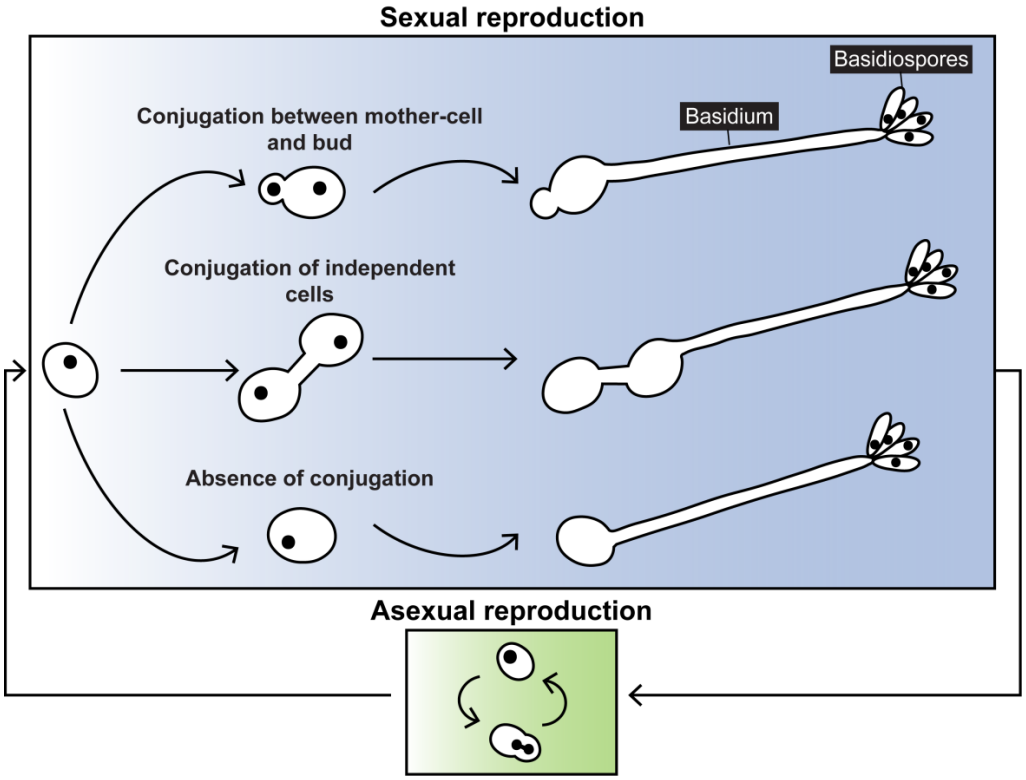


Figure 1.4. Life cycle of *P. rhodozyma*. This species can reproduce either sexually (blue box) or asexually (green box). Vegetative cells propagate by budding. Nitrogen depletion and the presence of polyols trigger the formation of an aerial basidium that gives rise to apical basidiospores. Three distinct cellular events were observed to give rise to formation of the basidium: conjugation of independent cells, conjugation between mother-cell and bud (pedogamy) and single cells.

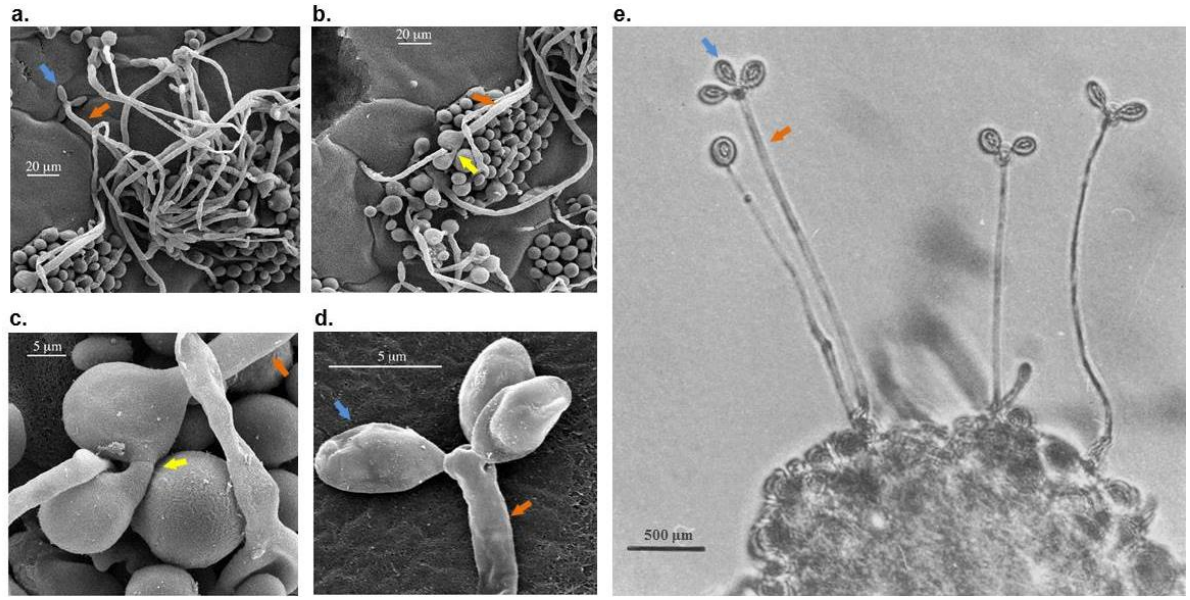


Figure 1.5. Sexual state of *P. rhodozyma*. (a-d) Photos from scanning electron microscopy showing a sporulating colony of *P. rhodozyma* cells. Yellow arrows point to two conjugated cells; orange arrows indicate basidia and blue arrows point to basidiospores. (Adapted from the site of the American Society for Microbiology, *http://202.195.144.50/ASM/106-Introduce.htm*. Photos by *Carlos Echavarri-Erasun* and *Eric Johnson*). **(e)** Micrograph of a colony with aerial basidiospores of strain CBS 6938. (Adapted from Kucsera et al., 2000).

1.2.4. Biotechnological relevance of *Phaffia rhodozyma*

Phaffia is the only basidiomycetous yeast that produces the antioxidant compound astaxanthin (Andrewes *et al.* 1976; Johnson & Lewis 1979), a high-value carotenoid, widely used in aquaculture rations, poultry feeds and with numerous applications in the pharmaceutical industry (Schmidt *et al.* 2011; Mata-Gomez *et al.* 2014; Sandmann 2015). Furthermore, *P. rhodozyma* represents one of the few commercially exploited natural sources of astaxanthin (Schmidt *et al.* 2011; Sandmann 2015). In addition to the production of astaxanthin, *P. rhodozyma* produces other molecules with potential biotechnological application, such as Phaffiol, an antioxidant compound (Jinno *et al.* 1998) and mycosporine-glutaminol-glucoside (MGG), a UVB-screening compound (Libkind *et al.* 2005; Libkind *et al.* 2011a; Moliné *et al.* 2014). Moreover this yeast has the ability to ferment different sugars (Reynders *et al.* 1997), a rare attribute for basidiomycetous yeasts, which adds to its biotechnological applicability (Lukács *et al.* 2006; Rodriguez-Saiz *et al.* 2010). In order to increase the economic viability of *P. rhodozyma* as a biological source of astaxanthin, it is important both to optimize the fermentation parameters and to improve the astaxanthin yields of the industrially used strains (Lukács *et al.* 2006; Rodriguez-Saiz *et al.* 2010). With these goals in mind, and taking advantage of the amenability of the species, different strategies have been developed, including random mutagenesis (Adrio *et al.* 1993; Miao *et al.* 2011) and genetic engineering of the biosynthetic route of astaxanthin (Breitenbach *et al.* 2011; Loto *et al.* 2012; Gassel *et al.* 2014; Ledetzky *et al.* 2014). Optimization of the growth conditions of *P. rhodozyma* has also been widely studied in order to increase the production of astaxanthin through the utilization of low cost carbon sources (Vázquez *et al.* 1997;

Dominguez-Bocanegra & Torres-Munoz 2004; Wozniak *et al.* 2011). All these studies spurred the development of different genetic tools including different yeast transformation protocols (Adrio & Veiga 1995; Rubinstein *et al.* 1996; Martinez *et al.* 1998; Wery *et al.* 1998; Visser *et al.* 2005) and mutant selection approaches, as well as, strategies for the construction of targeted knockout mutants and gene overexpression (Lin *et al.* 2012; Niklitschek *et al.* 2012; Hara *et al.* 2014a; Hara *et al.* 2014b).

1.3. Objectives and outline

The main objective of the work presented in this thesis was the elucidation of the molecular determinants and their underlying mechanisms governing the homothallic sexual cycle of *P. rhodozyma.* This work also aimed to provide the first insights regarding primary homothallism in basidiomycetes, which remain greatly understudied at the molecular level. **Chapter 1** consists of a general introduction highlighting the importance of sexual reproduction in fungi, summarizing its different modes of reproduction and indicating the most relevant molecular mechanisms that allow for the great diversity of sexual behaviors and mating type-determining systems observed in this group of organisms. Additionally, it describes the most significant aspects of the biology of the yeast *Phaffia rhodozyma.* **Chapter 2** describes the identification, in available draft genomes of *P. rhodozyma*, of genes pertinent to mating determination in basidiomycetous fungi, as well as, key genes involved in the meiotic process. Identified *MAT* genes were further characterized and also amplified in a group of *P. rhodozyma* strains encompassing the known population diversity of this species, allowing for phylogenetic analyses. **Chapter 3** describes the genetic manipulation of *MAT* genes in the haploid *P. rhodozyma* strain CBS 6938 by the construction of knockout mutants and their phenotypic characterization regarding their ability to undergo sexual reproduction. The obtained results allowed the proposition of a molecular model for the homothallic life cycle of *P. rhodozyma*. **Chapter 4** describes phenotypic and physiologic tests performed on two putative new species belonging to the genus *Phaffia*. Furthermore, it describes preparation of genomic DNA of strains from distinct species belonging to the Cystofilobasidiales, including the two new *Phaffia* species, for whole genome sequencing. Phylogenetic inferences were obtained using information collected from the assembled genomic data. **Chapter 5** explores the draft genomes of *Phaffia* and other Cystofilobasidiales species regarding their *MAT* gene content and organization, attempting to relate the genomic information gathered with the mating behavior of the studied species. Finally, in **Chapter 6**, some general conclusions of the research described in this thesis are presented and also some future perspectives.

CHAPTER 2

Identification of *MAT* and meiosis related genes in *Phaffia* rhodozyma genomes

Parts of the work presented in this chapter were published in:

Bellora N, Moliné M, **David-Palma M**, et al. (2016) Comparative genomics provides new insights into the diversity, physiology, and sexuality of the only industrially exploited tremellomycete: *Phaffia rhodozyma*. BMC Genomics 17, 901. doi: 10.1186/s12864-016-3244-7

David-Palma M, Sampaio JP, Gonçalves P (2016) Genetic Dissection of Sexual Reproduction in a Primary Homothallic Basidiomycete. PLOS Genetics 12(6): e1006110.
doi: 10.1371/journal.pgen.1006110

2.1. Introduction

In heterothallic basidiomycetous yeasts, sexual identity is usually determined by two genetically unlinked *MAT* loci. Mating of two compatible strains of distinct mating types is generally initiated by the interaction of pheromones (Mfa) and pheromone receptors (Ste3), encoded at the *PR* locus, that mediate cell-cell recognition leading to cell fusion. Subsequently, the progression through the sexual cycle is controlled by homeodomain transcription factors of two different homeodomain classes (Hd1 and Hd2), encoded at the *HD* locus (Fraser *et al.* 2007; Stanton & Hull 2007). Besides the *MAT* genes, which are directly involved in determining sexual identity, other mating related genes, not necessarily encoded at the *MAT* loci, are also required for completion of sexual reproduction. Examples of such genes include those involved in the pheromone response pathway that are activated upon pheromone/receptor interaction, namely the genes encoding the subunits of the heterotrimeric G protein and those composing the mitogen-activated protein kinase cascade itself (Bardwell 2004; Roach *et al.* 2014). The presence or absence of these genes in homothallic basidiomycetes, including *P. rhodozyma* remained largely unstudied. Therefore, this chapter focuses on the search and identification of *MAT* and mating related genes in the draft genomes (CBS 7918^{T} and CRUB 1149) generated by the study of Bellora et al. (2016). Additionally, a search for genes deemed necessary for meiosis in most eukaryotes (Ramesh *et al.* 2005; Malik *et al.* 2008; Schurko & Logsdon 2008; Halary *et al.* 2011) was also performed in order to ascertain if these were present in *P. rhodozyma*. Phylogenetic inferences were performed using the *MAT* genes found, including not only sequences obtained from the draft genomes available but also sequences obtained from a group of strains encompassing the different populations shown to exist within *P. rhodozyma* (David-Palma *et al.* 2014). Considering that most heterothallic basidiomycetes are often multiallelic for one or both *MAT* loci, it was pertinent to assess the possibility of the existence of cryptic molecular mating types among the different strains of *P. rhodozyma,* in spite of the homothallic behavior of the species. During this work additional genomic information pertaining to strain CBS 6938 (Sharma *et al.* 2015) was published and allowed further comparisons between the *MAT* region found in different strains of *Phaffia rhodozyma*.

2.2. Materials and Methods

2.2.1. Search for putative *MAT* and meiosis genes in *P. rhodozyma* genomes and their characterization

The genomic regions containing *MAT* genes, namely the homeodomain transcription factors (*HD1/HD2*) and the mating pheromones (*MFA*) and receptors (*STE3*) were searched in the draft genome assemblies of *P. rhodozyma* CBS 7918^T and CRUB 1149 (Bellora *et al.* 2016), or in local databases of proteins resulting from annotation of those draft genomes, by reciprocal BLASTP or TBLASTN, respectively. As queries, *MAT* proteins from *C. deneoformans* were used (Sxi1, Sxi2, Mfa1, and Ste3) (Table 2.3. and Table I.1, Appendix I). Pheromone precursor genes failing detection by BLAST due to their short length and highly variable sequences were identified manually upon inspection of the genomic regions near pheromone receptor genes, by searching for the existence of ORFs whose deduced protein sequences contained a conserved "CaaX" motif at the C-terminus. To ascertain the contiguity of the scaffolds harboring the two sets of pheromone and receptor genes, a pair of primers (MP100 5′-TCCATCCTCAACTGATTGC-3′ and MP103 5′-TTCATCTTGTCAGACAGC-3′) were used to amplify and partially sequence the region between both pheromone precursor genes. Standard PCR and cycling conditions were used with Phusion® High-Fidelity DNA Polymerase using an annealing temperature of 51 °C and extension for 90 s. Protein sequences of genes involved in the pheromone signaling cascade in *C. deneoformans* (Roach *et al.* 2014), were used to identify the corresponding putative orthologs in *P. rhodozyma* by the same approach detailed above (Table I.1, Appendix I). Similarly, protein sequences of 61 genes known to be involved in meiosis in *Saccharomyces cerevisiae*, 41 of which have inferred orthology in *C. deneoformans* and other basidiomycetes (Schurko & Logsdon 2008; Halary *et al.* 2011; Gioti *et al.* 2013a), were retrieved from GenBank, and the corresponding orthologs in *P. rhodozyma* were identified (Table I.2, Appendix I). Core meiotic genes and meiosis-specific genes were categorized according to previously published sources (Ramesh *et al.* 2005; Malik *et al.* 2008; Halary *et al.* 2011).

The transmembrane regions in the pheromone receptor proteins were predicted by HMMTOP software (Tusnady & Simon 2001). For the deduced Hd1 and Hd2 proteins, homeodomain regions were determined using InterPro server (Mitchell *et al.* 2015) while nuclear localization signals (NLS) were predicted using SeqNLS server (Lin & Hu 2013) with a 0.5 cutoff. Potential alpha-helixes were predicted by Jpred4 (Drozdetskiy *et al.* 2015) and searches for coiled-coil dimerization motifs were conducted using COILS with a sliding windows of 28, weighing option and probability ≥90% (Lupas *et al.* 1991), available at the Bioinformatics Toolkit server (https://toolkit.tuebingen.mpg.de/) (Alva *et al.* 2016).

2.2.2. Comparison of *MAT* gene sequences between the three available *P. rhodozyma* genomes

Complete sequences of *STE3-1, STE3-2, MFA1, MFA2, HD1 and HD2* genes were retrieved from all three *P. rhodozyma* genomes (Sharma *et al.* 2015; Bellora *et al.* 2016). Protein sequences were deduced after removal of putative introns according to proteins CED85384, CED85379, CDZ96688 and CDZ96689 from strain CBS 6938 (Sharma *et al.* 2015). Sequences from each of the predicted proteins were aligned with ClustalW as implemented in Bioedit (Hall 1999) and compared (Figure I.1, Appendix I). Predicted pheromone receptor proteins were aligned with ClustalW as implemented in Bioedit (Hall 1999) and amino acid sequence identity was calculated between alleles of each strain or between alleles of distinct mating types of the same species (Table I.3, Appendix I).

2.2.3. Phylogenetic analysis of the putative *MAT* genes in *P. rhodozyma* strains representing distinct populations

Mating type genes *STE3*-1, *STE3*-2, *HD1* and *HD2* were amplified and sequenced in several *P. rhodozyma* strains (Table 2.1 and Table 2.2). All sequences obtained were deposited in Genbank (Table I.4, Appendix I). Nucleotide unrooted maximum likelihood phylogenies were inferred with General Time Reversible model and 1000 bootstrap replications on MEGA5.1 software (Tamura *et al.* 2011). The trees with the highest *log* likelihood are shown with branch lengths measured in the number of substitutions per site (Figure 2.4).

Table 2.1. List of strains used in the phylogenetic study.

Strain	Geographic origin	Isolation substrate	Population*
CRUB 0853	Patagonia, Argentina	*Cyttaria hariotii* on *Nothofagus dombeyi*	A
CRUB 1149	Patagonia, Argentina	Water sample near *Cyttaria hariotii* on *Nothofagus pumilio*	
CRUB 1490	Patagonia, Argentina	*Cyttaria hariotii* on *Nothofagus dombeyi*	
CRUB 1151	Patagonia, Argentina	Water sample near *Cyttaria hariotii* on *Nothofagus pumilio*	
ZP 869	Tasmania, Australia	*Cyttaria gunnii* on *Nothofagus cunninghamii*	B
ZP 922	Haast Pass, New Zealand	*Cyttaria nigra* on *Nothofagus menziesii*	
GY13L04	Taiwan	Soil sample near *Stauntonia purpurea*	
KBP 2604	Novgorod, Russia	Exudate of *Betula* sp.	C
ATCC 24261	Yamagata, Japan	Exudate of *Betula maximowicziana*	
NRRL Y-17434	Illinois, USA	Exudate of *Betula populifolia*	
CBS 7918	Moscow, Russia	Exudate of *Betula verrucosa*	
CBS 6938	Finland	Exudate of *Betula* sp.	
ATCC 24201	Hiroshima, Japan	Exudate of *Cornus brachypoda*	D
ATCC 24229	Hiroshima, Japan	Exudate of *Cornus brachypoda*	

**Based on David-Palma et al. (2014) phylogenetic grouping using MLS*

Table 2.2. List of primers and PCR conditions used to amplify *MAT* genes in *P. rhodozyma* strains.

Primer	Sequence (5' – 3')	Annealing/ extension	Remarks
MP029	CGGTGGGCGTTCTGGTCGGAC	63ºC / 1 min	partial amplification of *STE3-1* gene (836 bp)
MP030	ACTCTGATGGCGAAGCAACGGC		
MP035	TTATGCATCAACCGGCGTCTGGCA	62ºC / 1 min	partial amplification of *STE3-2* gene (929 bp)
MP036	GGACACAGAGGCAACRGTAGTTCCA		
MP154	ATGAGATCGTTCAACAGC	50ºC / 2min 30 s	partial amplification of *HD1* gene (2580 bp)
MP086	GTCTTCCGTTCTTTCTCG		
MP160	TTCATCTAGACCATCCTTATCC	50ºC / 1 min 30 s	partial amplification of *HD1* gene (1582 bp)
MP161	AAGATGGCGAGAATGTAGGTATCG		
MP085	TGTACAGTTTCACGAAGC	46ºC / 1 min 30 s	partial amplification of *HD2* gene (1595 bp)
MP155	ACATTTCATTAAGGCTGG		
MP085	TGTACAGTTTCACGAAGC	46ºC / 1 min	partial amplification of *HD2* gene (916 bp)
MP086	GTCTTCCGTTCTTTCTCG		
MP049	AGATAGAAACCCAACACTCGC	50ºC / 2 min	partial amplification of N-terminal regions of *HD1* and *HD2* genes and intergenic region (1773 bp)
MP050	GGGAGATGATCGTCGATTTCG		

2.2.4. Phylogenetic analysis of the pheromone receptor proteins of *P. rhodozyma* and other Tremellomycetes

Previously published pheromone receptor sequences were used to reconstruct the phylogenetic tree of Ste3 (Figure 2.5.). The accession numbers of sequences used are listed in Table I.4 (Appendix I). Available draft genome sequences of *Kwoniella mangrovensis* CBS 10435, *K*. *heveanensis* BCC 8398 and *Tremella fuciformis* tr26 were searched for the presence of pheromone receptor homologues by TBLASTN using *P*. *rhodozyma* Ste3-1 as query. Genomic regions corresponding to positive hits were retrieved (GenBank accession numbers ASQD01000019.1, ASQB01000005 and LBGW01000351, respectively) and protein sequences were deduced after removal of putative introns, either manually or using AUGUSTUS (Stanke *et al.* 2004). The final protein dataset was aligned using an iterative refinement method (L-INS-i) in MAFFT v.7.221 (Katoh & Standley 2014). Poorly aligned regions were removed with trimAl v.1.2 (Capella-Gutierrez *et al.* 2009) using the "gappyout" option. The resulting alignment containing 338 positions was analyzed in ProtTest v.3.2 using the corrected Akaike information criterion (AICc) to determine the model of sequence evolution that best fitted the data. A maximum likelihood-based phylogenetic tree was built in RAxML v.8.1.24 (Stamatakis 2014) using PROTGAMMAILGF model of amino acid substitutions and branch support was determined using 1000 rapid bootstraps. The *Saccharomyces cerevisiae* Ste3 pheromone receptor was used to root the tree.

2.3. Results

2.3.1. *MAT* and mating related genes in *P. rhodozyma*

In the genome assembly of the *P. rhodozyma* type strain, CBS 7918^T, three scaffolds were identified as harboring putative *MAT* gene homologs (Stanton & Hull 2007), namely the pheromone receptors genes (*STE3*), mating pheromone precursor genes (*MFA*) and the *HD1* and *HD2* homeodomain transcription factors genes (Table 2.3.) Similar results were obtained for strain CRUB 1149 (Table 2.3.).

Table 2.3. List of *MAT* genes found in *P. rhodozyma* genomes

		Scaffold number in each draft genome		
Gene	Function	CBS 7918	CRUB 1149	CBS 6938*
STE3-1	Receptor gene for mating pheromones	226	186	LN483332
STE3-2	Receptor gene for mating pheromones	253	198	LN483332
MFA1	Pheromone precursor gene	226	186	LN483332
MFA2	Pheromone precursor gene	253	198	LN483332
HD1	Homeodomain transcription factor	60	63	LN483167
HD2	Homeodomain transcription factor	60	63	LN483167

*(Sharma *et al.* 2015)

The receptors and pheromone precursor genes were located on two different scaffolds, i.e. *STE3-1* and *MFA1* were found in scaffold 226 and *STE3-2* and *MFA2* were found in scaffold 253 (Figure 2.1). However, in the genome of strain CBS 6938 (Sharma *et al.* 2015), the two sets of genes (*STE3-1/MFA1* and *STE3-2/MFA2*) were located on the same scaffold approximately 5 kb apart. Using PCR, it was possible to amplify (and sequence) the stretch between the two *MFA* genes, thus confirming that the two gene sets were similarly positioned in strain CBS 7918^T, suggesting that these two scaffolds (226 and 253) are linked and at a similar distance as in strain CBS 6938 (Figure 2.1).

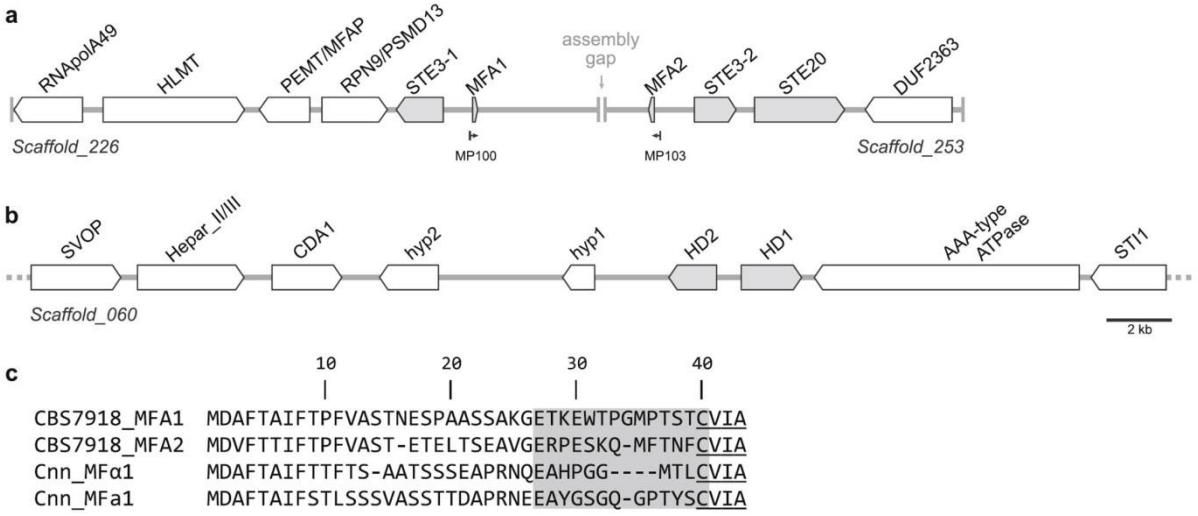


Figure 2.1. Gene content and organization around the putative mating type genes of *P. rhodozyma*. (**a)** Representation of the putative *PR* locus encoding mating pheromone receptors (*STE3-1* and *STE3-2*) and pheromone precursor genes (*MFA1* and *MFA2*). (**b)** Representation of the putative *HD* locus encoding homeodomain transcription factors (*HD1* and *HD2*). Genes are indicated by arrows showing the

direction of transcription. Arrows in grey indicate putative mating related genes while those in white depict additional genes apparently unrelated to mating. Hyp1 and hyp2 are hypothetical predicted proteins of uncertain function. The end of a scaffold is indicated by a solid vertical bar. The contiguity of the scaffolds harboring the two sets of pheromone and receptor genes was confirmed by PCR with primers MP100 and MP103. **(c)** Sequence alignment of the putative pheromone precursors from *P. rhodozyma* (CBS 7918^T) and *C. deneoformans* (Cnn Mfα1 - XP_570122.1; Cnn Mfa1- AAG42766.1). The predicted mature pheromones are shaded in grey and the "CaaX"-motifs for C-terminal processing are underlined.

The two pheromone receptor genes were predicted to encode different proteins with about 50 % sequence identity (Table I.3 and Figure I.1.c, Appendix I) and seven transmembrane domains were found in each (Figure I.2., Appendix I). The two pheromone precursor genes found (*MFA1* and *MFA2*) were predicted to encode proteins that are about 58 % identical and yield different mature pheromones that have a conserved "CaaX" motif at the C-terminal part of the precursor protein (Figure 2.1.c). A putative *HD* locus was found with two divergently transcribed genes, *HD1* and *HD2*, on scaffold 60 (Figure 2.1.b). The Hd1 and Hd2 predicted proteins both possess a homeodomain region, with Hd1 presenting a characteristic three amino-acid loop extension between helix I and helix II (Figure I.3. and Figure 2.2.)

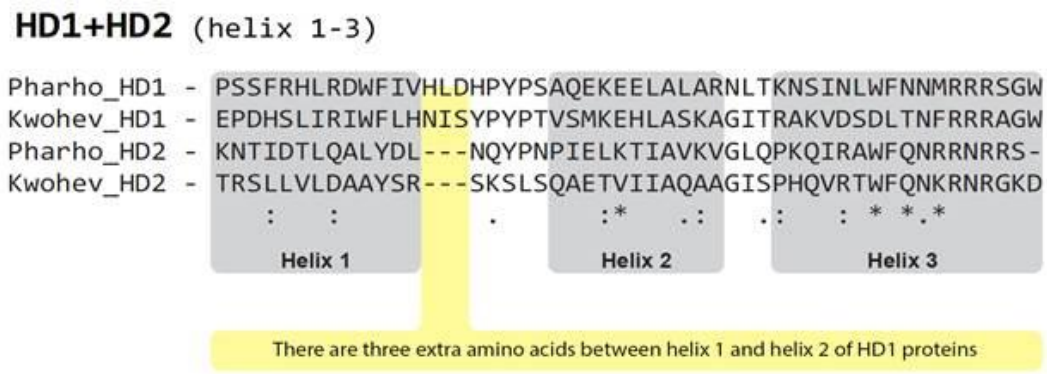


Figure 2.2. Features of Hd1 and Hd2 proteins of *P. rhodozyma*. Sequence alignment of the three-helical region of *P. rhodozyma* strain CBS 7918^T (Pharho_HD1 - G03139_P; Pharho_HD2 - G03138_P) and *Kwoniella heveanensis*, strain BCC15000 (Kwohev_HD1 - ACZ51528.1; Kwohev_HD2 - ACZ51529.1).

Of the two homeodomain transcription factors proteins, only for Hd2 was a nuclear localization signal (NLS) detected. Alpha helixes were detected at the N-terminal region of both proteins, however coiled-coil dimerization motifs were only found at the C-terminal region of Hd1 (Figure I.3., Appendix I). Orthologs of mating related genes encoding components of the conserved pheromone response pathway that is activated upon pheromone/receptor interaction were identified, namely the genes encoding the subunits of the heterotrimeric G protein (*GPA1-3*, *STE4* and *STE18*) and those that compose the mitogen-activated protein kinase module itself (*STE11*, *STE7* and *CPK1*) (Table I.1, Appendix I). Moreover, a p21-activated kinase gene (*STE20*) was identified in the vicinity of the *STE3-2* gene on scaffold 253 (Figure 2.1. and Table I.1, Appendix I). A set of putative orthologs encoding transcription factors, Ste12, Mat2, Znf2 and Prf1, that have key roles in mating in *S. cerevisiae*, *C. deneoformans* and *U. maydis* respectively, were also found (Table I.1, Appendix I). The analysis of the genome assembly of *P. rhodozyma* strain CRUB 1149, yielded identical results to those obtained with the type strain (CBS 7918^T).

2.3.2. Meiosis genes in *P. rhodozyma*

Overall, 42 out of the 61-searched meiosis-related genes were identified in *P. rhodozyma* (Table I.2., Appendix I). Of the 30 “core” meiotic genes, which encode a set of proteins of the conserved meiotic recombination machinery of eukaryotes, 29 orthologs were found in *P. rhodozyma* except for *MLH2* (Table I.2., Appendix I).

2.3.3. *MAT* gene variability within *P. rhodozyma*

A comparison of the three-full length amino acid sequences of Ste3-1 and Ste3-2 from the available draft genomes (Figure I.1. a and b) showed a maximum of seven and three amino acid substitutions between variants of Ste3-1 and Ste3-2, respectively. These differences were found between strains belonging to different populations (CRUB 1149 from population A vs. CBS 7918^T and CBS 6838 from population C), while the two strains belonging to the same population (CBS 7918^T and CBS 6838) exhibited identical amino acid sequences. The predicted pheromone precursor proteins Mfa1 of all three strains are 100% identical, with the same being true for Mfa2 (Figure I.1. d and e). To assess whether additional *STE3-1*, *STE3-2*, *HD1* or *HD2* alleles, potentially encoding different mating identities, could be retrieved in *P. rhodozyma*, sequences of these genes were obtained for a group of strains representing all four previously identified populations of *P. rhodozyma* (Figure 2.4a) (David-Palma *et al.* 2014). This survey uncovered 10 variants for both the *STE3-1* and the *STE3-2* gene; however, the variants found for each gene are extremely similar to each other. Comparison of the complete predicted protein sequences of the homeodomain transcription factors Hd1 and Hd2 of a group of strains encompassing all known populations of *P. rhodozyma* showed that the proteins are quite similar (Figure 2.3).

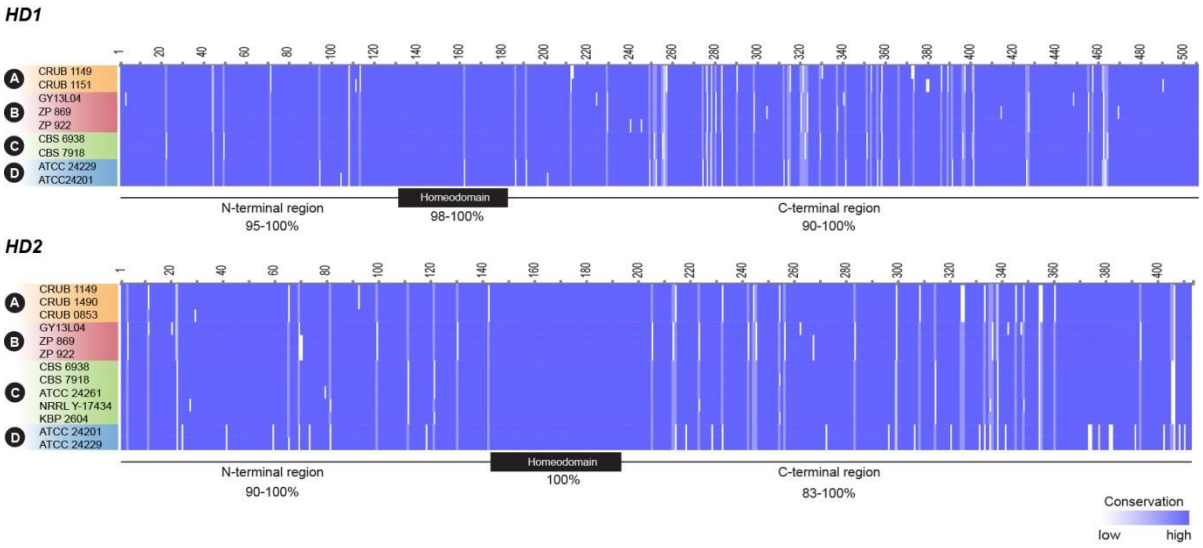


Figure 2.3. Alignments of the predicted proteins for Hd1 and Hd2 of distinct *P. rhodozyma* strains. For both alignments, strains are listed according to their population, delimitated by different color blocks (Population A: orange; Population B: red; Population C: green; Population D: blue). Conservation of the aligned proteins is depicted in different shades of blue, as indicated in the key bellow. Homeodomain regions are indicated below the alignments by black boxes. Amino acid identity of each region is indicated below each alignment.

Comparison of the Hd1 sequences of strains from distinct populations revealed a high level of conservation within the homeodomain, which presented a single conservative amino acid substitution in the strains from population D (Figure 2.3 and Figure I.1f) and an even distribution of polymorphic sites, in both N-terminal and C-terminal regions. Similarly, the homeodomain region of the Hd2 is also highly conserved being identical in all the variants identified (Figure 2.3). In contrast with all the Hd1 sequences studied, which all present the same length, Hd2 sequences representing different *P. rhodozyma* populations exhibited different C-terminal region lengths, except for the shortest variant (Figure I.1g) that was found in strains representing populations B and C.

2.3.4. Phylogenies of *MAT* genes within *P. rhodozyma* species

The *STE3-1* and *STE3-2* phylogenies including all the identified variants reproduce well the previously reported phylogenetic clusters (populations) within *P. rhodozyma* (Figure 2.4a). Similarly, to what was observed for the genes in the *PR* loci, *HD1* and *HD2* gene sequences, from the *HD* locus, retrieved from strains representing all four *P. rhodozyma* populations recapitulate previously reported phylogenetic relationships between the different strains (Figure 2.4b).

a.

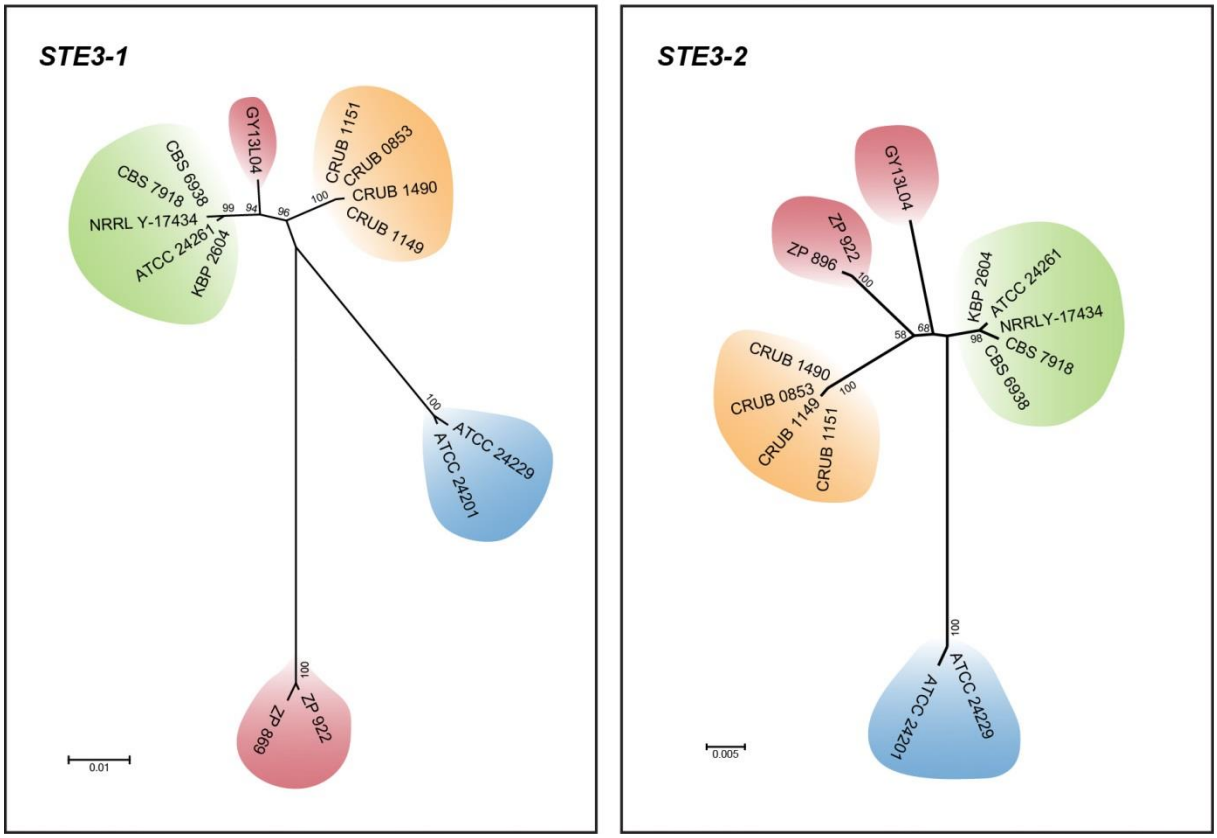


Fig 2.4. Phylogenies of *MAT* genes. Description is given in page 30.

b.

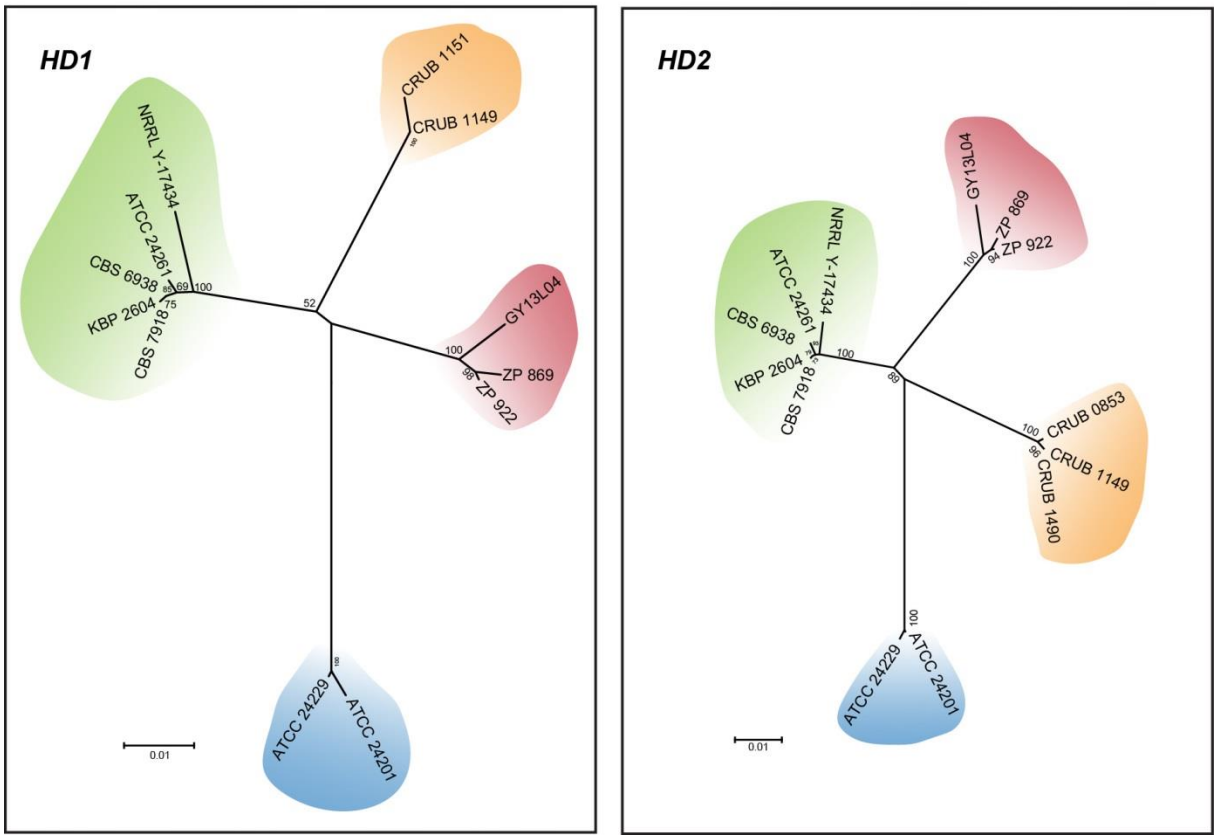


Fig 2.4. Phylogenies of *MAT* genes (continued). Distinct colors encompassing groups of strains indicate the *P. rhodozyma* populations to which the various strains belong. Population A: orange; Population B: red; Population C: green; Population D: blue. Nucleotide unrooted maximum likelihood phylogenies were inferred with General Time Reversible model and 1000 bootstrap replications on MEGA5.1 software (Tamura *et al.* 2011). The trees with the highest log likelihood are shown with branch lengths measured in the number of substitutions per site. In the final datasets, *STE3-1*, *STE3-2*, *HD1* and *HD2* have 703, 735, 814 and 1441 positions respectively.

2.3.5. Phylogenetic relationship of pheromone receptors of *P. rhodozyma* and other basidiomycetes

A phylogenetic analysis was conducted with the predicted receptor proteins Ste3-1 and Ste3-2 of *P. rhodozyma* CBS 7918^T and their closest known relatives from heterothallic species harboring two mating type specific Ste3 receptors, from the *Cryptococcus*, *Kwoniella* and *Tremella* lineages. The two *P. rhodozyma* receptors appear to share a more recent common ancestor with each other than with the receptors found in the other species (Figure 2.5). In line with this, the protein sequences of Ste3-1 and Ste3-2 receptors from *P. rhodozyma* have diverged considerably less (50% amino acid identity) than the two Ste3 alleles found in the other (heterothallic) species examined, e.g. the Ste3 alleles from opposite mating types of *C. deneoformans* have 31% amino acid identity with each other (Table I. 4).

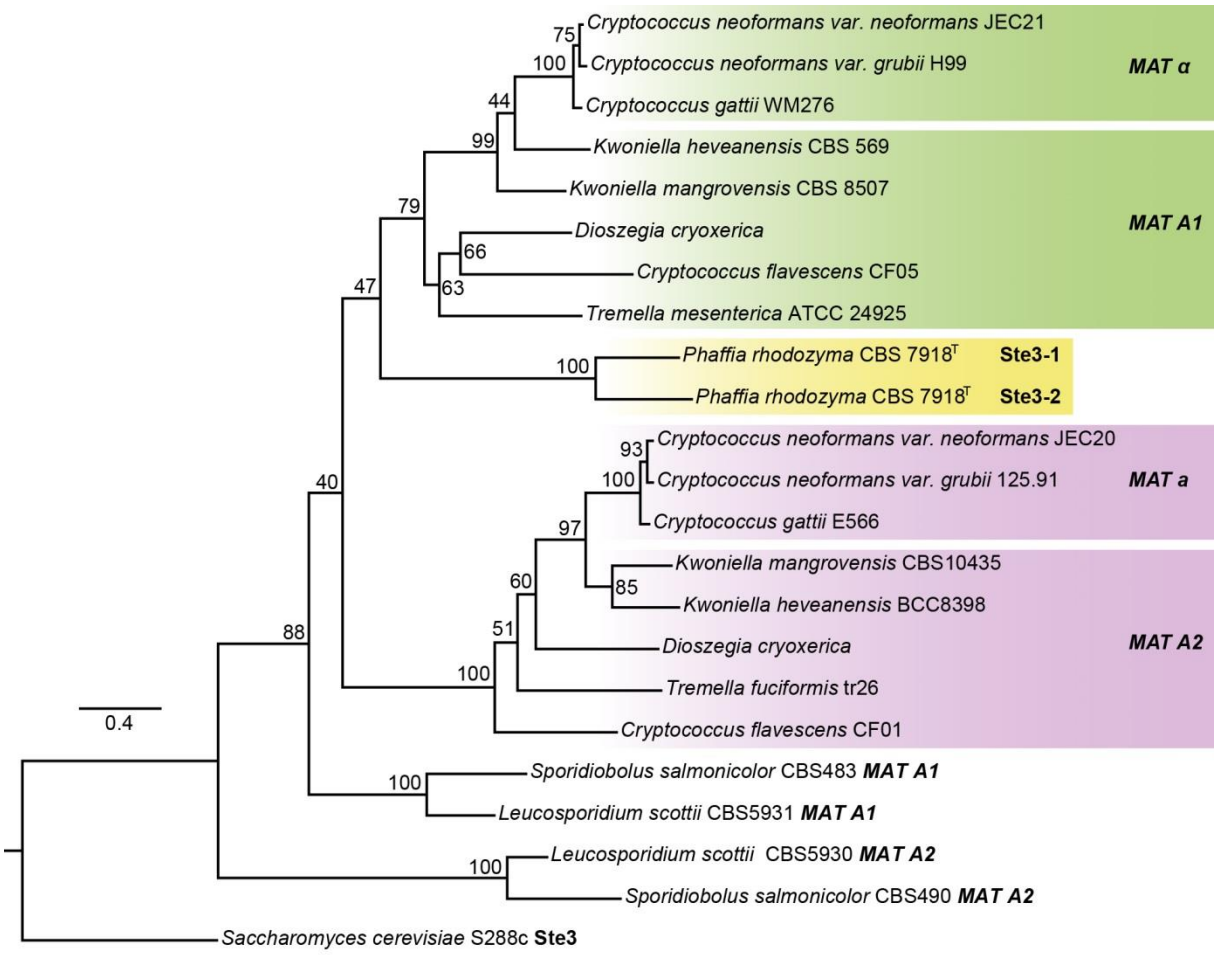


Figure 2.5. Maximum likelihood phylogenetic tree of pheromone receptors (Ste3) from Tremellomycetes. Phylogeny includes Ste3 protein sequences from the species within Tremellomycetes closest to *Phaffia* and is rooted with the sequences of both Pucciniomycotina species (*L. scottii* and *S. salmonicolor*) and *S. cerevisiae* Ste3 pheromone receptors.

2.4. Discussion

2.4.1. *MAT* genes are present in *P. rhodozyma*

In each of the three available genomes of *P*. *rhodozyma* (strains CBS 7918^T, CRUB 1149 and CBS 6938) putative *MAT* genes that determine sexual identity in basidiomycetes were found (Table 2.3).

The putative *PR* locus appears to be composed by two clusters, each encoding a pheromone precursor and a Ste3 type receptor (henceforth referred to as *PR1* and *PR2*) located at approximately 5 kb from each other (Figure 2.1). The Ste3 receptors and Mfa precursors encoded in the clusters *PR1* and *PR2* are clearly different from each other (Figure 2.1 and Figure I.1c). The pheromone receptors exhibit the characteristic seven transmembrane domains while the pheromone precursor proteins have the "CaaX" motif that signals the precursor for post-translation modifications (Kües *et al.* 2011). Considering that the occurrence in the same genome of multiple pheromone receptor genes is quite common in heterothallic species within the Agaricomycotina, and that different combinations of *PR* clusters may define distinct mating type identities (Riquelme *et al.* 2005; Kües *et al.* 2011; Kües 2015) it was necessary to assess whether additional pheromone receptors alleles, potentially encoding different mating identities, could be retrieved in *P*. *rhodozyma*. A comparison of the complete predicted proteins of Ste3-1 and Ste3-2 from the available draft genomes of *P. rhodozyma*, showed them to be very similar. The same degree of similarity was observed for the partial sequences obtained for the *STE3-1* and *STE3-2* genes of 14 *P. rhodozyma* strains (Figure 2.4). Both the *STE3-1* and *STE3-2* phylogenies (Figure 2.4a) reproduce the same strain clusters (populations) obtained with phylogenies of non-*MAT* genes (David-Palma *et al.* 2014). Given the similarity between the Ste3 encoded in strains of distinct populations and the fact the different strains encode identical Mfa1 and Mfa2 pheromone precursors, it appears that the *PR* locus is not bi- or multiallelic in *P*. *rhodozyma*, as opposed to heterothallic basidiomycetes. Many heterothallic basidiomycete species, both bipolar like *Cryptococcus deneoformans* (Lengeler *et al.* 2002) or tetrapolar, like *Leucosporidium scottii* (Maia *et al.* 2015) harbor two functionally distinct, and thus mating type determining, Ste3 receptors. When phylogenies are constructed using the amino acid sequences of these two pheromone receptors identified in various basidiomycete lineages, trans-specific polymorphism is usually observed, with ***a1*** alleles of the different species branching together rather than clustering with the ***a2*** allele from the same species (Devier *et al.* 2009; Coelho *et al.* 2010; Kües *et al.* 2011). The phylogenetic inference performed with the Ste3-1 and Ste3-2 receptors of *P*. *rhodozyma* in addition to receptor proteins from other Tremellomycetes species, showed (Figure 2.5) that, while reproducing the expected trans-specific polymorphism for all other species examined, the two *P*. *rhodozyma* receptors are an exception, sharing a more recent common ancestor with each other.

A putative *HD* locus was also identified in *P*. *rhodozyma*, being composed of one pair of divergently transcribed *HD1-HD2* genes (Figure 2.1). Comparison of complete sequences of Hd1, from representative strains of all known populations of *P. rhodozyma*, showed them to be highly conserved, the same being true for Hd2 sequences (Figure 2.3). The highest degree of conservation was

observed in the homeodomain regions, while few polymorphisms were distributed throughout the N-terminal and C-terminal regions of the HD proteins (Figure 2.3). This is in contrast to what is usually observed for different alleles of either Hd1 or Hd2 proteins in heterothallic species, where the N-terminal dimerization domain is much more divergent than the C-terminal domain, as a result of negative frequency dependent selection imposed by the functional constraints on the N-terminus of both the HD proteins (Badrane & May 1999; May *et al.* 1999). Although exceptions to this pattern exists, like *P. chrysosporium* that presents a hyper-polymorphic C-terminal and a more conserved N-terminal region (James *et al.* 2011), the pattern observed in *P. rhodozyma* is thus far unique in the polymorphic pattern presented (Figure 2.3). Similarly to what was observed for the phylogenies of the two pheromone receptor genes, the inferred phylogenies for gene *HD1* and *HD2* (Figure 2.4b) also reproduce the same populations obtained with phylogenies of non-*MAT* genes (David-Palma *et al.* 2014). Considering the results obtained for the *MAT* genes present at both the *PR* and *HD* loci and the variants encoded in either of them, in the group of strains studied, there appears to be no evidence of polymorphisms at the *MAT* loci that might represent cryptic molecular mating-types. Additionally, given that in all the different draft genomes available for *P. rhodozyma* (Sharma *et al.* 2015; Bellora *et al.* 2016), the *PR* and *HD* loci were always located in distinct scaffolds, this suggests that the two loci are probably genetically unlinked.

2.4.2. Mating related genes and meiotic genes are present in *P. rhodozyma*

Putative orthologs of mating related genes encoding components of the conserved pheromone response pathway that is activated upon pheromone/receptor interaction were found in all *P. rhodozyma* genomes (Table I.1, Appendix I). This indicates that similarly to heterothallic basidiomycetes there is conservation of the pheromone activated pathway (Roach *et al.* 2014). The homothallic life cycle of *P. rhodozyma* is presumed to involve meiosis (Kucsera *et al.* 1998; Kucsera *et al.* 2000). However, evidence exists for non-Mendelian segregation of markers, aneuploidy (Kucsera *et al.* 1998) and chromosomal length polymorphisms (Nagy *et al.* 1994; Adrio *et al.* 1995), which could be regarded as pointing against the occurrence of a typical meiosis. To get further insights regarding this subject, the CBS 7918^{T} draft genome was searched for genes shown to be involved in meiosis in other fungi. Of the set of 30 proteins that compose the conserved meiotic recombination machinery of eukaryotes (Villeneuve & Hillers 2001; Ramesh *et al.* 2005; Malik *et al.* 2008; Halary *et al.* 2011), 29 orthologs were found in *P. rhodozyma*, with *MLH2* being the only gene undetected (Table I.2., Appendix I). However, absence of this gene is not specific to *P. rhodozyma*, since it was previously noted in other fungal species with known extant sexual cycles (e.g. *U. maydis* and *C. deneoformans*). Altogether, *P. rhodozyma* retained most of the genes associated with mating and meiosis.

2.5. Conclusion

Phaffia rhodozyma appears to have two distinct *MAT* loci. The survey of variants of *MAT* genes in strains representing the four natural populations identified so far in *P*. *rhodozyma* failed to uncover additional potentially functionally divergent homologs of *MAT* genes that could represent cryptic molecular mating-types. Therefore, the results suggest that the species may be considered exclusively homothallic in the sense that no molecular mating-types were identifiable, although it admits both haploid selfing (i.e. mating among haploid cells with identical genotypes resulting from clonal division of a mother cell) and outcrossing modes of reproduction, the first being apparently much more frequent, as supported by population analysis studies (David-Palma *et al.* 2014). The genetic makeup of the *MAT* loci in *P*. *rhodozyma*, although suggestive of primary homothallism, departs in important aspects from what would be expected from a simple assemblage in one genome of two mating-types as typically found in the Agaricomycotina. Firstly, in the sexually reproducing species most closely related to *P*. *rhodozyma*, two lineages of receptors can be clearly discerned exhibiting trans-specific polymorphism (Kües *et al.* 2011), while the phylogenetic analysis performed in this chapter (Figure 2.5) shows, with a high degree of confidence, that the two receptors in *P*. *rhodozyma* are more closely related with each other than with receptors in other species. These results suggest the possibility that *P*. *rhodozyma* receptors, similarly to species in the Agaricales, may descend from only one of these receptor lineages (Kües *et al.* 2011). A consequence of this is that the two receptors seem to have diverged relatively recently, at least after the separation of the *Phaffia* and *Cryptococcus* lineages. Secondly, the extant *MAT* configuration of *P. rhodozyma* also departs from the simple gathering in a single genome of two mating types, because contrasting with the two *PR* clusters found, only one *HD1/HD2* gene pair was uncovered in this species. Finally considering that the two *MAT* loci appear to be genetically unlinked it is possible that the extant genetic makeup of *P. rhodozyma* is derived from a tetrapolar ancestor. Given that no inactivating mutations were detected in any of the putative *MAT* genes uncovered in *P. rhodozyma*, it is therefore likely that they play an active role in the homothallic sexual cycle of this yeast.

CHAPTER 3

Genetic dissection of the homothallic sexual reproduction of *P. rhodozyma*

Part of the work presented in this chapter was published in:

David-Palma M, Sampaio JP, Gonçalves P (2016) Genetic Dissection of Sexual Reproduction in a Primary Homothallic Basidiomycete. PLOS Genetics 12(6): e1006110.
doi: 10.1371/journal.pgen.1006110

3.1. Introduction

Inspection of the genomes of three *P. rhodozyma* strains (Sharma *et al.* 2015; Bellora *et al.* 2016) revealed the presence of two pheromone precursor and pheromone receptor gene clusters and one *HD1/HD2* pair similar to those commonly found in heterothallic basidiomycete *MAT* loci (Figure 2.1). The two pheromone receptor genes exhibit considerable sequence divergence and each is flanked by a unique pheromone precursor gene. The *PR* locus is seemingly genetically unlinked to the *HD* locus, since in all *P. rhodozyma* draft genome assemblies examined the *STE3*/*MFA* and *HD1*/*HD2* genes are located on different scaffolds.

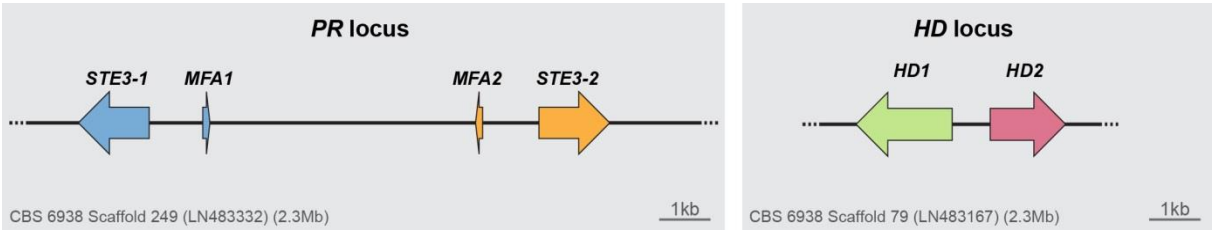


Figure 3.1. *MAT* loci in *P. rhodozyma*. Organization of the *PR* locus of strain CBS 6938 depicting the *STE3/MFA* gene clusters in blue and orange. Structure of the *HD* locus, depicting the two divergently transcribed homeodomain transcription factor genes, *HD1* and *HD2*.

The identification of the above-mentioned *MAT* genes raised the question of whether they had a role in the homothallic life cycle of *P. rhodozyma*. In chapter 2, the possibility that cryptic molecular mating types might exist among available *P. rhodozyma* strains, despite their homothallic behavior, was examined and discarded. In this chapter, genetic tools available for this species were used to undertake the dissection of the genetic underpinnings of sexual reproduction in *P. rhodozyma*. Although ploidy of *P. rhodozyma* strains is thought to vary among distinct strains from haploid to diploid or higher (Medwid 1998), all genetic studies performed so far in strain CBS 6838 indicate that it is haploid (Wery *et al.* 1997). This strain was, therefore, used as genetic background for the construction of various deletion mutants to assess if all six *MAT* genes identified were relevant for sexual reproduction. Based on the results obtained a mechanistic model for primary homothallism in a basidiomycete was proposed for the first time.

3.2. Materials and Methods

3.2.1. Strains and culture conditions

Escherichia coli DH5α (Table II.1, Appendix II) strain was used for all cloning steps and was grown in LB medium (10 g/L Tryptone, 5 g/L Yeast Extract and 5 g/L NaCl) with 100 µg/ml ampicillin at 37°C. Wild type *P. rhodozyma* strain CBS 6938 was grown in YPD medium (10 g/L Yeast Extract, 20 g/L Peptone and 20 g/L glucose) at 20°C, while mutants strains were grown in YPD medium supplemented with appropriate antifungal drugs (50 µg/ml geneticin, 50 µg/ml hygromycin or 100 µg/ml zeocin). Wild type and mutant strains of *P. rhodozyma* used or generated in this work are listed in Table 3.1.

3.2.2. Construction of the gene deletion fragments and complementation plasmids

Different gene deletion fragments (GDF) were generated to construct *P. rhodozyma* deletion mutants, using a common strategy, consisting in cloning the upstream and downstream flanking regions of the selected gene upstream and downstream of an antifungal resistance cassette, to promote integration of the GDF in the targeted genomic locus (Lin *et al.* 2012). Standard molecular biology methods were employed (Sambrook & Russell 2001; Nelson & Fitch 2011) and three distinct plasmids were used as backbone, namely pPR2TN (Verdoes *et al.* 1999), pBS-HYG (Niklitschek *et al.* 2008) and pJET1.2+ZEO (Table II.2, Appendix II), encoding geneticin, hygromycin and zeocin resistance genes, respectively (Figure II.1, Appendix II). Specific primers for the flanking regions of *STE3-1*, *HD1*, *HD2*, *SPO11*, regions *STE3-1/MFA1* and *STE3-2/MFA2* were designed to include restriction sites that allowed cloning of the amplified regions onto the chosen plasmids (Table II.2, Appendix II). The GDF for the *STE3-2* gene was constructed by overlap extension PCR and then cloned into the pJET1.2 vector using CloneJET PCR Cloning Kit (Thermo Scientific) (Table II.3, Appendix II). Nested primers were used to amplify the complete GDFs by PCR using Phusion High-Fidelity DNA Polymerase (Thermo Scientific) (Table II.2 and Table II.3, Appendix II). PCR products were purified using GeneJET Gel Extraction Kit (Thermo Scientific) or Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) and finally used to transform *P. rhodozyma*. Complementation of selected deletion mutants was accomplished using plasmid pUC18+rDNA+ZEO that was constructed by inserting the constructed 1.8 Kb zeocin resistance cassette (Table II.4, Appendix II) and a 3 Kb fragment of the rDNA from plasmid pPR2TN in plasmid pUC18 at the SmaI and SacI restriction sites respectively. DNA fragments containing the ORFs' of the genes pertaining to the complementation with approximately 300 bp flanking regions, were subsequently cloned into the PstI and BamHI restriction sites of pUC18+rDNA+ZEO (Figure II.1, Appendix II). Each of the four complementation plasmids were then linearized with ClaI within the rDNA sequence to promote integration of the plasmid into the ribosomal DNA of the mutant strains to be complemented (Table II.5, Appendix II) (Visser *et al.* 2005).

3.2.3. Transformation of *P. rhodozyma*

Linearized plasmids and GDF were used to transform *P. rhodozyma* by electroporation as previously described (Visser *et al.* 2005). Transformants were selected in YPD medium with the appropriate antifungal drugs. Mutants with multiple deletions were obtained by transforming a confirmed deletion mutant with a second or third GDF. Correct integration of the disruption cassettes was verified by PCR (Table II.2., Table II.3. and Table II.5., Appendix II) as previously described (Lin *et al.* 2012). Briefly, a primer inside the resistance cassette and a primer outside of the flanking region present in the GDF at both the 5' and 3' extremities were used in diagnostic PCR reactions to identify the desired mutants. Absence of the gene targeted for deletion was also assessed by PCR for each mutant. Key deletion mutants were also confirmed by Southern blot in order to ensure that integration of the GDF occurred only once and in the correct locus. For Southern blot, 5 µg of genomic DNA was digested with ClaI and run in a 0.8% agarose gel. Southern blot was performed using standard methods.
Primers MP091-MP092 and MP062-MP063 (5'-AAGATGGATTGCACGCAGGTTCTCC-3', 5'-TTCCACCATGATATTCGGCAAGCAGG-3', 5'-ACGTCTGTCGAGAAGTTTCTGATCG-3' and 5'-TTTGCCCTCGGACGAGTGCTGG-3', respectively) were employed to amplify fragments of the resistance genes present in the geneticin and hygromycin cassettes to be used as probes. Labeling of the probes was performed with (α-^{32}P) dATP using the Prime-a-Gene Labelling system (Promega). Signals were detected on X-ray films (Hyperfilm MP, GE Heathcare Life Sciences) (Figure II.2, Appendix II).

3.2.4. Sporulation assays

To test the ability of the deletion mutants to sporulate, the various mutants were inoculated on DWR (2.5% agar and 0.5% ribitol) solid medium, incubated at 18°C and observed regularly for up to two months. Sporulation efficiency assays were performed as described previously by Kucsera (Kucsera *et al.* 1998). Briefly, cells were grown on YPD medium overnight (180 rpms, 20°C, in 10% of the volume of the flask), collected by centrifugation and washed thoroughly with sterile distilled water to remove culture medium. Cells were subsequently distributed in 10 µl drops over the surface of DWR plates that were subsequently incubated at 18°C for 10 days. The number of basidia on each plate was determined by direct observation of the perimeter of the colonies using an optical microscope. Three independent assays were performed with CBS 6938 wild type (WT) strain and with all deletion mutants (in triplicate). Student's t-test was performed (with a significance of 99%) to ascertain the statistical significance of the differences observed between the WT and each of the sporulating deletion mutants (Table II.6 And Table II.7; Appendix II). Additional assays were conducted to determine the viability of F1 progeny of the *spo11Δ* mutant and wild type strain CBS 6938. Basidiospores were recovered by micromanipulation and were transferred to YPD solid medium to determine viability. Chi-square statistic was performed to verify if the difference between WT and *spo11Δ* mutant was statistically significant.

3.2.5. Crosses between deletion mutant strains

Mutant strains to be crossed were firstly cultivated and washed as previously described. Secondly were mixed 1:1 and finally distributed in 10 µl drops on the surface of DWR plates. Additionally, direct mixture of strains on DWR plates was also performed. Plates were incubated at 18°C and observed daily (Table II.8, Appendix II).

3.2.6. Bacterial two-hybrid assays

In order to ascertain the possible interaction between Hd1 and Hd2 protein bacterial two-hybrid assays were performed using the Bacterial adenylate cyclase two-hybrid system kit from Euromedex. The cDNA of each of the *HD* genes (Figure II.3, Appendix II) was synthesized at Eurofins and delivered as an insert in the commercial vector pEX-K4. Synthetic *HD* genes were amplified by PCR and sub-cloned into plasmids pKNT25 and pUT18 using the Hind III and Pst I restriction sites present in the multiple cloning site of both of plasmids, according to standard molecular biology techniques and the Euromedex manual (Table II.9 and Table II.10, Appendix II) (Battesti & Bouveret 2012). Genes were cloned in frame at the N-terminus end of the T25 and T18 peptides. After transformation, integrity of the constructs was assessed by PCR amplification and sequencing of the cloned fragments (primers MP191/MP192 and MP193/MP194 (Table II.10, Appendix II). Different combinations of the recombinant plasmids (Figure II.4.a, Appendix II) were co-transformed into BTH101 *E. coli* cells (Table II.1, Appendix II). After successful transformation, the phenotype of 8 clones (identified as 1.1–10.8) of each of the different co-transformations was assessed in X-gal and MacConkey/maltose media upon incubation at 30°C for 48h. Results were scored after 24h (Figure II.4.b and c, Appendix II) and 48h incubation times. The presence of both fragments and correct plasmids in each of the clones was assessed by PCR (Table II.10, Appendix II).

3.2.7. Yeast two-hybrid assays

Matchmaker Gold Yeast Two-Hybrid System from Clontech was also used to assess a possible interaction between the homeodomain proteins of *P. rhodozyma*. Synthetic coding sequence of the *HD1* and *HD2* genes (Figure II.3, Appendix II) were cloned into pGBKT7 and pGADT7 plasmids by transforming each of the PCR fragments and the digested plasmid (Table II.11 and Table II.12, Appendix II) into the pertinent yeast strain (Figure II.5.a, Appendix II). All *S. cerevisiae* transformations were performed according to Yeastmaker Yeast Transformation System 2 User Manual from Clontech. Primers used to amplify the *HD* genes carried 40 bp 5' tails homologous to the ends of the linearized plasmids to promote recombination. Plasmids pGBKT7 and pGADT7 were linearized with Pst I and Cla I, respectively. Two distinct versions of the synthetic coding sequences of the *HD* genes were used, a shorter one comprising the complete N-terminal and homeodomain region of each of the genes (corresponding to the first 183 amino acids of Hd1 protein and the first 196 amino acids of Hd2

protein respectively) and another, comprising the complete proteins. Each of the haploid *S. cerevisiae* strains generated was tested for the ability of the fusion protein expressed to activate any of the reporter genes on their own (Figure II.5.b, Appendix II). Haploid *S. cerevisiae* strains (Figure II.5.c, Appendix II) were them mated and diploid strains were selected as described in the Matchmaker Gold Yeast Two-Hybrid System user manual. Three diploid strains (named X, Y and Z), selected from each mating experiment were tested for their ability to activate the reporter genes. Plates were incubated at 30°C for 72h and photographed daily (Figure II.5.d and e, Appendix II).

3.2.8. Search for potential alternative homeodomain transcription factor proteins

Inspection of all BLASTP hits obtained in *P. rhodozyma* draft genome (CBS 6938) with the *SXI1* and *SXI1* of *C. deneoformans* was performed. The resulting proteins were evaluated in HMMER software (Finn *et al.* 2011) and InterProScan (Mitchell *et al.* 2015) in order to identify their main features and domains. Proteins obtained were also used for BLASTP at the NCBI server to identify corresponding proteins with significant *E*-values that could aid in inferring the role of each of the studied proteins.

3.3. Results

3.3.1. Pheromone receptors are required for sexual reproduction

In heterothallic systems, pheromones and pheromone receptors are usually involved in the process of cell-cell compatibility recognition preceding plasmogamy (Tsuchiya & Fukui 1978; Lee *et al.* 2010). This process might conceivably be dispensable in self-fertile sexual reproduction but was, on the contrary, found to be relevant for homothallic systems in species of the genera *Neurospora* (Nygren *et al.* 2012) and *Sordaria* (Ascomycota) (Mayrhofer *et al.* 2006). To address the question of whether pheromone receptor genes were required for sexual reproduction of *P*. *rhodozyma*, deletion mutants of each of the two receptor genes *STE3-1* and *STE3-2* were produced in turn, using homologous recombination to target chromosomal integration of antibiotic resistance markers to the *PR* locus so as to delete each of the receptor genes, but leaving the pheromone precursor genes intact (Figure 3.2 and Table 3.1).

Table 3.1. List of *P. rhodozyma* strains used and correspondent genetically manipulated derivatives generated in this study.

Strains / Mutants	Relevant genotypes
CBS 6938	*Phaffia rhodozyma* wild type strain
*ste3-1*Δ	*ste3-1Δ::G418*[1]
*ste3-2*Δ	*ste3-2Δ::HYG*[2]
*ste3-1*Δ*ste3-2*Δ	*ste3-2Δ::HYG*[2]*/ste3-1Δ::G418*[1]
*hd1*Δ	*hd1Δ::HYG*[2]
*hd2*Δ	*hd2Δ::G418*[1]
*hd1*Δ*hd2*Δ	*hd1Δ::HYG*[2]*/hd2Δ::G418*[1]
*ste3-1*Δ*mfa1*Δ	*ste3-1/mfa1Δ::G418*[1]
*ste3-2*Δ*mfa2*Δ	*ste3-2/mfa2Δ::G418*[1]
*ste3-1*Δ*mfa2*Δ	*ste3-1Δ::G418*[1]*/mfa2Δ::ZEO*[3]
*ste3-2*Δ*mfa1*Δ	*ste3-2Δ::HYG*[2]*/ mfa1Δ::ZEO*[3]
*ste3-1*Δ*mfa1*Δ*hd1*Δ	*ste3-1/mfa1Δ::G418*[1]*/hd1Δ::HYG*[2]
*ste3-2*Δ*mfa2*Δ*hd2*Δ	*ste3-2/mfa2Δ::G418*[1]*/hd2Δ::ZEO*[3]
*spo11*Δ	*spo11Δ::HYG*[2]
*ste3-1*Δ*ste3-2*Δ*+STE3-1*	*ste3-2Δ::HYG*[2]*/ste3-1Δ::G418*[1]*/STE3-1::ZEO*[3]*+ rDNA*
*ste3-2*Δ*mfa2*Δ*+MFA2*	*ste3-2/mfa2Δ::G418*[1]*/MFA2::ZEO*[3]*+ rDNA*
*hd1*Δ*+HD1*	*hd1Δ::HYG*[2]*/HD1::ZEO*[3]*+ rDNA*
*hd1*Δ*hd2*Δ*+HD1*	*hd1Δ::HYG*[2]*/hd2Δ::G418*[1]*/HD1::ZEO*[3]*+ rDNA*

[1] *Geneticin resistance cassette;* [2] *Hygromycin resistance cassette and* [3] *Zeocin resistance cassette.*

Assessment of the phenotype of the individual pheromone receptor mutants, *ste3-1Δ* and *ste3-2Δ*, in sporulation medium showed that their sporulation capabilities were similar to the wild type, suggesting

that neither receptor is required *per se* for sporulation (Figure 3.2.e). On the contrary, the double mutant, *ste3-1Δste3-2Δ*, lacking both receptors, failed completely to sporulate indicating that the two receptors are functional and redundant. Subsequently the *STE3-1* gene was reintroduced in the double mutant *ste3-1Δste3-2Δ*, through integration in the rDNA locus. As expected, functional complementation restored the capability of the resulting strain *ste3-1Δ ste3-2Δ +STE3-1* to sporulate (Figure 3.2.c).

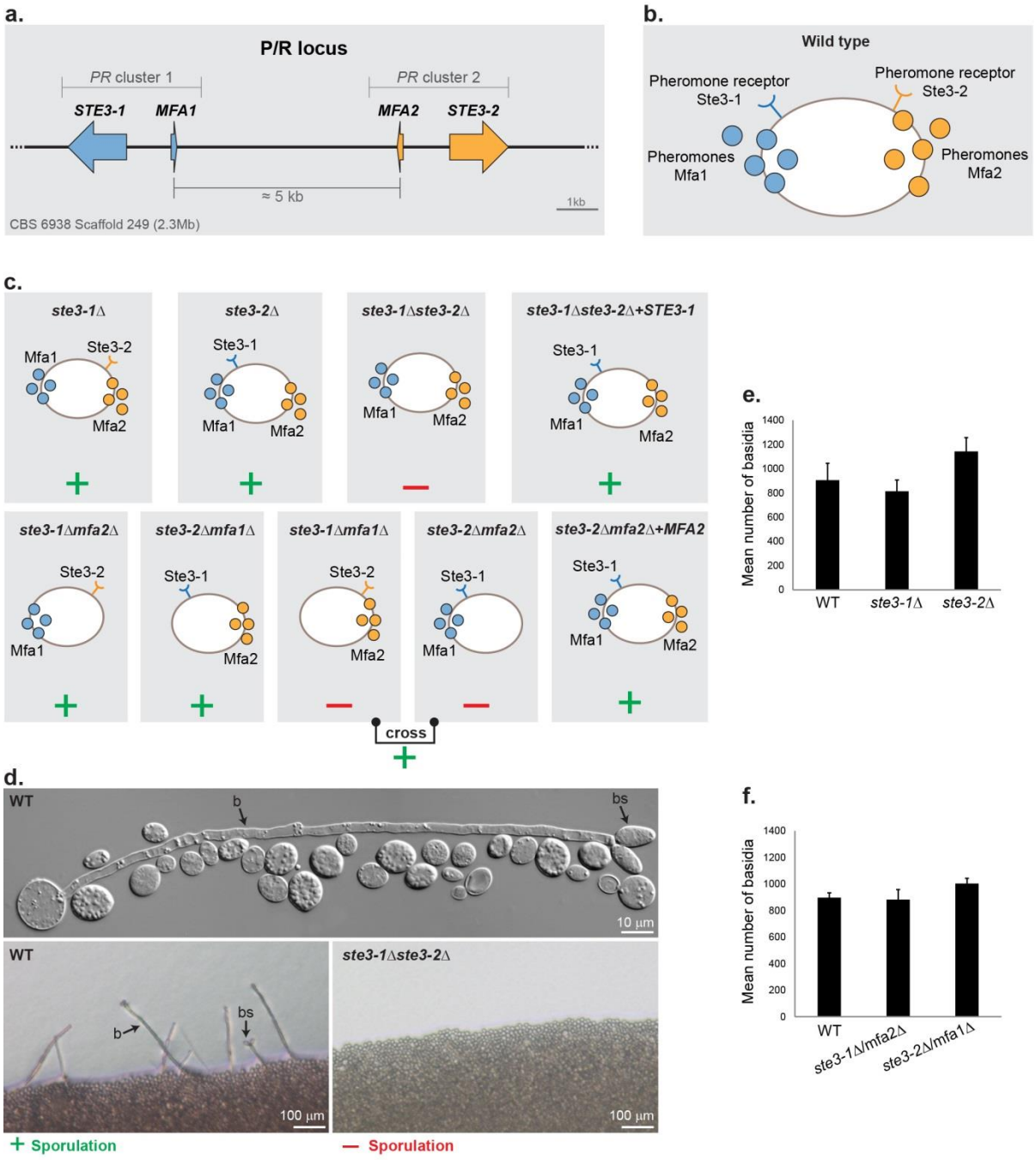


Figure 3.2 Deletion mutants in the *PR* locus. (a) Organization of the *PR* locus of strain CBS 6938 depicting gene clusters *PR1* (blue) and *PR2* (orange). **(b)** Pheromones and pheromone receptors as are expected to be expressed in the wild type cells. **(c)** *PR* locus genes expressed and sporulation phenotypes of mutants in which one or more *PR* locus genes have been deleted. Plus signs (dark green) indicate that formation of basidia and

basidiospores was observed; minus (red) signs denote complete failure of the mutant to sporulate. **(d)** Light microscopy photographs showing an example of positive and negative sporulation phenotypes, as indicated. Letters *b* and *bs* denote basidium and basidiospore, respectively. **(e)** and **(f)** Mean number of basidia observed per plate in sporulation proficient mutants after 10 days of incubation in sporulation medium at 18°C. Error bars represent standard deviations from the mean for the three biological replicates. No significant differences were observed between the WT and the sporulation proficient mutants (Student's t-test).

3.3.2. Evidence for reciprocal compatibility between receptors and pheromones encoded by the two *PR* gene clusters

Genes encoding distinct pheromone precursors (*MFA1* and *MFA2*) flank each of the two pheromone receptor genes (*STE3-1* and *STE3-2*) that are located at a short distance from each other in the genome (Figure 3.2.a). To find out how, if at all, the pheromones encoded in the *PR* clusters interacted with the receptors, two deletion mutants were constructed, *ste3-1Δmfa1* and *ste3-2Δmfa2*. Neither of the resulting strains could sporulate, strongly suggesting that each receptor is activated by the pheromone encoded in the other cluster, i.e. Ste3-1 is for example very likely activated by Mfa2 (Figure 3.2.c). In accordance with this, reintroduction of the *MFA2* gene in the *ste3-2Δ mfa2Δ* mutant complemented its sporulation defect (Figure 3.2.c). Mutants with double deletions were subsequently constructed, *ste3-1Δmfa2Δ* and *ste3-2Δmfa1Δ*, each expressing a distinct, presumably interacting pheromone receptor/pheromone pair. Both double mutants were able to sporulate at normal levels (Figure 3.2.c and f, Table II.7, Appendix II) showing that, as predicted, Mfa1 interacts with Ste3-2 while Mfa2 activates Ste3-1. Finally, when the *ste3-1Δ mfa1Δ* and *ste3-2Δ mfa2Δ* mutants were co-cultured in suitable medium, a low level of sporulation two orders of magnitude lower than wild type was observed (Figure 3.2.c and Table II.8, Appendix II), which is consistent with the functional receptor remaining in each strain being activated by the pheromone produced and secreted by the other strain.

3.3.3. Genes *HD1* and *HD2* are required for sexual development in *P. rhodozyma*

Divergently transcribed candidate *HD1* and *HD2* genes were also identified in the *P. rhodozyma* genomes, resembling the genomic arrangement of homologous genes found in most tetrapolar species (Figure 3.1). To investigate whether the putative *HD1* and *HD2* genes uncovered in the genomes were involved in sexual development, single mutants, *hd1Δ* and *hd2Δ*, in which each of the two genes was deleted in turn, as well as a double mutant *hd1Δhd2Δ*, were constructed (Table 3.1) Sporulation was abolished in the *hd1Δ* single mutant and in the *hd1Δhd2Δ* double mutant, but vestigial sporulation was observed for the *hd2Δ* mutant (Figure 3.3 and Table II.6). Reintroduction of the *HD1* gene in the rDNA of the *hd1Δ* strain restored the wild type sporulation phenotype, confirming that loss of sporulation was truly a consequence of *HD1* deletion (Figure 3.3d). Similarly, ectopic expression of *HD1* in the *hd1Δhd2Δ* mutant restored a limited ability to sporulate (Table II.7) to levels like those observed for the *hd2Δ* mutant (Figure 3.3 and Table II.6). Taken together, these results suggest that sporulation requires both proteins but that the absence of *HD2* does not completely block completion

of the sexual cycle. The most likely explanation for these observations is that the divergently transcribed *HD1* and *HD2* genes in *P. rhodozyma* work together to regulate genes required for sexual development, unlike similarly arranged genes found in tetrapolar species across the Basidiomycota.

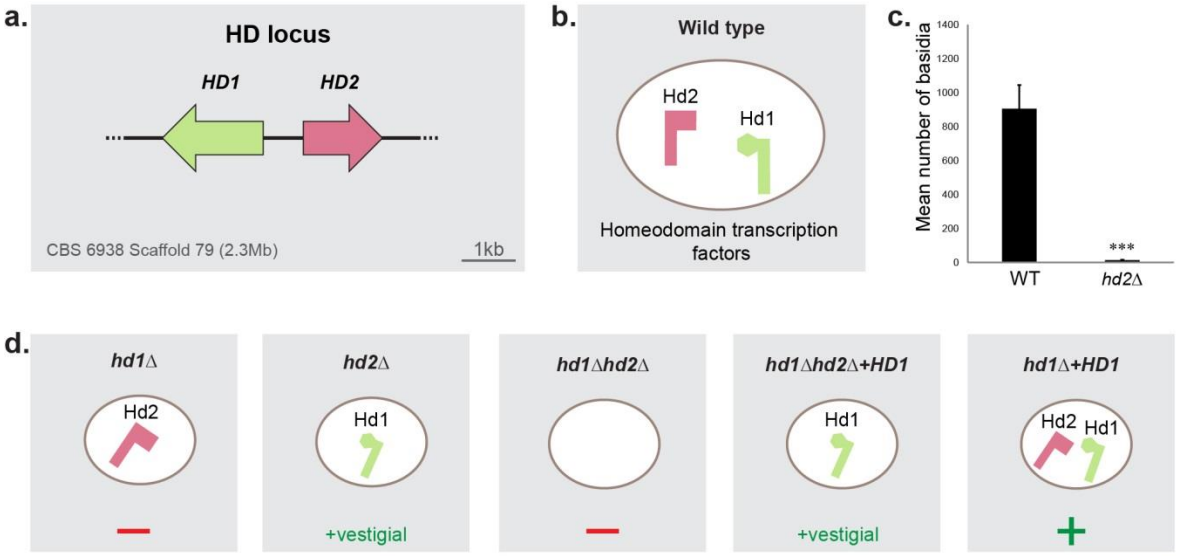


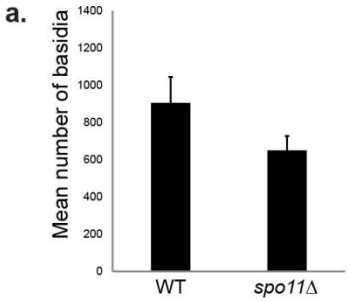
Fig 3.3. Deletion mutants in the *HD* locus. (a) Organization of the *HD* locus of strain CBS 6938 depicting the *HD1* (green) and *HD2* (pink) genes. **(b)** Homeodomain transcription factors as are expected to be expressed in wild type cells. **(c)** *HD* locus genes expressed and sporulation phenotypes of mutants in which one or both *HD* locus genes have been deleted. Plus signs (dark green) indicate that formation of basidia and basidiospores was observed, while minus signs (red) denote complete failure of the mutant to sporulate. Smaller plus sign was used to denote that only vestigial sporulation was observed in mutants expressing only *HD1*. **(d)** Number of basidia observed per plate in sporulation proficient mutants after 10 days of incubation in sporulation medium at 18°C. Error bars represent standard deviations from the mean for three biological replicates. Asterisks denote significant difference between the WT and the *hd2Δ* mutant (Student's t-test, p = 0,0004).

To assess whether the Hd1 and Hd2 proteins might be capable of forming a heterodimer, a bacterial two-hybrid assay was performed (Battesti & Bouveret 2012) to try to detect an interaction between the Hd1 and Hd2 proteins of *P. rhodozyma*. The results, shown in Figure II.4 (Appendix II), failed to demonstrate the occurrence of an interaction between the two proteins sufficiently strong to be detected by this assay. To further verify this apparent absence of interaction, a second assay was performed using the yeast two-hybrid system, which had been previously used to detect interactions between Hd1 and Hd2 proteins (known as *bW* and *bE*) of *U. maydis* (Kämper *et al.* 1995) and *C. deneoformans* (Hull *et al.* 2005). To this end, fusions were constructed mimicking those successfully employed to detect interactions between the *U. maydis* proteins, including fusion proteins comprehending the complete HD proteins as well as shorter versions including solely the N-terminal domains normally involved in dimerization and the homeodomain region (Table II.11 and Table II.12). In line with results obtained for the bacterial two-hybrid system, a clear interaction was not detected using the two possible combinations of the short dimerization domains of *HD1* and *HD2* (Figure II.5) with only one of the four transformants expressing the Hd1 and Hd2 N-terminal domains showing some activation of the *MEL1* reporter gene (Figure II.5.e) and no discernible activation of the remaining two reporter genes. However, a weak interaction signal denoted only by the *MEL1* reporter

gene was consistently detected in all combinations involving fusion proteins that comprehended one complete coding region of either *HD1* or *HD2* (Figure II.5, Appendix II). Taken together, these results could indicate that the two *P. rhodozyma* proteins may interact, albeit weakly. Interestingly, in the yeast assay, we also observed interactions that might support the formation of homodimers (Figure II.5, Appendix II).

3.3.4. Genetic evidence for the involvement of meiosis in the sexual cycle of *P. rhodozyma*

The homothallic sexual cycle of *P. rhodozyma* is considered to involve meiosis (Kucsera *et al.* 1998; Kucsera *et al.* 2000) and close inspection of the draft genomes available (Sharma *et al.* 2015; Bellora *et al.* 2016) revealed the presence of all but one (*MLH2*) of the core genes required to complete meiosis (Bellora *et al.* 2016). Given that the ploidy of different *P. rhodozyma* strains is known to vary (Medwid 1998) and segregation of molecular markers were observed to deviate from a typical Mendelian distribution (Kucsera *et al.* 1998), some doubts existed regarding the occurrence of a typical meiosis. It was therefore relevant to examine the dependence of *P. rhodozyma* sporulation on *SPO11*, a core meiosis gene encoding an endonuclease (Keeney 2008) shown to be required for meiotic recombination in *C. deneoformans* (Feretzaki & Heitman 2013a). To achieve that, a comparison of the ability of a *spo11Δ* mutant to complete the sexual cycle and the viability of the spores produced with that of the wild type strain was performed. The results, shown in Figure 3.4, indicate that sporulation was less efficient in the *spo11Δ* mutant than in the wild type, although this difference did not reach statistical significance. In addition, the viability of F1 spores isolated from the *spo11Δ* mutant was significantly lower than that of the wild type. These results suggest that meiosis is indeed part of the sexual life cycle of *P. rhodozyma.*



b.

Strain	Basidiospores dissected	Basidiospores germinated	Viability (%)	
WT	41	32	78.0	***
***spo11*Δ**	64	27	42.2	

Fig 3.4. *SPO11* deletion mutant. (a) Number of basidia observed per plate in sporulation proficient mutants after 10 days of incubation in sporulation medium at 18°C. Error bars represent standard deviations from the mean for the three biological replicates. No significant differences were observed between the sporulation ability of the WT and *spo11Δ* mutant. **(b)** Total number of basidiospores recovered from WT and *spo11Δ* mutant and fraction of viable spores (asterisks denote significant difference, *P*-value = 0.000302, *Chi-square* statistic).

3.3.5. Construction of heterothallic strains

The results described in the previous sections show that the genetic determinants and mechanisms involved in the homothallic life cycle of *P*. *rhodozyma* are similar to those of heterothallic, tetrapolar basidiomycetes. From the point of view of strain improvement of *P*. *rhodozyma* for biotechnological applications, it would be very useful to be able to generate strains in which outcrossing is strongly favored, because it would facilitate selection of strains harboring desirable combinations of characteristics using a selective mating approach. To try to generate obligate outcrossing *P*. *rhodozyma* strains, mutants with artificial "mating-types" were generated harboring complementary components of the two mating recognition systems, *ste3-1Δmfa1Δhd1Δ* and *ste3-2Δ mfa2Δhd2Δ* (Table 3.1). Given the previous results, a cross between such triple mutants should almost exclusively yield spores resulting from conjugation between independent cells belonging to complementary "mating-types", while formation of basidia originating from single cells or from pedogamy would almost completely be prevented by the absence of complete *HD1/HD2* gene pair in each of the strains with the different "mating-types". Hence, while in a cross between *ste3-1Δ mfa1Δ* and *ste3-2Δ mfa2Δ* mutants, extracellular diffusion of the pheromones would permit any of the three possible modes for formation of basidia (Figure 1.4, Chapter 1) because each mutant possessed a complete *HD1/HD2* pair, in the triple mutant cross, sporulation would be expected to occur almost exclusively after cell fusion joined together the Hd1 and Hd2 partners originating from different "mating-types". Indeed, in a cross between the strains with the artificially created "mating types", basidia with basidiospores were formed, albeit at levels three orders of magnitude lower than wild type and two orders of magnitude lower than the cross between *ste3-1Δ mfa1Δ* and *ste3-2Δ mfa2Δ* (Table II.8, Appendix II).

3.4. Discussion

The construction of several deletion mutants allowed the characterization in further detail of the *MAT* loci of *P*. *rhodozyma*. Primary homothallism has been observed in a considerable number of species across the entire fungal kingdom (Lin & Heitman 2007; Wilson *et al.* 2015b), sometimes in a few strains of an otherwise heterothallic species (Coelho *et al.* 2011; Maia *et al.* 2015) or, alternatively in lineages consisting mainly of species formed entirely of homothallic individuals with a few heterothallic species, as observed in *Aspergilli* (Wilson *et al.* 2015b). In a number of cases, like for *A. nidulans*, it was possible to demonstrate that homothallism was associated to the presence in one individual of the entire complement of genetic information normally present in the two opposite mating types of heterothallic individuals of closely related species (Paoletti *et al.* 2007). Likewise, the purposeful introduction of *MAT* genes of one mating type into the opposite mating type resulted in the rise of a homothallic phenotype (Hull & Heitman 2002). Nevertheless, a thorough explanation of how these genes interact to produce the homothallic phenotype is still lacking in most cases (Lin & Heitman 2007; Kues & Navarro-Gonzalez 2010; Lee *et al.* 2010; Wilson *et al.* 2015b). In *P. rhodozyma* both receptors are functional and are activated by the pheromone encoded in the other cluster. It

would appear that the most likely setting for these pheromone/receptor specificities to have evolved would be a heterothallic system, which supports that an ancestor of the *Phaffia* lineage was likely heterothallic.

Homothallism is thought to be rather uncommon in Basidiomycetes (Whitehouse 1949), especially primary homothallism (Lemke 1969; Koltin *et al.* 1972), which is in line with the complex genetic underpinnings of the mating system in this phylum, involving two independent non-self-recognition checkpoints. Hence, a transition from heterothallism to primary homothallism would require a genomic rearrangement gathering in the same genome two compatible versions of both the *PR* and *HD* loci, or that one or both loci would become self-compatible (Lin & Heitman 2007). It seems more likely that homothallism in *P. rhodozyma* is derived from a tetrapolar ancestor, rather than bipolar, because the *HD* and *PR* loci are apparently genetically unlinked (Chapter 2) and are both required for sporulation (Figure 3.2 and Figure 3.3). Yet, the extant *MAT* locus in *P. rhodozyma* is not simply a gathering of all *MAT* genes normally present in two compatible mating types in tetrapolar systems because only one *HD1/HD2* gene pair is present, encoding proteins that are both required to promote sexual development, as shown by the absence of sporulation in the *hd1Δ* and by a dramatic drop in sporulation (approximately 2% of wild type sporulation levels remaining) observed for the *hd2Δ* mutant. In heterothallic systems, HD proteins form heterodimers in which the interacting partners are encoded by different mating types, thereby enforcing mating between genetically distinct individuals (outcrossing). The domain structures of HD proteins and the molecular interactions they are likely to undergo have been examined in detail in *U. maydis* (Kämper *et al.* 1995; Romeis *et al.* 1997; Schlesinger *et al.* 1997; Yee & Kronstad 1998), *C. cinerea* (Asante-Owusu *et al.* 1996), *S. commune* (Robertson *et al.* 2002) and *C. deneoformans* (Hull *et al.* 2005). These studies showed that the protein domains required for proper functioning of the heterodimeric transcription factor, such as high affinity DNA binding, nuclear localization signal (NLS) and transcriptional activation were not present in a single protein. For example, the *C. cinerea* Hd1 protein harbors a transcriptional activation domain and a NLS but its homeodomain is dispensable for sexual development, while its Hd2 counterpart possesses no NLS but its homeodomain is absolutely required for DNA binding by the heterodimer (Spit *et al.* 1998). In this manner, undimerized partners are doomed to be unsuccessful as transcription factors: Hd2 is unable to get transported into the nucleus on its own while Hd1 can get transported into the nucleus due to its NLS, but will fail to effectively bind DNA (Spit *et al.* 1998). Conversely, in the Hd2 proteins of *Heterobasidion*, no *bona fide* homeodomain could be found, so that DNA binding probably relies entirely on the homeodomain of the Hd1 protein (van Diepen *et al.* 2013). Interestingly, in *P. rhodozyma, a* weak interaction, at most, was detected between the two HD proteins. This is in contrast to the results obtained when *U. maydis* (Kämper *et al.* 1995) and *C. deneoformans* (Hull *et al.* 2005) HD proteins were examined in a yeast two-hybrid assay. However, phenotypes of the various deletion mutants showed clearly that HD1 and HD2 are involved in the regulation of an overlapping set of genes essential for sporulation. Taking into account the well-established mode of operation of model species within the Agaricomycotina, this leaves room for two possible interpretations: i) the two *P. rhodozyma* HD proteins do not form heterodimers, as normally observed for proteins encoded by the same locus, and hence, they have independent contributions to

regulate a set of genes essential for sexual development; ii) the two proteins interact to promote transcriptional regulation of genes essential for sporulation, but the interaction is much weaker than those previously characterized in heterothallic species. The obtained results appear to favor the latter possibility for a number of reasons. Firstly, it has been shown for mutant alleles in *U. maydis* that they can promote sexual development in vivo despite their failure to interact detectably in the yeast two-hybrid assay (Kämper *et al.* 1995). Secondly, it was found that Hd1 alone supported the ability to sporulate (mutant *hd2Δ*), albeit at low levels (Figure 3.3 and Table II.6, Appendix II). This is consistent with, for example, a scenario in which Hd1 is dependent on Hd2 for high affinity DNA-binding, but can bind independently sufficiently well to support a low level of sporulation in the absence of the Hd2 partner. In line with this, a prominent DNA binding role has been ascribed to the Hd2 proteins of other members of the Agaricomycotina, such as *C. cinerea* (Spit *et al.* 1998), *S. commune* (Luo *et al.* 1994) and *C. deneoformans* (Stanton *et al.* 2009). The third reason in favor of a weak interaction between the two HD proteins pertains to how self-compatibility of the *P*. *rhodozyma* HD protein pair, if it exists, may have evolved. For this, two possibilities may be considered. Firstly, a *HD* gene pair encoding self-compatible proteins could conceivably form through recombination between two distinct heterothallic ancestor alleles, in which case the interaction between the two proteins would probably be expected to be sufficiently strong to be unequivocally detected in the two hybrid assays, as observed for the large majority of naturally occurring *HD* alleles in the heterothallic species examined (Kämper *et al.* 1995; Hull *et al.* 2005). However, in face of previous findings (Kämper *et al.* 1995), a very likely second evolutionary path to generate a compatible *HD1/HD2* pair would be the emergence of one or more mutations relieving the structural hindrance normally preventing interaction between HD proteins encoded by the same locus. It has been previously shown that a single amino acid mutation may be sufficient to remove the obstacle for self-dimerization (Kämper *et al.* 1995). In that case, a weak interaction permitting for example cooperative DNA binding would probably suffice for normal function provided the two proteins can reach the nucleus independently. In fact, Hd1 is apparently able to reach the nucleus independently of Hd2, because it was capable of promoting some sporulation in the absence of Hd2, although inspection of the sequences of both proteins revealed that only Hd2 possessed a candidate NLS (Chapter 2). The idea of homodimerization of the HD proteins in *P. rhodozyma* was also considered. However in basidiomycetes, no biological roles have been ascribed so far to homodimers of HD proteins, as opposed to what was described for *S. cerevisiae* (Mukai *et al.* 1997) where homodimers of homeodomain proteins exist and have a well-defined function. In the absence of an unequivocal experimental demonstration of an interaction, the establishment of a definite mode of action for the HD proteins will have to await a detailed dissection of the functional domains present in each of the two proteins and the identification of their DNA binding sites, as previously accomplished in other systems (Luo *et al.* 1994; Kämper *et al.* 1995; Romeis *et al.* 1997; Schlesinger *et al.* 1997; Yee & Kronstad 1998; Badrane & May 1999; Hull *et al.* 2005; Stanton *et al.* 2009). While contemplating the several possibilities to reconcile the lack of a strong interaction between the Hd1 and Hd2 proteins of *P*. *rhodozyma* with the phenotypes exhibited by the various mutants, the prospect that Hd1 might engage an alternative dimerization partner in the absence of HD2 was considered. However, close inspection of *P*. *rhodozyma* (CBS 6938) draft genome, failed to

detect genes encoding homeodomain proteins with the appropriate domain architecture (Table II.13, Appendix II). Hence, this possibility was considered very unlikely.

The genetic arrangement of the two *PR* clusters is strongly reminiscent of *PR* clusters in many basidiomycetes, each functional receptor gene being in the vicinity of a gene encoding a functional pheromone that nevertheless fails to activate its receptor counterpart in the same cluster or locus (Riquelme *et al.* 2005; Fraser *et al.* 2007). It can therefore be hypothesized that this arrangement resulted from the fusion of two heterothallic loci.

Based on the overall obtained results and considering available information for other species in the Agaricomycotina, a model (Figure 3.5) is proposed summarizing the most likely roles of the six *MAT* genes in the sexual cycle of *P*. *rhodozyma*.

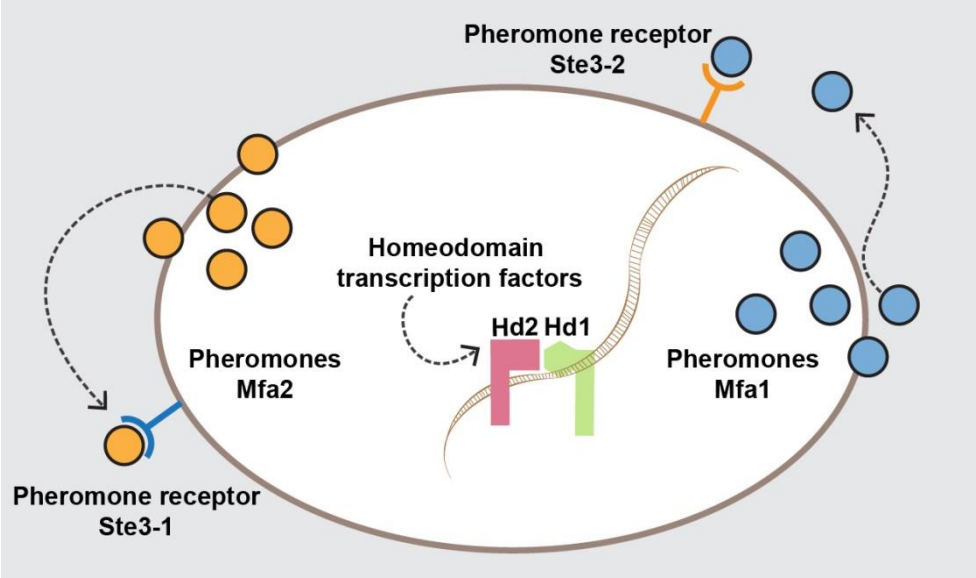


Fig 3.5. Model describing the molecular interactions required for sexual reproduction. Pheromones and receptors encoded in the *PR* locus were both shown to be reciprocally compatible, forming two functionally redundant pairs. The Hd1 and Hd2 proteins encoded at the *HD* locus are proposed to contribute synergistically for sexual development, possibly through a weak interaction allowing them to form a stable heterodimer upon DNA binding.

Considering this model, it was hypothesized that it would be possible to generate artificial heterothallic "mating types" in *P*. *rhodozyma*, by constructing strains containing one *PR* cluster and one *HD* gene complementary to those present in the other "mating type". This was confirmed in successful heterothallic crosses between the *ste3-1Δ mfa1Δ hd1Δ* and *ste3-2Δ mfa2Δ hd2Δ*. The possibility of outcrossing in *P*. *rhodozyma* had been previously put forward based on the observation of fusion between independent cells and on genetic evidence, including strains that seem to be hybrids between the known *P*. *rhodozyma* populations (Kucsera *et al.* 1998; David-Palma *et al.* 2016). However, in line with previous observations (David-Palma *et al.* 2016), results indicate that outcrossing is probably infrequent, since sporulation is greatly diminished when the possibility of selfing is genetically prevented (Table II.8, Appendix II). Nevertheless, the triple mutant crosses open a new avenue to facilitate improvement of *P*. *rhodozyma* strains for production of astaxanthin or other

relevant molecules, for example by combining industrially attractive features like growth at higher temperatures with genetic alterations that improve astaxanthin production.

Additional evidence that *P*. *rhodozyma* probably undergoes meiosis during its life cycle, was also gathered since deletion of a core meiosis gene shown to be required for meiosis in *C*. *deneoformans* (Feretzaki & Heitman 2013a) the model organism most closely related to *P*. *rhodozyma*, affected sporulation efficiency and significantly decreased spore viability. It is therefore unlikely that in *P*. *rhodozyma* ploidy changes are achieved by distinct mechanisms as postulated to occur in the parasexual cycles of other fungi, like the yeast *C*. *albicans* (Bennett & Johnson 2003).

The fact that the species in which this study was performed, is considered one of the most promising microbial source of the carotenoid astaxanthin, may create the possibility of further strain enhancement through selective breeding of strains with genetically created mating types bearers of desirable traits.

CHAPTER 4

Two new species in the genus *Phaffia* - *Phaffia novazelandica* sp. nov. and *Phaffia tasmanica* sp. nov.

4.1. Introduction

Currently, a single species is recognized in the genus *Phaffia*, namely *Phaffia rhodozyma* (Miller *et al.* 1976). This genus belongs to the order Cystofilobasidiales (Tremellomycetes, Agaricomycotina). *Phaffia* was originally found in association with exudates of deciduous trees native to certain northern regions of the northern hemisphere. It was for long considered that *Phaffia* was confined to that ecological niche and region (Phaff 1972; Golubev *et al.* 1977; Golubev 1995; Fell & Blatt 1999). An important expansion of the known genetic diversity of *P. rhodozyma* started in 2007 when this species was detected in the Southern Hemisphere in a new habitat, the stromata of *Cyttaria* spp., an ascomycetous biotrophic parasite of southern beech trees (*Nothofagus* spp.) (Libkind *et al.* 2008; Libkind *et al.* 2011b). The discovery of this new *Phaffia* habitat led to the hypothesis that the distribution range of this yeast could be even broader (Libkind *et al.* 2007). Presuming there was indeed an association between *P. rhodozyma* and the *Nothofagus*-*Cyttaria* system, and since the extant distribution of *Nothofagus* included also Australasia (Peterson & Pfister 2010), the existence of additional *P. rhodozyma* populations in Australasia was hypothesized (Libkind *et al.* 2007). Indeed, subsequent field work carried out in Queensland, Tasmania (Australia) and in New Zealand's South island lead to the detection of an unprecedented diversity of *Phaffia* yeasts (David-Palma *et al.* 2014). Multilocus sequencing (MLS) revealed four main lineages of *P. rhodozyma* (Figure 1.3, Chapter 1) and a signal for niche-dependent assortment of genotypes that was stronger than that of geographic partitioning (David-Palma *et al.* 2014). Populations A and B, from South American and Australasian *Nothofagus*, respectively, were phylogenetically related. Clade C grouped European, Asian and North American strains from *Betula*, whereas the most divergent group was population D from *Cornus* (Figure 1.3, Chapter 1). Furthermore, this sampling effort in Australasia yielded two additional *Nothofagus*-associated *Phaffia*-like lineages that were considered to represent new species (*Phaffia* sp.I and *Phaffia* sp.II) (David-Palma *et al.* 2014). Although a formal taxonomic description was not provided at that time, the phylogenetic inferences obtained using an MLS dataset (Figure 1.3, Chapter 1) composed by seven different genes and also the ITS region (Figure 4.1) corroborated the placement of the two lineages in the genus *Phaffia* (David-Palma *et al.* 2014). At the time these results were published, phylogenetic analysis using rDNA sequences had shown that *Cystofilobasidium* represented the sister genus of *Phaffia* (Scorzetti *et al.* 2002; Boekhout *et al.* 2011; Kurtzman *et al.* 2011a). Newer phylogenetic studies using a larger dataset for phylogenetic inference, indicated the genus *Krasilnikovozyma* as a sister group of *Phaffia* (Figure 4.2), including *Phaffia* in a new family, *Mrakiaceae*, that excludes the *Cystofilobasidium* genus. However, these phylogenetic relations were poorly supported (Liu *et al.* 2015a; Liu *et al.* 2015b). Aiming at the formal description of the two Australasian *Phaffia*-like lineages and its correct placement within Cystofilobasidiales, a broader phylogenetic study based on draft genome sequences of representatives of *Phaffia* and other taxa within the Cystofilobasidiales was performed, additionally to conventional yeast taxonomy analyses. Moreover, as the two lineages were previously reported to have a homothallic life cycle (David-Palma *et al.* 2014), search for *MAT* genes was also performed. Furthermore, given the

biotechnological relevance of this genus, the presence of genes pertaining to the biosynthetic route of astaxanthin was also investigated.

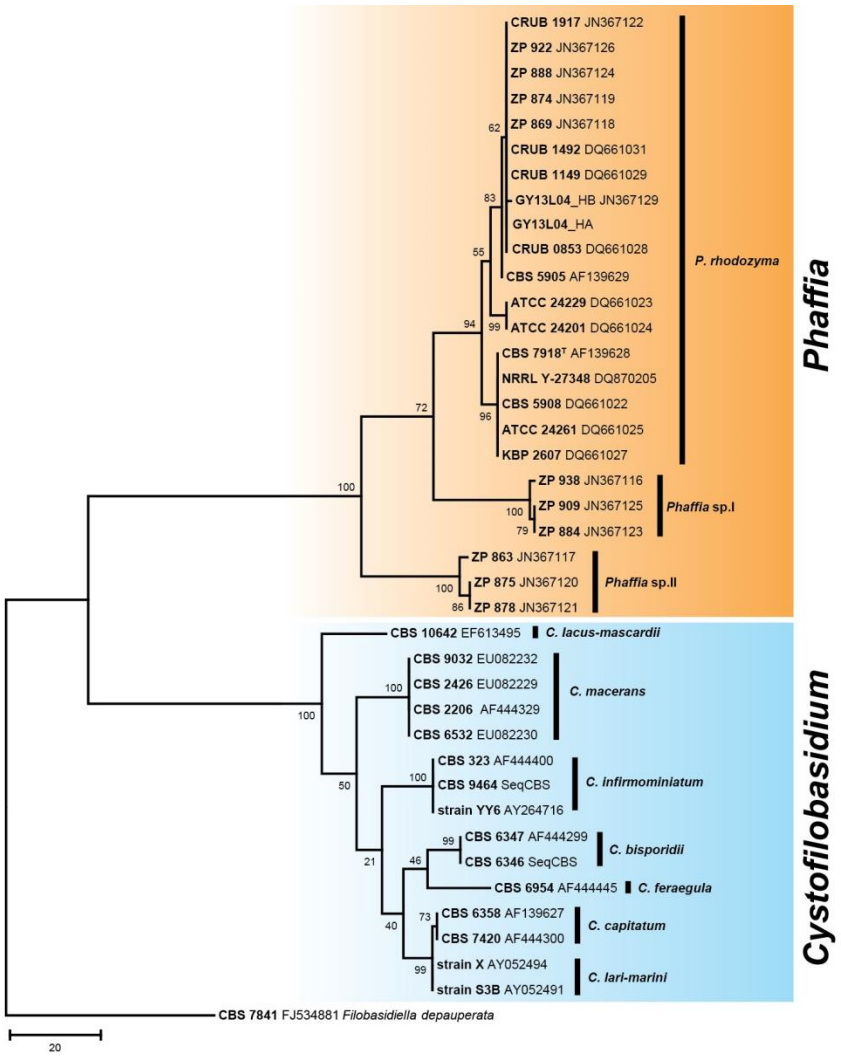


Figure 4.1. ITS phylogeny inferred with maximum likelihood for *Phaffia* and *Cystofilobasidium* species. Accession number of each of the sequences used are presented in the tree after strain number. (Adapted from David-Palma et al., 2014).

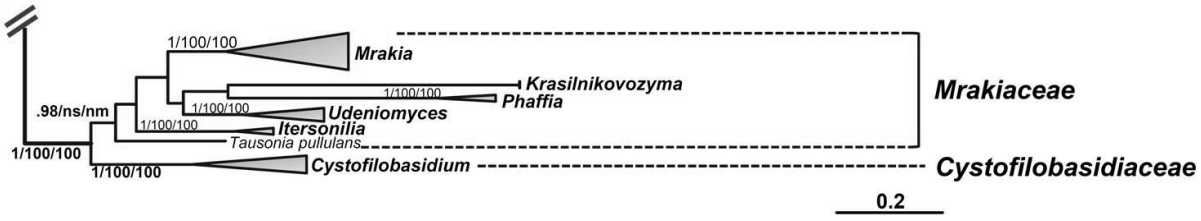


Figure 4.2. Detail of the Tremellomycetes phylogenetic tree based on a seven genes dataset. Branch lengths are scaled in terms of expected numbers of nucleotide substitutions per site. The Bayesian posterior probabilities (PP) and bootstrap percentages (BP) of maximum likelihood and neighbour-joining analyses from 1000 replicates are shown respectively from left to right on branches resolved. Note: ns, not supported (PP < 0.9 or BP < 50 %); nm, not monophyletic. (Adapted from Liu et al., 2015).

4.2. Materials and methods

4.2.1. Phenotypical characterization

Morphological characteristics of the cultures were assessed in YPD (10 g/L Yeast Extract, 20 g/L Peptone and 20 g/L glucose), in Corn Meal Agar (Difco) media and in DWR (2.5% agar and 0.5% ribitol) solid media incubated at 18°C and observed regularly (Golubev 1995; Kucsera *et al.* 1998). Microscopic observations were made using a microscope (Leica DMR) equipped with brightfield and differential interference contrast optics, and microphotographs were recorded using a Leica DFC320 digital camera while dissection of basidiospores was performed using a Zeiss Scope.A1 Micromanipulator. Physiological and biochemical characteristics were examined according to standard protocols (Kurtzman *et al.* 2011b) and were performed in triplicate for the three isolates chosen from each new species (Table III.1, Appendix III).

4.2.2. Genome sequencing

Genomic DNA was extracted from single cell derived cultures using the ZR Fungal/Bacterial DNA MiniPrep kit from ZYMO RESEARCH (Table 4.1). Purity check of the gDNAs was performed on Nanodrop ND-1000 and quantification was performed in Qubit 2.0. Libraries preparation (Nextera kit) and sequencing for 2 × 300 cycles (paired-end short reads), using the Illumina MiSeq technology, was performed by a commercial provider.

Table 4.1. List of strains selected for whole genome sequencing.

Species	Strain number
Phaffia novazelandica	CBS 14095^{T} (ZP 938)
Phaffia tasmanica	CBS 14096^{T} (ZP 875)
Cystofilobasidium capitatum	CBS 7420 (PYCC 4530)
Cystofilobasidium bisporidii	CBS 6347 (PYCC 5604)
Cystofilobasidium ferigula	CBS 7201 (PYCC 4410)
Cystofilobasidium macerans	CBS 6532
Krasilnikovozyma huempii	PYCC 5836

4.2.3. Genome assembly, prediction of coding regions and search of relevant genes

Illumina reads obtained from the strains indicated in Table 4.1. were pre-processed with Trimmomatic (v0.32) (Bolger *et al.* 2014) to remove adaptors and low-quality bases below a defined threshold (Phred score 20), using a sliding-window approach. In addition, Illumina reads (mate-pairs and paired-ends) for *T. pullulans* (JCM 9886), *T. pamirica* (JCM 10408), *M. aquatica* (JCM 1775), *M. frigida* (JCM7857), *U. megalosporus* (JCM 5269) and *U. pyricola* (JCM 2958) (Table 4.2.) were obtained from

the RIKEN BioResource Center repository (http://www.jcm.riken.jp/cgi-bin/nbrp/nbrp_list.cgi) and pre-processed as above, except for the mate pair reads which were trimmed with NxTrim (v0.4.0) (O'Connell *et al.* 2015) using the default parameters.

Table 4.2. List of draft genomes from species of Cystofilobasidiales.

Species	Strain	Origin of the data	Project
P. rhodozyma	CBS 6938	NCBI	PRJEB6925
P. novazelandica	CBS 14095^T	This study	PRJNA371751
P. tasmanica	CBS 14096^T	This study	PRJNA371754
C. capitatum	CBS 7420	This study	PRJNA371774
C. bisporidii	PYCC 5604	This study	PRJNA371778
C. ferigula	PYCC 4410	This study	PRJNA371786
C. macerans	CBS 6532^T	This study	PRJNA371809
M. blollopis	SK-4	NCBI	PRJDB3253
M. aquatica	JCM 1775	RIKEN BioResource Center and RIKEN Center for Life Science Technologies through the Genome Information Upgrading Program of the National Bio-Resource Project of the MEXT, Japan	PRJDB3647
T. pamirica	JCM 10408^T		PRJDB3689
T. pullulans	JCM 9886^T		PRJDB3678
U. megalosporus	JCM 5269^T		PRJDB3720
U. pyricola	JCM 2958^T		PRJDB3672
M. frigida	JCM 7857^T (CBS 5270)		PRJDB3713
K. huempii	PYCC 5836	This study	PRJNA371818

Assemblies were performed with SPAdes (v3.1.1) with two sets of parameters. For the strains indicated in Table 4.1, the parameters "--careful" and "-k 47,57,65,77,81" were used, whereas for the other strains the k-mer values were adjusted to 21, 33, 55, 77 to account for a smaller read size. Quality metrics for all assembled genomes were obtained with QUAST (v3.2) (Table III.2, Appendix III). Coding regions were predicted with Augustus (v.3.2.1) with the training annotation files of *C. deneoformans* (Stanke *et al.* 2008) after which, protein databases were created for each of the genomes. BLASTP was performed against those protein databases using as query the proteins sequences of six RNA polymerase subunits (Rpa1, Rpa2, Ppb1, Rpb2, Rpc1 and Rpc2) from *P. rhodozyma* (CDZ97905, CDZ97151, CED83722, CDZ96772, CDZ98872 and CED85369). In the *P. novazelandica* and *P. tasmanica* draft genomes the search for genes pertaining to the production of astaxanthin (*IDI*, *CRTE*, *CRTYB*, *CRTI*, *CRTS* and *CRTR*) and *MAT* genes was performed by TBLASTN using as query the respective proteins from *P. rhodozyma* (CAA75796.1, AAY33922.1, AAY33923.1, CAA75240.1, ABA43719.1, ACI43097.1 and CED85384.1, CED85379.1, CDZ96688.1, CDZ96689.1 respectively). Scaffolds encoding *MAT* genes were compared with that of *P. rhodozyma.*

Putative orthologs were named according to the protein accession number of *P. rhodozyma* (CBS 6938). Synteny conservation across species was assessed manually based on the predicted annotations and confirmed by high-scoring BLASTP hits in GenBank. Secondary features of the predicted protein sequences of the pheromone receptors and the homeodomain transcription factors were inferred using the same methods as described in Chapter 2 (section 2.2.1). Predicted pheromone precursor proteins sequences from all three *Phaffia* species were aligned with ClustalW as implemented in BioEdit (Hall 1999) in order to predict the likely location of the N-terminal residue of the mature pheromone.

4.2.4. Phylogenetic analysis

Phylogenetic relationships between the different Cystofilobasidiales species were inferred using a concatenated dataset composed of the amino acid sequences of six RNA polymerase subunits (Rpa1, Rpa2, Ppb1, Rpb2, Rpc1 and Rpc2), using as outgroup the respective sequences from *C. deneoformans* (strain JEC21). Proteins were aligned with MAFFT (v7.221) (Katoh & Standley 2013) using the L-INS-I strategy and trimming of the obtained alignments was performed with trimAl (v1.3) ("gappyout" option) (Capella-Gutierrez *et al.* 2009). Concatenation of the six alignments was performed with ElConcatenero (v2.0.4) (https://github.com/ODiogoSilva/ElConcatenero). The final dataset was composed of 16 species sequences with 7879 residues. Phylogeny was inferred with IQ-TREE software (v1.4.3) (Nguyen *et al.* 2015) using the best model (LG+F+G4) according to AICc. Both ultrafast (Minh *et al.* 2013) and non-parametric bootstrapping was performed with 10000 and 1000 replicates respectively. Branch support test (SH-like aLRT) (Guindon *et al.* 2010) was also performed with 10000 replicates. Phylogenetic relationships between the different Mfa proteins predicted in all *Phaffia* species were inferred using as outgroup the sequences from *C. deneoformans* MFa1 (AAG42766.1) and MFα1 (XP_570122.1). Final dataset was composed by 35 residues having all positions containing gaps or missing data been eliminated. Phylogenetic inference was carried out by Neighbor Joining using JTT matrix-based method in MEGA 5.1 (Tamura *et al.* 2011) with 1000 replicates. Using an identical methodology, the phylogenetic analysis of the receptors proteins predicted of all three *Phaffia* species was performed, with a final dataset of 366 positions, using as outgroup the sequences from *C. deneoformans* Ste3a (AAN75624.1) and Ste3α (XP_570116.1).

4.3. Results

4.3.1. Morphological and physiological characteristics of *P. novazelandica* and *P. tasmanica*

Several isolates from the *Nothofagus*-*Cyttaria* system collected in Australasia (Figure 4.3) were shown by phylogenetic analysis to form two divergent lineages, hypothesized to be distinct species from *P. rhodozyma*. Three strains from each lineage (Table 4.3) were evaluated using classical physiological and morphological tests. For the new species *Phaffia novazelandica*, strain ZP 938 was chosen as the type strain and was deposited in the Portuguese Yeast Culture Collection (PYCC 6859^{T}) and in the Centraalbureau voor Schimmelcultures (CBS-KNAW) (CBS 14095^{T}) culture collection. Sequences for the LSU and ITS regions were made publicly available in Genbank, with the accession numbers KR108929 and JN637116 respectively. Similarly, for species *Phaffia tasmanica*, strain ZP 875 was chosen as the type strain and was deposited in both culture collections (PYCC 6858^{T} and CBS 14096^{T}). Sequences for the LSU and ITS regions were also deposited in Genbank, with the accession numbers KT223097 and JN637120 respectively.

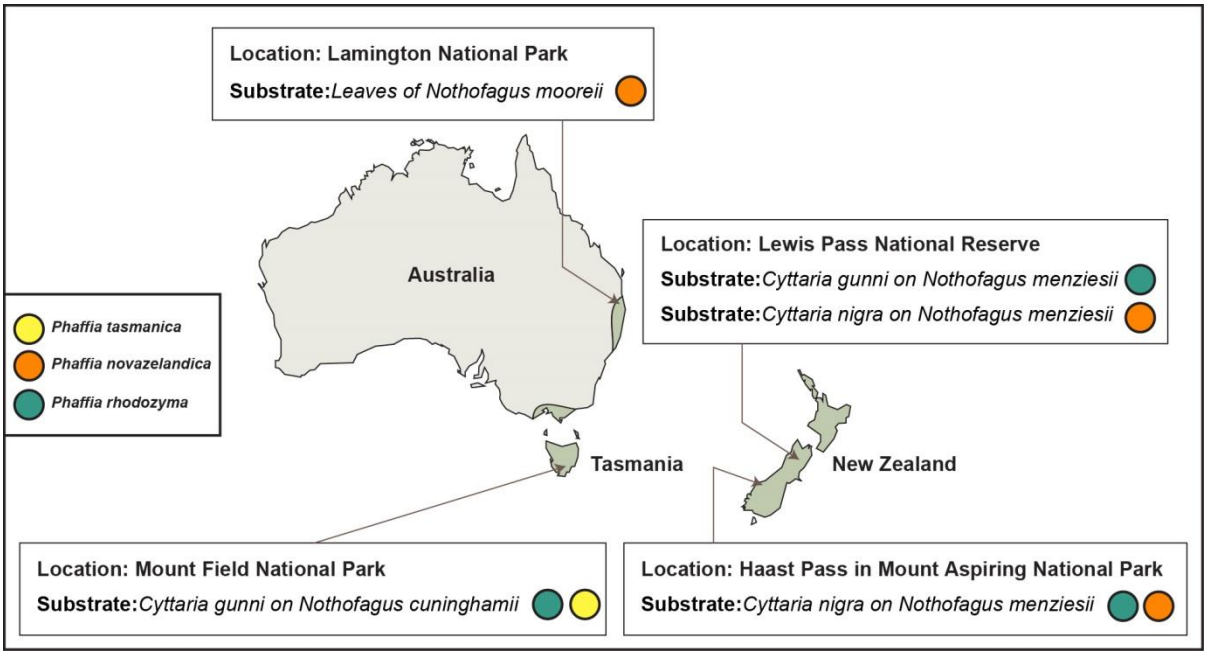


Figure 4.3. Australasia and the collection sites of *Phaffia* spp. isolates. Australasian map showing present-day Nothofagus distribution (darker green) and collection localities (arrows).

Table 4.3. List of strains used in phenotypic tests.

Species	Strain	Collection locality	Substrate
Phaffia novazelandica	ZP 938 (CBS 14095^T)	Lamington Park, Queensland, Australia	Leaves of *N. mooreii*
	ZP 909	Haast Pass, New Zealand	*C.nigra* on *N. menziesii*
	ZP 884	Lewis Pass, New Zealand	*C.nigra* on *N.menziesii*
Phaffia tasmanica	ZP 863	Mount Field National Park, Tasmania, Australia	*C.gunnii* on *N. cunninghamii*
	ZP 878	Mount Field National Park, Tasmania, Australia	*C. gunnii* on *N.cunninghamii*
	ZP 875 (CBS 14096^T)	Mount Field National Park, Tasmania, Australia	*C.gunnii* on *N. cunninghamii*

The two new *Phaffia* species were capable of fermenting glucose and sucrose, a characteristic shared with *P. rhodozyma*. Additionally, *P. tasmanica* is also capable of weakly fermenting raffinose, melezitose and starch. *Phaffia novazelandica* differs from the other two *Phaffia* species for being able to grow on xylitol and ethylamine as sole carbon and energy source (Table III.1, Appendix III). Maximum growth temperatures (MGT) were determined (Table 4.4). Strain CBS 14096^T from *P. tasmanica* presented a MGT of 29ºC, while the average MGT for this species is 27.3ºC. The type strain of *P. novazelandica* (CBS 14095^T) grows up to 27ºC while the average MGT for the species is 25.6ºC. These results are similar to the MGT of *P. rhodozyma* which is 27°C (Miller *et al.* 1976).

Table 4.4. Maximum growth temperatures of *P. novazelandica* and *P. tasmanica*.

Species	Strain	Maximum growth temperature
Phaffia novazelandica	ZP 938 (CBS 14095^T)	27°C
	ZP 909	25°C
	ZP 884	25°C
Phaffia tasmanica	ZP 863	26°C
	ZP 878	27°C
	ZP 875 (CBS 14096^T)	29°C

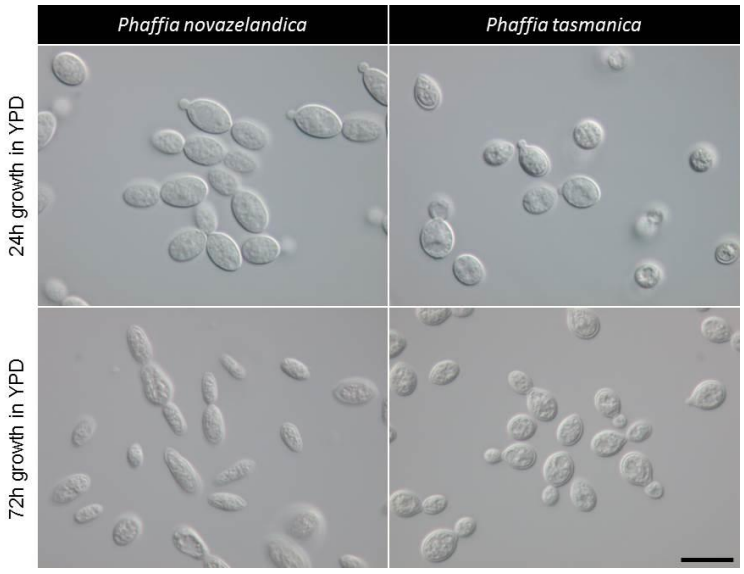


Figure 4.4. Vegetative cells of *P. novazelandica* (CBS 14095^T) and *P. tasmanica* (CBS 14096^T). Strains were grown in liquid YPD media at 20°C. Scale bar = 10 µm.

Cultivated in YM or YPD liquid media for 3 days at 20ºC, *P. novazelandica* presented elongated ellipsoidal cells (4-8×8-16 µm in size) that became more elongate with increased incubation time (7 days). *Phaffia tasmanica* presented ellipsoidal cells (4-7×7-8 µm in size) similar to those of *P. rhodozyma* when cultivated in rich media (Figure 4.4). Vegetative cells from both species multiplied by enteroblastic budding and neither presented true hyphae although pseudohyphae occurred in most solid media. Sexual reproduction structures were characterized *in P. rhodozyma* by the formation of a slender aerial basidium with terminal basidiospores (Golubev 1995). Both new species present sexual structures identical to those described for *P. rhodozyma* (Figure 4.5). However, the structure from which the basidium originates most frequently is distinct between species. In the case of *P. novazelandica*, conjugation between two independent cells is the most commonly observed event, while in *P. tasmanica*, most basidia appear to arise from a single cell, with no apparent conjugation (Figure 4.5). Pedogamy (conjugation between mother-cell and its bud), which is the most common occurrence in *P. rhodozyma* (Golubev 1995) was also observed in both novel species but rarely. Similarly to *P. rhodozyma* (Golubev 1995), in both novel species the sexual stage is triggered by the presence of ribitol in the culture medium and individual basidiospores of *P. novazelandica* and *P. tasmanica* were shown to retain the ability to complete the life cycle on their own, typifying these species as homothallic.

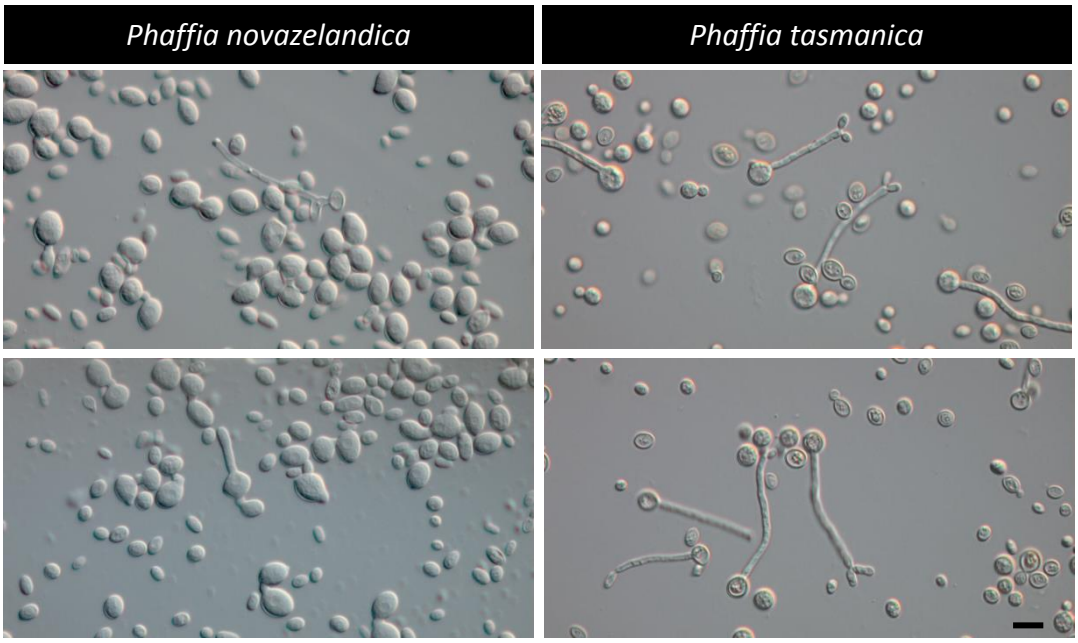


Figure 4.5. Sexual structures of *P. novazelandica* (CBS 14095^T) and *P. tasmanica* (CBS 14096^T). Strains were grown in DWR media at 18°C for 3 weeks. Scale bar represents 10 µm.

The sizes of the basidium and basidiospores of both new species are similar to those of *P. rhodozyma*. For *P. novazelandica* the maximum length of basidia was approximately 77 µm with a diameter of 2-3 µm, whereas for *P. tasmanica* the maximum length of basidia was approximately 100 µm with a similar diameter to that of *P. novazelandica*. The basidiospores of *P. novazelandica* measured 5-6×10-11 µm while those of *P. tasmanica* were slightly smaller 2-3×4-6 µm.

4.3.2. The divergent Australasian lineages belong to the genus *Phaffia*

A phylogeny of the three species of *Phaffia* and related genera was inferred using an alignment of concatenated amino acid sequences of six RNA polymerase subunits (Figure 4.6) obtained from the draft genome sequences of four *Cystofilobasidium* species that resulted from this study, or from public databases (Table 4.6). This analysis confirmed that *Cystofilobasidium* and *Phaffia* are sister genera and yielded a well-supported topology for the remaining lineages within the order Cystofilobasidiales, which represents a step forward in relation to a recently published phylogeny of the Tremellomycetes (Liu *et al.* 2015a; Liu *et al.* 2015b). Moreover, it shows that the two divergent Australasian lineages cluster with *P. rhodozyma* with high bootstrap support, corroborating previous findings obtained using both ITS sequence data (Figure 4.1) and genes encoding enzymes of the astaxanthin biosynthetic pathway (Figure 1.3., Chapter 1) (David-Palma *et al.* 2014).

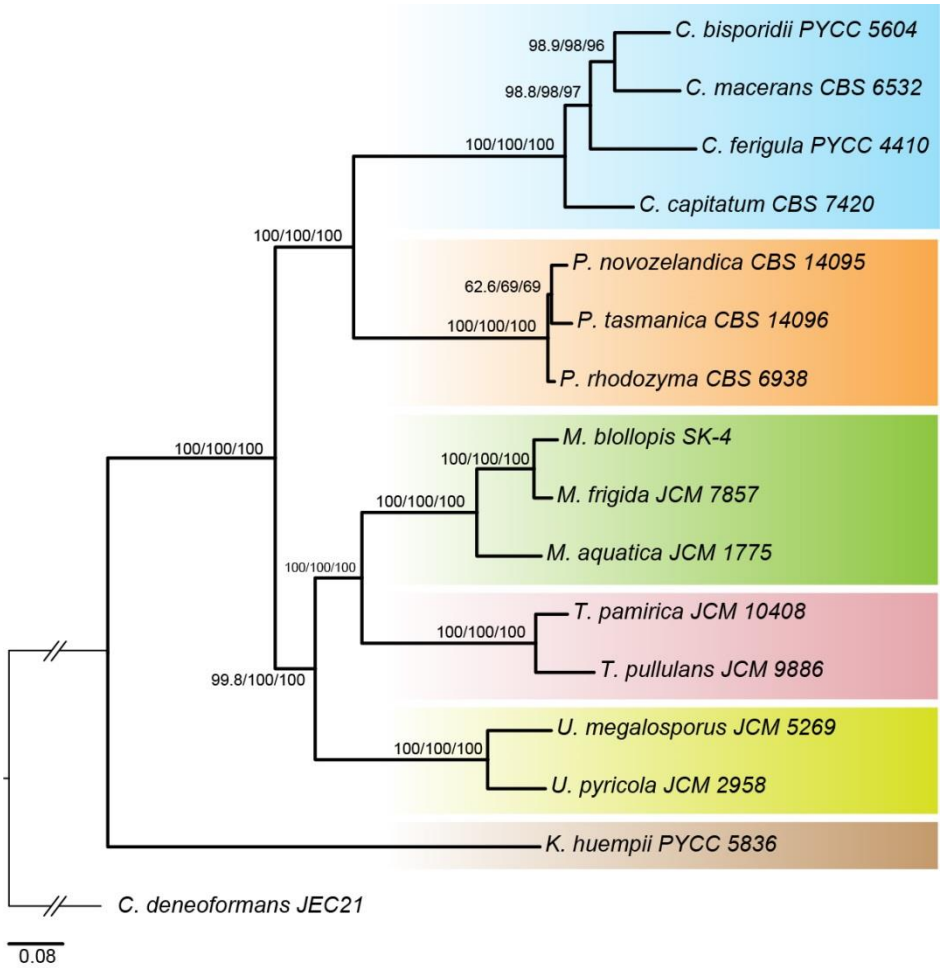


Figure 4.6. Phylogeny of the species of the order Cystofilobasidiales. Phylogeny of representative strains of six genera encompassed in the Cystofilobasidiales order. Numbers on branches represent SH-aLRT support (%), non-parametric bootstrap support (%) and ultrafast bootstrap support (%), respectively (see section 4.2.4 for details). Each of the genera is highlighted with a different color: blue for *Cystofilobasidium*, orange for *Phaffia*, green for *Mrakia*, yellow for *Udeniomyces*, pink for *Tausonia* and brown for *Krasilnikovozyma*.

4.3.3. Astaxanthin production by *P. tasmanica* and *P. novazelandica*

Cultures of representative strains of the three species have a similar orange to salmon-red color (Figure 4.7) that can vary in intensity, depending on the age of the culture and exposure to light, i.e. older cultures or those more exposed to light tend to became more pigmented. *Phaffia rhodozyma* was so far, the only yeast species known to produce astaxanthin. To ascertain whether this characteristic was also present in the new *Phaffia* species, the ability of *P. tasmanica* and *P. novazelandica* to produce this carotenoid was evaluated. With that aim, sequences of the six *P. rhodozyma* genes involved in the biosynthetic pathway of astaxanthin (*IDI*, *CRTI*, *CRTYB*, *CRTE*, *CRTR* and *CRTS*) were used to query the genomes of the type strains of the new species. All six genes could be readily retrieved from both draft genome sequences (Table 4.5). The genes appeared to encode functional enzymes, exhibiting 90-98% amino acid sequence identity to those of *P. rhodozyma* (Figure III.1, Appendix III). The presence of astaxanthin in extracts obtained from isolates of the two-new species was assessed by HPLC-PAD and compared with the results of the type strain of *P. rhodozyma* (CBS 7918^{T}), indicating that *P. novazelandica* and *P. tasmanica* also produce this carotenoid (Diego Libkind personal communication).

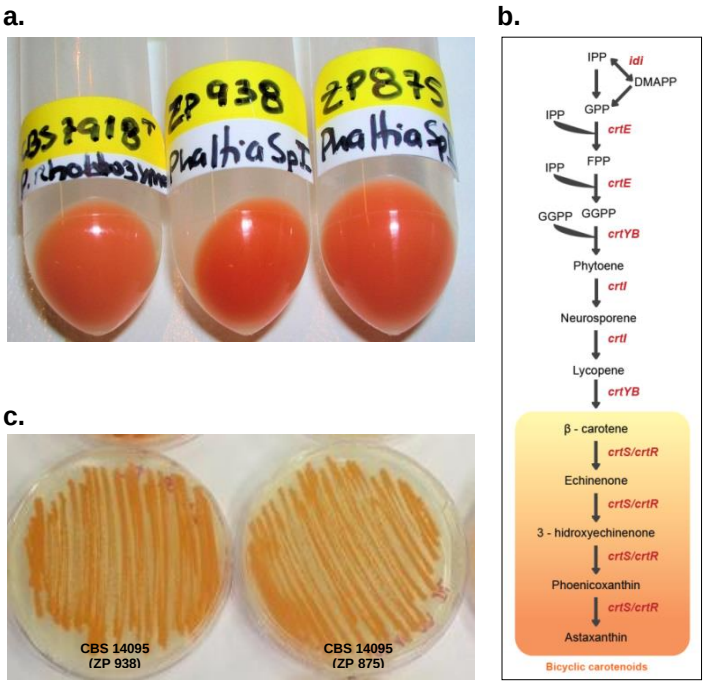


Figure 4.7. *Phaffia* and astaxanthin production. (a) Cell pellets of the type strains of all three *Phaffia* species after 48h growth in liquid YDP; **(b)** Biosynthetic route of astaxanthin in *P. rhodozyma*. The enzymes involved are in red; **(c)** cultures of *P. novazelandica* and *P. tasmanica* in YPD media after 1 week at 20°C.

Table 4.5. Location of genes of the biosynthetic route of astaxanthin in the draft genomes of *P. novazelandica* and *P. tasmanica.*

	Genes of the biosynthetic route of astaxanthin					
Species	*IDI*	*CRTE*	*CRTYB*	*CRTI*	*CRTS*	*CRTR*
P. novazelandica CBS 14095	scaffold 365	scaffold 17	scaffold 304	scaffold 59	scaffold 63	scaffold 182
P. tasmanica CBS 14096	scaffold 103	scaffold 17	scaffold 183	scaffold 27	scaffold 273	scaffold 137

4.3.4. Organization of *Phaffia MAT* loci

MAT genes similar to those present in *P. rhodozyma* were found in the draft genomes of *P. novazelandica* and *P. tasmanica.* Predicted amino acid sequences of the pheromone receptors, Ste3-1 and Ste3-2, revealed the characteristic seven transmembrane domains in both species, with the exception of Ste3-2 of *P. novazelandica* for which eight domains were predicted, although considering the sequence similarity between the Ste3-2 of all the *Phaffia* strains the prediction is most likely not accurate (Figure III.2., Appendix III). The predicted amino acid sequences of the homeodomain transcription factors, Hd1 and Hd2, of *P. novazelandica* presents features similar to the Hd1/Hd2 pair of *P. rhodozyma*, with a nuclear localization signal (NLS) present only in the Hd2 protein and coiled-coils only on the C-terminal region of Hd1. For *P. tasmanica* NLS were predicted in both proteins, while coiled-coils were only present in the C-terminal region of Hd1, as in the other *Phaffia* species (Figure III.3., Appendix III).

Examination of the *MAT* loci of *P. novazelandica* and *P. tasmanica* showed them to be like that of *P. rhodozyma*, with synteny being maintained throughout most of the *MAT* regions, despite a large inversion that took place in *P. tasmanica* close to the *PR* locus (Figure 4.8.a).

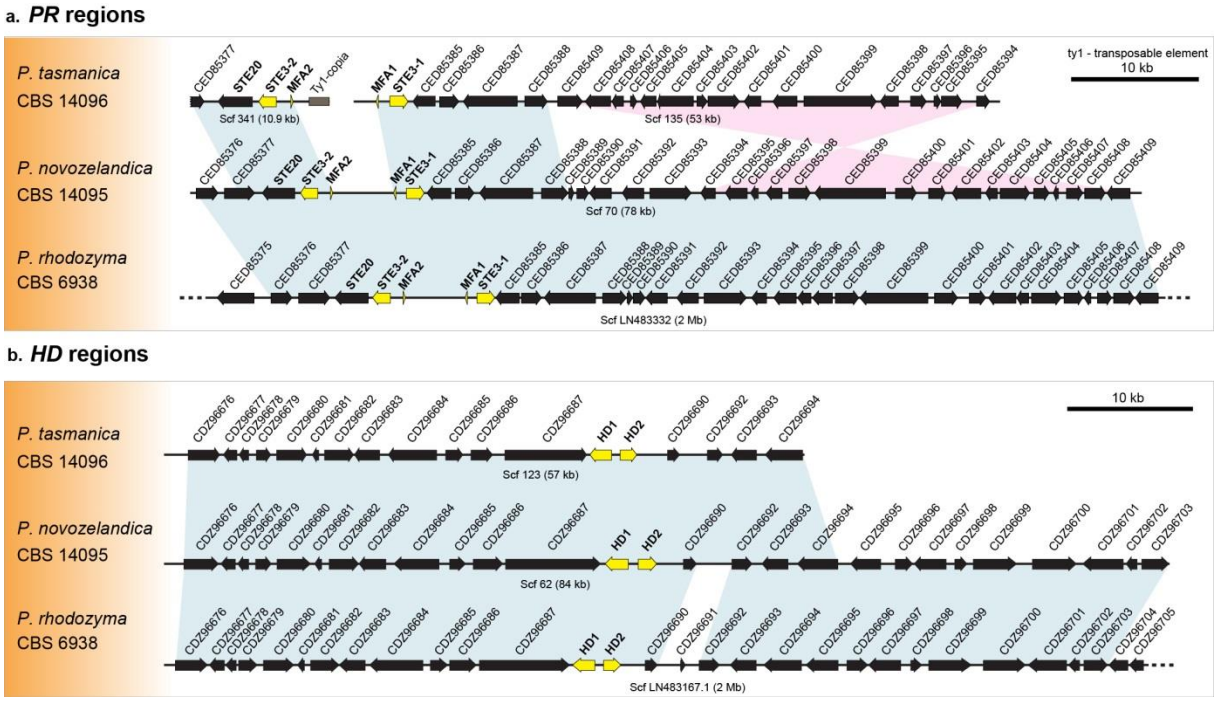


Figure 4.8. Scaffolds encompassing *PR* and *HD* genes in the different *Phaffia* species. Genes are depicted as arrows indicating the direction of transcription and are identified by the protein accession number of their putative orthologs in *P. rhodozyma* (CBS 6938). *MAT* genes are indicated in bold. Below each scaffold, its number and approximate size is indicated. Scaffolds ending with a dotted line indicate that the scaffold is only partially represented. Pheromone receptor genes, pheromone precursor genes and homeodomain transcription factor genes are colored in yellow. Next to one of the scales a key is given for specific features presented in the panel. Blocks of synteny are indicated between the scaffolds in blue and inversions are marked in light pink.

A phylogeny of the pheromone receptor genes of the three *Phaffia* species revealed a trans-specific polymorphism with each receptor having a closer counterpart in the other two species than the second receptor within the same species (Figure 4.9).

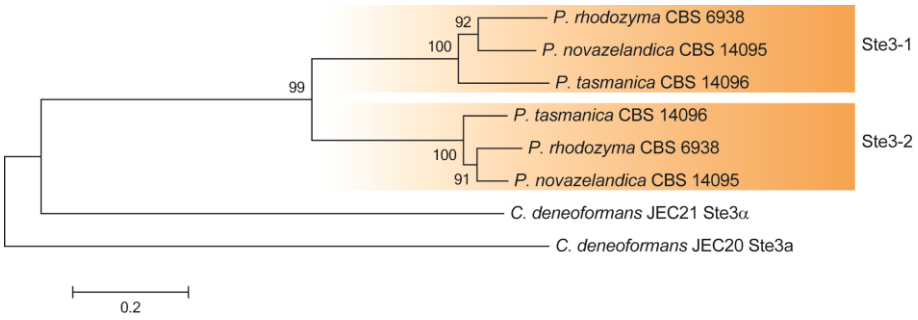
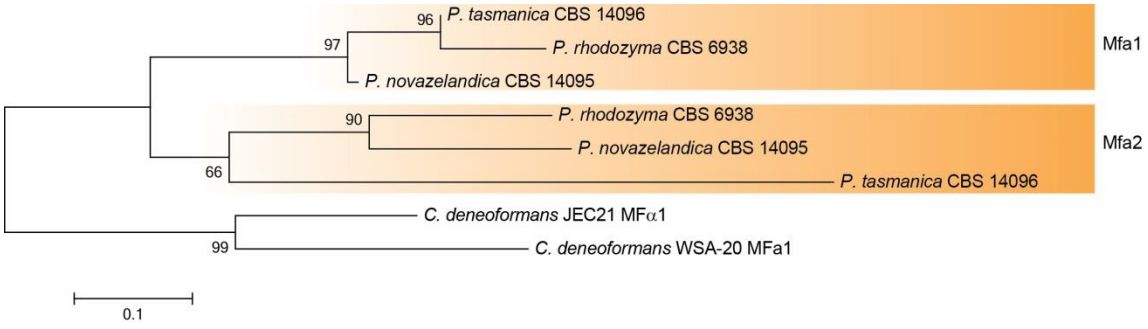


Figure 4.9. Phylogeny of *Phaffia* pheromone receptors. Phylogenetic inference was carried out by Neighbor Joining phylogeny using the JTT matrix-based method as implemented in MEGA 5.1 (Tamura *et al.* 2011) with 1000 bootstrap replicates and with a final dataset of 366 positions. As outgroup the sequences from *C. deneoformans* Ste3a (AAN75624.1) and Ste3α (XP_570116.1) were used.

Interestingly, a similar trans-specific pattern is observed in a phylogeny of the pheromone precursor proteins (Figure 4.10.a). Comparison of the pheromone precursor proteins of the three *Phaffia* species showed Mfa1 to be more conserved and Mfa2 more diverse, especially Mfa2 of *P. tasmanica* (Figure 4.10.b).

a.



b.

```
                                        10        20        30        40
                                ....|....|....|....|....|....|....|....|...
P. rhodozyma CBS 6938 MFA1      MDAFTAIFTPFVASTNESPAASSAKGETKEWTPGMPTSTCVIA
P. novazelandica CBS 14095 MFA1 MDAFTAIFTPFVASTNETTVASSAKGETQGWSSGTPTSTCVIA
P. tasmanica CBS 14096 MFA1     MDAFTAIFTPFVASTNESPAASSAKGETQGWQSGTQTSTCVIA
P. rhodozyma CBS 6938 MFA2      MDVFTTIFTPFVAST-ETELTSEAVGERPE-SKQMFTNFCVIA
P. novazelandica CBS 14095 MFA2 MDAFTGIFTPFIAST-QTEVASEAVGERS--GQNMYTGFCVIA
P. tasmanica CBS 14096 MFA2     MDAFTAVFTLSIPSA-ETEIISQASGQSIK-NAQSSAIYCVIT
```

Figure 4.10. Pheromone precursor proteins from all *Phaffia* species. (a) Phylogenetic inference was carried out by Neighbor Joining phylogeny using the JTT matrix-based method as implemented in MEGA 5.1 (Tamura *et al.* 2011) with 1000 replicates and with a final dataset of 35 positions. As outgroup, the sequences from *C.* deneoformans MFa1 (AAG42766.1) and MFα1 (XP_570122.1) were used. **(b)** Alignment of the pheromone precursor proteins of all three *Phaffia* species. Mature pheromones are highlighted in grey and the C- terminal motif for posttranslational processing is underlined.

4.4. Discussion

4.4.1. Phylogenetic assignment of *P. tasmanica* and *P. novazelandica* in the genus *Phaffia* and its placement within the Cystofilobasidiales

Phylogenetic inferences performed in a previous study using the ITS region, housekeeping genes and genes involved in the biosynthetic route of astaxanthin, had already revealed that the two most divergent *Phaffia*-like lineages found in Australasia, belonged in the genus *Phaffia* (David-Palma *et al.* 2014). The species tree inferred in this chapter from the six subunits of RNA polymerases, although missing the *Itersonilia* genus, strongly supports the existence of six well differentiated genera and clearly shows that both *P. novazelandica* and *P. tasmanica* were correctly assigned to the genus *Phaffia*. The phylogenetic relations inferred between the genera are in agreement with the ones obtained from rDNA datasets containing ITS, D1/D2 and SSU rDNA sequences (Liu *et al.* 2015b) although being very distinct from the topology observed in the seven-gene dataset presented by the same authors (Liu *et al.* 2015b). Given that the phylogenetic relationships depicted in the Cystofilobasidiales tree (Figure 4.6) are strongly supported, it seems likely that they represent the correct organization within the order, especially regarding the position of *Phaffia* as a sister clade of *Cystofilobasidium*.

4.4.2. Phenotypic and *MAT* genomic traits of *Phaffia*

Phenotypic analyses brought to light similarities between the three *Phaffia* species, like the rare ability among basidiomycetes to ferment sugars, as well as some unique traits that set them apart from each other. Morphology of the vegetative cells for *P. novazelandica* was found to be slightly distinct from the other two species but that was only apparent after more prolonged cultivation. Although both *P. tasmanica* and *P. novazelandica* demonstrated a homothallic life cycle, some differences were observed in comparison with the life cycle of *P. rhodozyma*. Each of the three species appears to have a favored way of giving rise to the basidium. While pedogamy was described by Golubev (Golubev 1995) as being the most frequently observed structure preceding the formation of the basidia in *P. rhodozyma*, conjugation between two distinct cells is most often observed in *P. novazelandica*, while in *P. tasmanica* the formation of basidia occurs predominantly from a single cell. Comparison of the *PR* regions of the three species showed a large inversion near the *PR* locus of *P. tasmanica* and the presence of a transposable element downstream of the *MFA2* gene. Overall, both *PR and HD loci* were shown to be very similar within the genus. The recent characterization of the two *MAT* loci of *P. rhodozyma* (Bellora *et al.* 2016) and its genetic dissection (David-Palma *et al.* 2016), demonstrated that all six *MAT* genes identified played a role in sexual reproduction. The *P. rhodozyma HD1* and *HD2* transcription factors genes were shown to be required for completion of the sexual cycle while the two pairs of pheromones and pheromone receptors were both functional, reciprocally compatible and redundant (David-Palma *et al.* 2016). One striking feature of the two distinct pheromone receptors present in *P. rhodozyma* is that they apparently shared a relatively recent common ancestor and

hence did not exhibit the trans-specific polymorphism observed for closely related species of the Tremellales (Kües *et al.* 2011; David-Palma *et al.* 2016). Interestingly, however, the pheromone receptors phylogeny depicted in Figure 4.9, which includes the four newly identified receptors in *P. tasmanica* and *P. novazelandica*, revealed a trans-specific polymorphism within the genus *Phaffia* (Figure 4.9). A similar topology was also observed in the phylogeny pertaining to the pheromone precursor proteins of each species (Figure 4.10). This indicates that the event leading to this similar genomic configuration had a deepen origin, likely extending back to the common ancestor of the extant *Phaffia* species. Additionally, to being syntenic, the *HD* loci of *P. tasmanica* and *P. novazelandica* also encode HD proteins with similar features to the ones found in *P. rhodozyma*. Taken together these results may suggest that the functional roles of *MAT* genes in the three *Phaffia* species may be conserved and that their most recent common ancestor was possibly homothallic.

4.4.3. Association of *P. tasmanica* and *P. novazelandica* with *Nothofagus*

The extant species of *Nothofagus* have a disjunctive distribution in the Southern Hemisphere, occurring in South America and Australasia (Knapp *et al.* 2005). *Cyttaria* stromata and *Nothofagus* leafs from two distinct sites in Australia and in New Zealand where *Nothofagus* trees are endemic were sampled in 2009 yielded *Phaffia* isolates (Figure 4.3). In all but one (Lamington National Park) of the four collection sites, two distinct species of *Phaffia* were found. While *P. tasmanica* appears to be restricted to Tasmania, *P. novazelandica* is present in Australia mainland and in both collection sites in the New Zealand South Island (Lewis Pass and Haast Pass). Different species of *Nothofagus* colonize distinct locations in Australasia (Knapp *et al.* 2005), which translates in the fact that *P. novazelandica* was isolated in association with two distinct *Nothofagus* species (*N. mooreii* in Australia and *N. menziesii* in New Zealand) while *P. tasmanica* was only found in association with *N. cuninghamii*. Because in each of the locations no other trees were sampled it is not possible to exclude the possibility that *Phaffia* isolates may also be associated with other trees, but given the high success rate of isolation from samples of *Nothofagus* (David-Palma *et al.* 2014) it seems likely that *Phaffia* is stably associated with *Nothofagus*.

4.4.4. Production of astaxanthin and the biotechnological potential of *Phaffia* diversity

The genes encoding the enzymes involved in the biosynthetic pathway of astaxanthin were present in the draft genomes o *P. novazelandica* and *P. tasmanica* and had a high identity with their *P. rhodozyma* homologues. No function impairing mutations were found in their predicted coding sequences, suggesting that all genes are functional. The production of astaxanthin was confirmed by HPLC-PAD, indicating that *P. novazelandica* and *P. tasmanica* also produce this carotenoid (Diego Libkind personal communication).

CHAPTER 5

MAT loci evolution in *Phaffia* and in other Cystofilobasidiales

5.1 Introduction

Fungi exhibit a high versatility of mating systems, which often vary between closely related species. The ever-increasing number of fungal genomes becoming available, make possible the exploration of new lineages of fungi regarding their *MAT* regions and the mechanisms underlying the transitions between different mating systems and behaviors also becomes more tangible (Scazzocchio 2014).

Phaffia rhodozyma belongs to the order Cystofilobasidiales, which is the earliest-branching lineage within the Tremellomycetes. Presently, this order encompasses seven genera that, compared to other basidiomycete lineages, have a preponderance of species with homothallic sexual behavior (Kurtzman *et al.* 2011a; Liu *et al.* 2015a; Liu *et al.* 2015b). Of the species currently recognized in Cystofilobasidiales (Liu *et al.* 2015a; Pontes *et al.* 2015) for which a sexual state is known, 10 are homothallic, four are heterothallic and for one species, *Cystofilobasidium macerans*, both homothallic and heterothallic strains are known (Libkind *et al.* 2009; Kurtzman *et al.* 2011a). Most of the homothallic species are included in the genera *Phaffia* and *Mrakia* which, besides the mating behavior, also share the ability to ferment sugars and the association with low temperature environments (Xin & Zhou 2007; Kurtzman *et al.* 2011a; David-Palma *et al.* 2014). Species of *Phaffia*, including those newly described in the present study (Chapter 4), are all homothallic (Golubev 1995; David-Palma *et al.* 2014). Until recently, the same held true for species of the genus *Mrakia,* but with the implementation of the "One Fungus = One Name" nomenclatural principle, the genus has been emended to include species previously belonging to the anamorphic genus *Mrakiella*, namely, *M. aquatica*, *M. cryoconiti* and *M. niccombsii* (Liu *et al.* 2015a). On the other hand, within *Cystofilobasidium*, four species are strictly heterothallic (*C. bisporidii*, *C. ferigula*, *C. infirmominiatum* and *C. lacus-mascardii*) and only two (*C. capitatum* and *C. intermedium*) are homothallic (Fell *et al.* 1973; Sampaio *et al.* 2001; Libkind *et al.* 2009; Pontes *et al.* 2015). In *C. macerans* both homothallic and heterothallic strains are found among the available isolates. However, whereas abundant formation of sexual structures (i.e. mycelium and teliospores) is observed in homothallic strains, the heterothallic strains showed very low sexual competence when paired (Libkind *et al.* 2009). For the remaining species within the Cystofilobasidiales the existence of a sexual cycle is still a matter of debate. In the case of *Itersonilia perplexans*, although a full life cycle had not been described, the species was said to have sexual reproduction and possibly a tetrapolar mating system (Boekhout 1991; Kurtzman *et al.* 2011a). However, it has been recently reaffirmed that no clear sexual stage is observed for this species (Liu *et al.* 2015a). Similarly, *Krasilnikovozyma huempii*, formerly known as *Mrakia curviuscula*, was described as having a homothallic sexual cycle (Bab'eva *et al.* 2002), however the results were disputed in a latter publication (Kurtzman *et al.* 2011a) and the species, now included in the newly described genus *Krasilnikovozyma* (Liu *et al.* 2015a), was considered asexual. The remaining genera of the order Cystofilobasidiales, namely *Tausonia* and *Udeniomyces* are all composed of species for which a sexual cycle has not been elucidated thus far (Kurtzman *et al.* 2011a).

Taking advantage of available and newly generated draft genome sequences of representatives of almost all genera within Cystofilobasidiales (except for *Itersonilia*), the analyses conducted in this

chapter aim to elucidate the *MAT* loci structure in species of this group, namely to understand if there are common genetic signatures underlying the widespread homothallism observed in this order.

5.2 Materials and Methods

5.2.1 Genome sequencing, identification of *MAT* genes and synteny analysis

Strains from three *Cystofilobasidium* species were chosen for whole genome sequencing, namely, *Cystofilobasidium bisporidii* CBS 6346^{T}, *Cystofilobasidium ferigula* PYCC 5628 and *Cystofilobasidium macerans* CBS 2425. Together with strains PYCC 5604, PYCC 4410 and CBS 6532 (Chapter 4), it was possible to compile a genome dataset (Table 5.1) in which strains representing different mating types of these three heterothallic species are represented. Draft genomes were obtained as previously described in section 4.2.2 (Chapter 4) and quality metrics were assessed with QUAST v3.2 (Table IV.1, Appendix IV). Gene coding regions were predicted with Augustus v.3.2.1 with the training annotation files of *C. neoformans* (Stanke *et al.* 2008). In addition to the draft genomes indicated above and the ones used in Chapter 4, other Cystofilobasidiales genomes publicly available as of June 2016, were also included (Table 5.1), namely *Mrakia frigida* ATCC 22029 (JGI) and *Mrakia frigida* Nwmf-AP1 (GenBank). The latter is here proposed to be re-classified as *Mrakia blollopis* Nwmf-AP1 following a taxonomic reevaluation (given as supplementary information in Appendix IV).

Table 5.1. List of all draft genomes from distinct Cystofilobasidiales species used in this chapter.

Species	Strain	Origin of the data	Project
Phaffia rhodozyma	CBS 6938	NCBI	PRJEB6925
Phaffia novazelandica	CBS 14095^{I}	This study	PRJNA371751
Phaffia tasmanica	CBS 14096^{I}	This study	PRJNA371754
Cystofilobasidium capitatum	CBS 7420	This study	PRJNA371774
Cystofilobasidium bisporidii	PYCC 5604	This study	PRJNA371778
Cystofilobasidium bisporidii	CBS 6346	This study	PRJNA371780
Cystofilobasidium ferigula	PYCC 4410	This study	PRJNA371786
Cystofilobasidium ferigula	PYCC 5628	This study	PRJNA371793
Cystofilobasidium macerans	CBS 6532	This study	PRJNA371809
Cystofilobasidium macerans	CBS 2425	This study	PRJNA371814
Mrakia blollopis	SK-4	NCBI	PRJDB3253
Mrakia blollopis[(a)]	Nwmf-AP1	NCBI	PRJNA268263
Mrakia frigida	ATCC 22029	JGI	1040531
Mrakia aquatica	JCM 1775	RIKEN BioResource Center and RIKEN Center for Life Science Technologies	PRJDB3647
Tausonia pamirica	JCM 10408^{I}		PRJDB3689
Tausonia pullulans	JCM 9886^{I}		PRJDB3678
Udeniomyces megalosporus	JCM 5269^{I}		PRJDB3720
Udeniomyces pyricola	JCM 2958^{I}		PRJDB3672
Krasilnikovozyma humepii	PYCC 5836	This study	PRJNA371818

[(a)] Name change following taxonomic reevaluation.

Search for *MAT* genes in all the Cystofilobasidiales genomes was performed as described in Chapter 2. For the scaffolds containing the *PR* and *HD* genes, the predicted gene models in the vicinity of the *MAT* genes, were annotated based on the examination of BLASTP or TBLASTN best hits in GenBank. Putative orthologs were named according to the protein accession number of *P. rhodozyma* (CBS

6938) whenever possible. Synteny conservation across species was assessed manually based on the predicted annotations and confirmed by high-scoring BLASTP hits in GenBank.

5.2.2. *MAT* gene comparisons across Cystofilobasidiales

Predicted pheromone receptor proteins were aligned with ClustalW as implemented in Bioedit (Hall 1999) and amino acid sequence identity was calculated between alleles of each species. Similarly, the predicted homeodomain transcription factor proteins of distinct strains of the same species were aligned and sequence identity was determined separately for the N-terminal, the homeodomain and the C-terminal regions and graphically represented using UGENE v.1.23.1. (Okonechnikov *et al.* 2012). Pheromone precursor proteins were also aligned with ClustalW and mature pheromones were predicted upon comparison with *Phaffia rhodozyma* sequences.

5.2.3. Phylogenetic analyses

Pheromone receptor proteins were aligned with MAFFT v7.221 (Katoh & Standley 2013) using the E-INS-i strategy and poorly aligned regions were trimmed with trimAl v.1.3 using the 'automated1' option (Capella-Gutierrez *et al.* 2009). The final dataset was composed of 26 sequences encompassing 305 residues. Phylogeny was inferred with IQ-TREE software v1.4.3 (Nguyen *et al.* 2015) using the LG+F+I+G4 as the best-fit model of amino acid substitution, as inferred based on the corrected Akaike Information Criterion (AICc). Branch support was determined using ultrafast (Minh *et al.* 2013) and standard non-parametric bootstrapping from 10000 and 1000 replicates, respectively, and also using the approximate likelihood ratio test (SH-aLRT) (Guindon *et al.* 2010) performed with 10000 replicates (Figure 5.1).

5.3. Results

5.3.1 Identification of *MAT* genes in Cystofilobasidiales

Draft genome assemblies, either generated in this study or retrieved from public repositories, of representatives of Cystofilobasidiales, encompassing both homothallic and heterothallic species were inspected for the presence of the genes typically involved in mating type determination in basidiomycetes. This survey revealed that except for *Krasilnikovozyma huempii* (PYCC 5836), for which no putative pheromone precursor genes were found, all other species have at least one pheromone receptor gene (*STE3*), one pheromone precursor gene (*MFA*) and one pair of divergently transcribed homeodomain transcription factors (*HD1* and *HD2*). Failure to identify *MFA* homologues in the genome *K. huempii* could be due to the overall lower quality of the assembly (Table IV.1 in Appendix IV), possibly hampering identification through similarity searches. Since the pheromone genes are usually found near the pheromone receptor, the scaffold harboring *STE3* was examined thoroughly. Again, this failed to identify any putative *MFA* homologue, suggesting that these genes are missing from the genome altogether. The number of receptors and pheromone precursor genes per genome was subsequently investigated and the results are summarized in Table 5.2.

Table 5.2. List of *MAT* genes found in the analyzed Cystofilobasidiales genomes.

Species (strain)	Mating	*STE3*	*MFA*	*HD1/HD2* pairs
Phaffia rhodozyma (CBS 6938)	Homothallic	2	2	1
Phaffia novazelandica (CBS 14095)	Homothallic	2	2	1
Phaffia tasmanica (CBS 14096)	Homothallic	2	2	1
Cystofilobasidium bisporidii (CBS 6346)	Heterothallic (A1B1)	1	4	1
Cystofilobasidium bisporidii (PYCC 5604)	Heterothallic (A2B2)	1	3	1
Cystofilobasidium ferigula (PYCC 4410)	Heterothallic (A1)	1	2	1
Cystofilobasidium ferigula (PYCC 5628)	Heterothallic (A2)	1	2	1
Cystofilobasidium macerans (CBS 6532)	Heterothallic (A1)	1	6	1
Cystofilobasidium macerans (CBS 2425)	Heterothallic (A2)	1	5	1
Cystofilobasidium capitatum (CBS 7420)	Homothallic	1	1	1
Mrakia frigida (ATCC 22029)	Homothallic	2	8	1
Mrakia blollopis (SK-4)	Homothallic	2	14	1
Mrakia blollopis (Nwmf-AP1)	Homothallic	2	8	1
Mrakia aquatica (JCM 1775)	Unknown	1	3	1
Tausonia pamirica (JCM 10408)	Unknown	2	3	2
Tausonia pullulans (JCM 9886)	Unknown	1	2	1 and *HD2**
Udeniomyces megalosporus (JCM 5269)	Unknown	2	6	1
Udeniomyces pyricola (JCM 2958)	Unknown	1	2	1
Krasilnikovozyma huempii (PYCC 5836)	Unknown	1	n.f.	1

**T. pullulans has a pair of divergently transcribed HD1/HD2 genes and an additional HD2.*
n.f. indicates that the gene was not found.

In genome assemblies of the heterothallic species *C. bisporidii, C. ferigula* and *C. macerans* as well as in the homothallic species *C. capitatum* only one receptor gene (*STE3*) was identified in each of the analyzed strains. In contrast, two distinct *STE3* alleles per genome were detected in homothallic species of *Phaffia* and *Mrakia*, while in species with an unknown sexual cycle, such as *Tausonia* spp. and *Udeniomyces* spp., the number of *STE3* genes varied between one and two (Table 5.2). The number of pheromone precursor genes varied even more widely among the different species. For instance, whereas in *C. capitatum* only one *MFA* gene was detected, in *M. blollopis* SK-4 a total of 14

putative pheromone precursor genes were identified. In fact, it is among the homothallic *Mrakia* species that the highest number of *MFA* genes is observed, with a varying number of genes per species, and possibly even between strains (Table 5.2). Finally, in the heterothallic species, both *C. ferigula* strains present two *MFA* genes, but the same pattern is not maintained for the remaining heterothallic species. For species with an unknown sexual cycle, considerable variation in the number of *MFA* genes per genome was also observed (Table 5.2).

Regardless of the mating behavior, the genomes analyzed possess only one pair of divergently transcribed *HD1/HD2* genes, except for *T. pamirica* that has two *HD1/HD2* gene pairs and *T. pullulans* that has one *HD1/HD2* pair and an additional *HD2* gene (Table 5.2).

5.3.2. Pheromone receptors in the Cystofilobasidiales

Using Ste3 sequences identified from the 18 Cystofilobasidiales strains under study, the evolutionary history of the various pheromone receptors was phylogenetically evaluated. The resulting phylogenetic tree (Figure 5.1a) is in general concordant with the species tree (Figure 5.1b), except for the relationship between *Mrakia*, *Udeniomyces* and *Tausonia*, which did not form a monophyletic group as observed in the species tree. It should be noted, however, that the support values of these branches are relatively low in the Ste3 phylogeny.

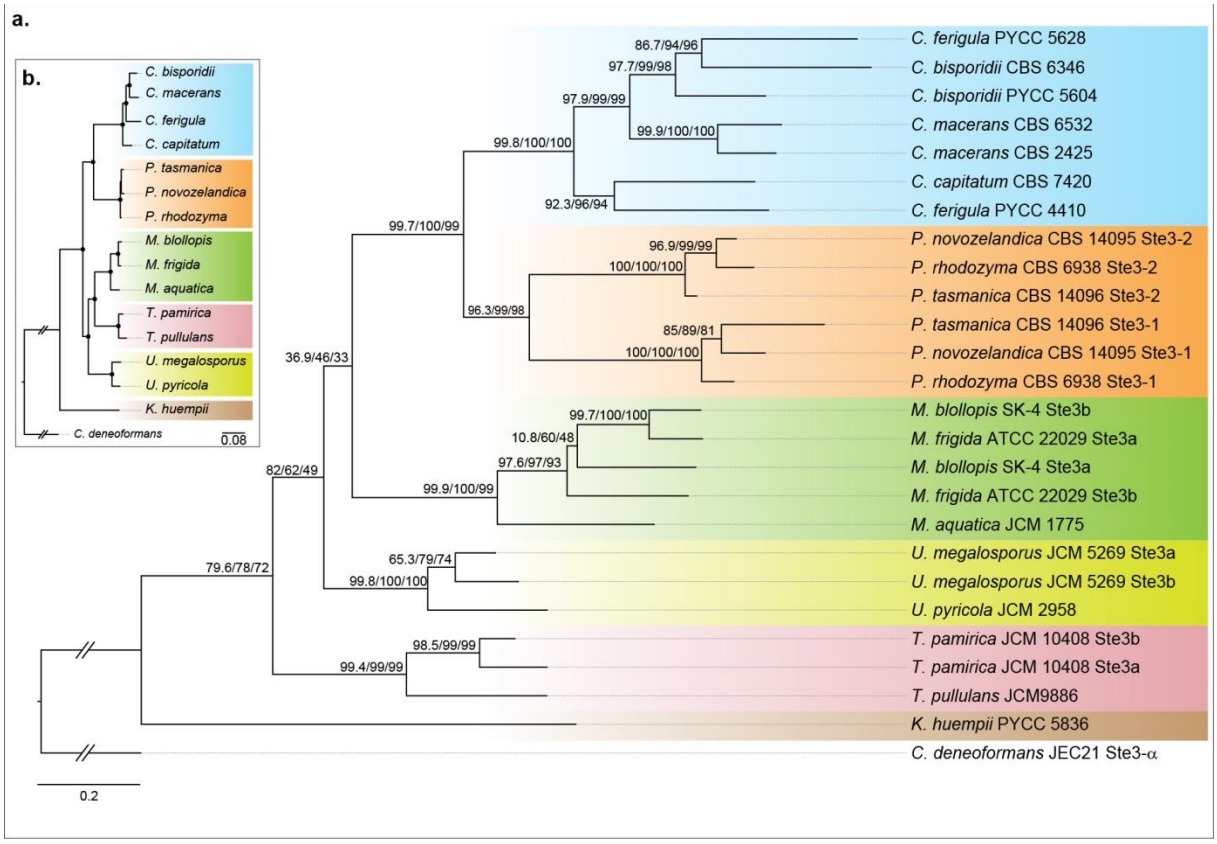


Figure 5.1. Phylogeny of pheromone receptors of the Cystofilobasidiales. (a) Numbers on branches represent SH-aLRT support (%), non-parametric bootstrap support (%) and ultrafast bootstrap support (%), respectively (see section 5.2.2 for details). **(b)** Simplified representation of the phylogeny of Cystofilobasidiales inferred using a concatenated dataset composed of the amino acid sequences of six RNA polymerase subunits

(see section 4.2.4 for details). Black circles in nodes represent branch support > 95%. Each of the genera in both phylogenies is highlighted with a different color: blue for *Cystofilobasidium*, orange for *Phaffia*, green for *Mrakia*, yellow for *Udeniomyces*, pink for *Tausonia* and brown for *Krasilnikovozyma*.

All the pheromone receptors are placed in well supported phylogenetic clusters that coincide with the organization at genus level (SH-aLRT support ≥ 80%; non-parametric bootstrap support ≥ 80% and ultrafast bootstrap support ≥ 95%). Whithin the genus *Phaffia* the Ste3-1 receptors found in the different species were much more similar to each other than to the Ste3-2 receptors of their own species, therefore representing a case of recent trans-specific polymorphism, where allele divergence predates speciation (Figure 5.1 and Chapter 4). A slightly different situation seems to exist for the two homothallic species *M. blollopis* and *M. frigida*. Here, the Ste3b from *M. blollopis* and the Ste3a from *M. frigida* are more closely related, but the same is not observed for the other two pheromone receptors alleles (Figure 5.1). When compared, the two Ste3 putative proteins found in each of the *M. blollopis* strains (SK-4 and Nwmf-AP1), were virtually identical, i.e. Ste3a of strain SK-4 has 100% identity to the Ste3a of strain Nwmf-AP1, while the Ste3b of strain SK-4 has 99,7% identity to the Ste3b of strain Nwmf-AP1, displaing only one different amino acid (Figure IV.2). Within the *Cystofilobasidium* clade, a more complex pattern of evolution is observed (Figure 5.1a). For instance, the two pheromone receptors present in strains PYCC 5628 and PYCC 4410, corresponding to mating types of *C. ferigula*, cluster in distinct sub-clades and represents the most dissimilar allele pair among the various *Cystofilobasidum* species, showing only 51% amino acid sequence identity (Table 5.3). This also represents the oldest trans-specific polymorphism detected within this lineage. Quite the opposite is observed when comparing the Ste3 sequences of the two strains of opposite mating types of *C. macerans*, which cluster tightly together and have diverged considerably less (73% amino acid identity). An intermediate scenario is observed for Ste3 sequences of CBS 6346 and PYCC 5604 from *C. bisporidii*, that share only 56% amino acid sequence identity. For *Tausonia pamirica* and *Udeniomyces megalosporus*, two pheromone receptor were present in the same strain. These pairs of receptors clustered together and had high amino acid identity in relation to each other (Table 5.3).

Table 5.3. Percentage of amino acid identity between different Ste3 in several species.

Species	Mating	Amino acid sequence identity (%)
Phaffia rhodozyma	Homothallic	50
Phaffia novazelandica	Homothallic	49
Phaffia tasmanica	Homothallic	50
Cystofilobasidium bisporidii	Heterothallic	56
Cystofilobasidium ferigula	Heterothallic	51
Cystofilobasidium macerans	Heterothallic	73
Mrakia frigida	Homothallic	68
Mrakia blollopis	Homothallic	63
Tausonia pamirica	Unknown	70
Udeniomyces megalosporus	Unknown	81

5.3.3 Pheromones in the Cystofilobasidiales

For each pheromone precursor gene found in the Cystofilobasidiales species the respective pheromone precursor protein and mature pheromone was determined (Table 5.4).

Table 5.4. *MFA* genes from Cystofilobasidiales, their pheromone precursor proteins and putative mature pheromones.

Species (strain)	MFA genes	Different pheromone precursor proteins	Different mature pheromones
Phaffia rhodozyma (CBS 6938)	2	2	2
Phaffia novazelandica (CBS 14095)	2	2	2
Phaffia tasmanica (CBS 14096)	2	2	2
Cystofilobasidium bisporidii (PYCC 5604)	3	2	2
Cystofilobasidium bisporidii (CBS 6346)	4	1	1
Cystofilobasidium ferigula (PYCC 4410)	2	1	1
Cystofilobasidium ferigula (PYCC 5628)	2	2	2
Cystofilobasidium macerans (CBS 6532)	6	6	6
Cystofilobasidium macerans (CBS 2425)	5	5	3
Cystofilobasidium capitatum (CBS 7420)	1	1	-
Mrakia frigida (ATCC 22029)	8	5	5
Mrakia blollopis (SK-4)	14	11	9
Mrakia blollopis (Nwmf-AP1)	8	7	6
Mrakia aquatica (JCM 1775)	3	2	2
Tausonia pamirica (JCM 10408)	3*	at least 2	at least 2
Tausonia pullulans (JCM 9886)	2	2	2
Udeniomyces megalosporus (JCM 5269)	6*	at least 4	at least 3
Udeniomyces pyricola (JCM 2958)	2	1	1

**The sequence of MFA2 of T. pamirica is partial and only the first 12 amino acids of the pheromone precursor protein are known, therefore it was excluded from the determinations of the number of different precursor proteins and mature pheromones (last two columns of the table). A similar situation occurs to MFA1 of U. megalosporus.*

The number of mature proteins present in each draft genome varies greatly within the heterothallic species that belong to *Cystofilobasidium*. For both *C. bisporidii* and *C. ferigula*, one strain has two different mature pheromones while the other has one. In these species, the two mature pheromones found in one mating type are different from that found in the opposite mating type (Figure 5.2).

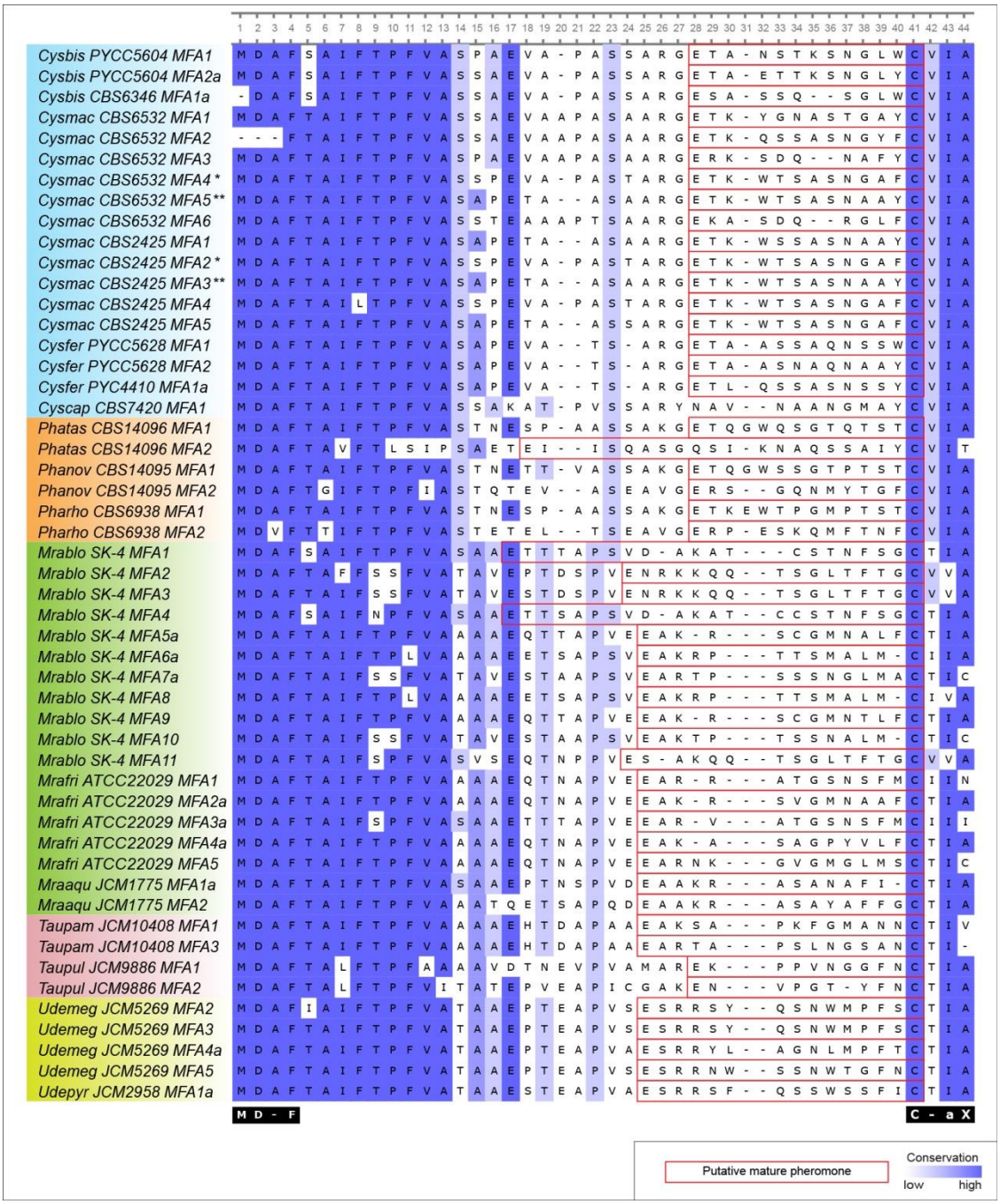


Figure 5.2. Protein sequence alignment of the pheromone precursors of members of the Cystofilobasidiales. Pheromone precursor proteins found in each of the Cystofilobasidiales strains were aligned with ClustaW. Conservation is depicted in different shades of blue in the alignment and putative mature pheromones are enclosed by a red line. Under the alignment, two black boxes highlight the conserved motifs found: ”–“ stands for variable residues within the motif; “a” stands for any aliphatic amino acid and “X” for any amino acid. The pheromone precursor proteins are listed according to genera, which are delimitated by different color (using the same color code as in Figure 5.1). Asterisks next to the pheromone precursor proteins names indicate identical sequences. Redundant pheromone precursor proteins or partial ones, too small to provide relevant information were excluded from the alignment.

In *C. macerans*, a dichotomy is also observed regarding the number of mature pheromones possessed by each of the two mating types. *Cystofilobasidium macerans* strain CBS 6532 has six different mature pheromones, while strain CBS 2425 has three (Table 5.4). Interestingly, the two strains of *C. macerans* share identical pheromone precursor proteins and mature pheromones (Figure

5.2). The pheromone precursor proteins of *MFA4* and *MFA5* from strain CBS 6532 are identical to the pheromone precursor proteins of *MFA2* and *MFA3* from strain CBS 2425, respectively (Figure 5.2). Considering the mature pheromones of these two *C. macerans* strains, with the exception of pheromone Mfa1, all the mature pheromones from CBS 2425 are shared with strain CBS 6532 (Figure 5.2). Regarding the homothallic species, three distinct situations are observed: (i) in *C. capitatum*, it was not possible to identify the possible site for N-terminal processing in order to predict a putative mature protein; (ii) in the genus *Phaffia*, all species show the same number of different mature pheromones; (iii) in *Mrakia* a variable number of mature pheromones is observed not only between different species (*M. blollopis versus M. frigida*) but also, between distinct strains of the same species (strains Sk-4 and Nwmf-AP1 from *M. blollopis*) (Table 5.4 and Figure 5.3).

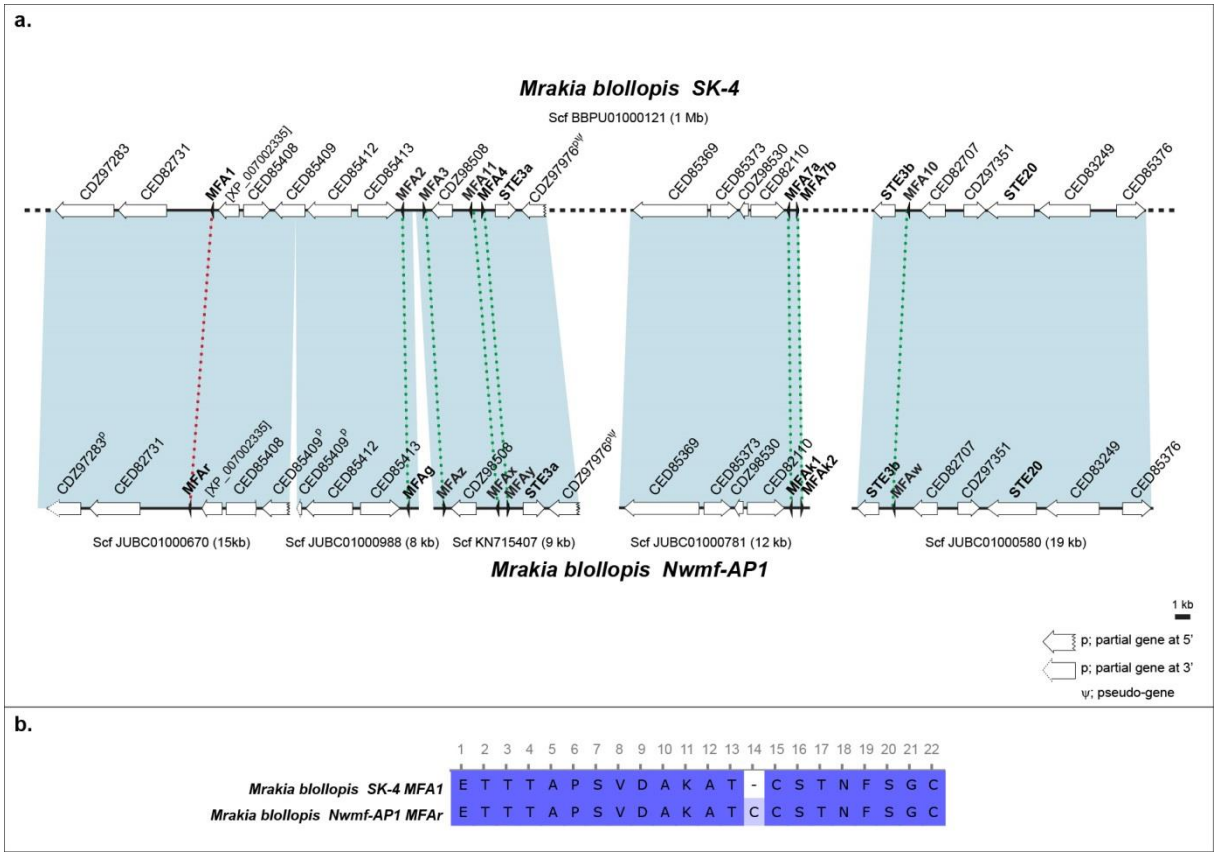


Figure 5.3. Genomic regions of *M. blollopis* strains, depicting the *MFA* genes found and their similarity. **(a)** Scaffold regions from strains SK-4 and Nwmf-AP1 where *MFA* genes were found. Genes are depicted as arrows indicating the direction of transcription and are identified by the protein accession number of their putative orthologs in *P. rhodozyma* (CBS 6938) whenever possible; the accession number in brackets indicates that the putative ortholog belongs to another organism. *MAT* related genes (*MFA*, *STE3* and *STE20*) are indicated in bold. Dotted line in the upper scaffold indicates that parts of it are omitted. Above or below each scaffold (scf), its number and approximate size is indicated. Blue color blocks uniting the scaffolds from SK-4 and Nwmf-AP1 strains indicate presence of synteny. Pheromone genes (in black) connected by a dotted green line indicate that those genes encode identical mature pheromones while the ones connected by a dotted red line indicate that those genes encode distinct mature pheromones **(b)** Alignment of the two distinct mature pheromones found in *Mrakia blollopis*.

Mrakia blollopis SK-4 strain has nine different putative mature pheromones, while for strain Nwmf-AP1 only six were found. To the exception of Mfar, all the mature pheromones found in *M. blollopis* strain Nwmf-AP1 are shared with strain SK-4, however there is high similarity shared between Mfar and

Mfa1 from strain SK-4 (Figure 5.3b). Concerning the species without a known sexual cycle the number of putative mature pheromones varies between one and three. The alignment of the pheromone precursor proteins shows high conservation at the N-terminus and C-terminus regions of all the proteins (Figure 5.2). At the N-terminus, almost all species have the motif MDAF, apart from Mfa2 from *P. rhodozyma* that presents the motif MDVF. At the C-terminal region, the pheromone precursor proteins of *Cystofilobasidium* and *Phaffia* show conservation of the "CaaX" motif (where "a" stands for an aliphatic amino acid and X for any amino acid), depicting mainly a CVIA motif (Figure 5.2). For the pheromone precursor proteins of the genera *Mrakia*, *Tausonia* and *Udeniomyces* the "CaaX" motif is maintained in some cases, but the majority of the precursors show a C-terminal motif that instead of having two aliphatic amino acids after the cysteine residue, present only one, while the other is a threonine residue similar to pheromone precursors found in fungi like *Sporisorium reilianum* or *Rhodotorula toruloides* (Schirawski *et al.* 2005; Coelho *et al.* 2008).

5.3.4. Homeodomain transcription factors in homothallic and heterothallic Cystofilobasidiales

To compare the results obtained for *P. rhodozyma* regarding the degree and distribution of polymorphisms observed in *HD1* and *HD2* homeodomain transcriptions factors from distinct strains (Chapter 2) with those of other Cystofilobasidiales, the predicted proteins of *HD1* and *HD2* genes from different strains of homothallic species (*M. blollopis*) and heterothallic species (*C. bisporidii, C. macerans* and *C. ferigula*) were compared (Figure 5.4). The N-terminal region of the Hd1 proteins of *M. blollopis* show an amino acid identity of 61%, while the C-terminal region shows a 99% amino acid identity between the two proteins. A similar result is obtained when the Hd2 of the two strains are compared, with the N-terminal region showing an even lower amino acid identity of only 38% (Figure 5.4a). These results differ greatly from the ones obtained when the same comparison was performed among *P. rhodozyma* strains, a species also homothallic (Chapter 2).

For strains PYCC 5604 and CBS 6346 belonging to the heterothallic species *C. bisporidii,* the N-terminal regions of both Hd1 and Hd2 are highly polymorphic, with just 34% and 31% amino acid identity, respectively. In contrast, the C-terminal regions of Hd1 and Hd2 show high levels of conservation with 99% and 95% of amino acid identity, respectively (Figure 5.4b). For *C. macerans* strains CBS 2425 and CBS 6532, the predicted Hd1 and Hd2 proteins show a distinct distribution of polymorphisms compared to the one observed for *C. bisporidii*. Although in the N-terminal regions of both Hd1 and Hd2 the amino acid identity values are similar to those observed in *C. bisporidii* (35% and 36%), in the C- terminal regions, amino acid identity is considerably lower (61% and 65%). In fact, even the homeodomain region of both Hd1 and Hd2 are less conserved than in all other examined species (Figure 5.4). For the heterothallic species, *C. ferigula*, both *HD1* and *HD2* genes are incomplete in either one or both strains (PYCC 5628 and PYCC 4410) not allowing a comparison between the Hd1 and Hd2 proteins over their entire length (Figure 5.4d). Nevertheless, for the regions that can be compared, part of the C-terminal region of Hd1 and the complete C- terminal region of

HD2, a high level of conservation is observed. The pattern of polymorphism distribution is thus more similar to that observed in *C. bisporidii* than to the pattern uncovered in *C. macerans*.

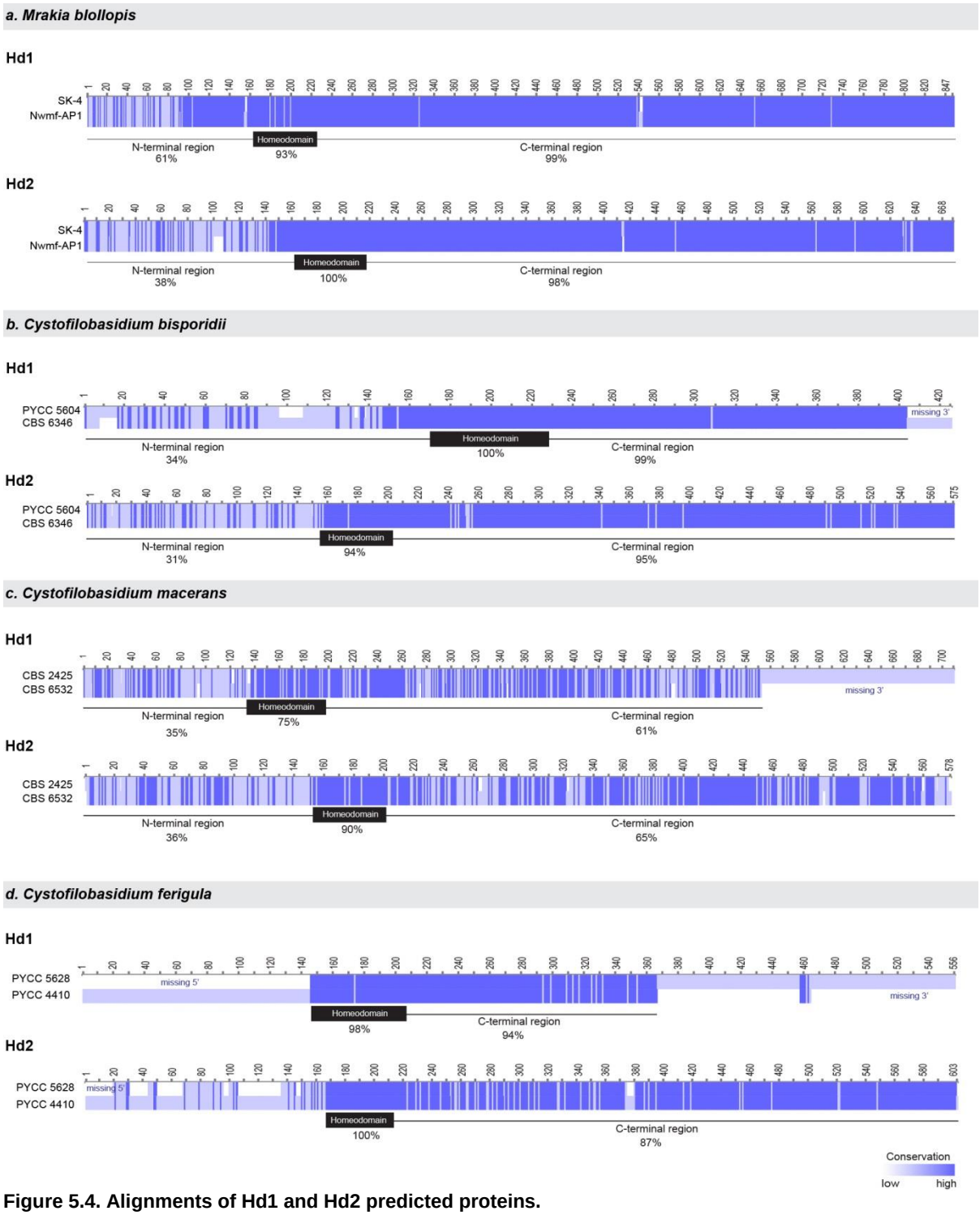


Figure 5.4. Alignments of Hd1 and Hd2 predicted proteins.
Alignments of predicted Hd1 and Hd2 proteins are shown for different strains the same species. **(a)** strains SK-4 and Nwmf-AP of homothallic species *M. blollopis*. **(b)** strains PYCC 5604 and CBS 6346 of the heterothallic species *C. bisporidii* **(c)** strains CBS 2425 and CBS 6532 of the heterothallic species *C. macerans.* **(d)** strains PYCC 5628 and PYCC 4410 of the heterothallic species *C. ferigula*. Conservation of the aligned proteins is depicted in different shades of blue, as depicted in the key bellow. Homeodomain regions are indicated below the alignments by black boxes. Amino acid identity of each region is indicated below each alignment. When only partial proteins were available, calculation of amino acid identity was performed using only well aligned regions shared by both proteins.

5.3.5 The *PR* locus in Cystofilobasidiales

Comparison of the *PR* scaffolds of *P. tasmanica*, *P. novazelandica* and *P. rhodozyma*, shows considerable conservation of both gene content and synteny in all three species (Figure 5.5a). The two clusters composed by *STE3-1/MFA1* and the *STE3-2/MFA2* are present in all *Phaffia* species and the location of *STE20* relatively to *STE3-2/MFA2* cluster is also maintained in the three species. In *P. tasmanica* the two *STE3/MFA* clusters are present on two separate scaffolds. One of the scaffolds (numbered 341) exhibits a transposable element at its end while in the other (scaffold 135) there has been an inversion of a region localized downstream of the *STE3-1/MFA1* cluster (Figure 5.5a). For *Cystofilobasidium* scaffolds harboring pheromone receptor and pheromone precursor genes some conservation of gene content is observed, although the small size of the scaffolds does not support a more informative synteny analysis (Figure 5.5b). Comparison of the two strains of distinct mating types of *C. macerans* (CBS 6532 and CBS 2425) shows that genes flanking the pheromone receptor genes in each strain are different, with the same being true for the *C. ferigula* strains. This suggests that synteny is not maintained between different mating types in the region immediately surrounding the *STE3* gene. In most *Cystofilobasidium* strains (*C. bisporidii* strain PYCC 5604, *C. macerans* strains CBS 6532 and CBS 2425 and *C. capitatum* strain CBS 7420) it is possible to observe a common feature shared also with all *Phaffia* species: the *STE20* gene is flanked downstream by gene CED85377 (Figure 5.5b).

The *PR* region of the three *Mrakia* species shows some conservation of gene content and even some syntenic blocks of genes, especially between *M. blollopis* and *M. frigida* (Figure 5.5c). Among the syntenic blocks identified are the regions encoding the two *STE3* alleles of each strain and also some of the several pheromone precursor genes found in these species (Figure 5.5c). A closer inspection of these regions shows the presence of several repeats both in *M. blollopis* and *M. frigida* (Figure 5.6). In the *M. blollopis* strain SK-4 the repeats are associated with two particular genes: CDZ97976 and CED85376 (Figure 5.6a). These genes appear as direct and/or inverted repeats and in some cases the genes are incomplete. Similarly, in *M. frigida* ATCC 22029 these two genes are also present in repeated regions (Figure 5.6b). The repeated region involving gene CDZ97976 also encompasses two pheromone precursor genes (*MFA3* and *MFA4*; *a* and *b* copies) and the last 207 nucleotides of both *STE3* alleles (*STE3a* and *STE3b*) (Figure 5.6 b and c). In fact, the first 80 nucleotides of *STE3a* and *STE3b* are also identical and part of a distinct repeat region, which additionally encompasses *MFA2* gene (*a* and *b* copies) and the first 77 nucleotides of *MFA5* and *MFA1* (Figure 5.6c).

As for *T. pamirica* and *T. pullulans* the genomic content of the *PR* regions appear to be conserved between the two strains analyzed (Figure 5.5d). In *T. pullulans,* the *PR* region that in *T. pamirica* is present in on single scaffold, is divided in two (scaffolds 98 and 8). Some parts of both scaffolds are syntenic with the one from *T. pamirica* but also a possible inversion is apparent between the two species (Figure 5.5d). In *T. pamirica* in addition to the *MAT* genes found in scaffold 9, a partial *STE3* gene and a *MFA* gene, were found in smaller individual scaffolds (Figure 5.5d). Similarly, for *T. pullulans* an additional *MFA* gene was found in a smaller scaffold, outside of scaffold 8, where the pheromone receptor was found (Figure 5.5d). For *Udeniomyces megalosporus* and *U. pyricola*, the

comparison of the *PR* regions shows some gene content conservation but synteny is limited when compared to previous genera (Figure 5.5e). For *U. pyricola*, an additional copy of *MFA1* gene (*MFA1b*) is found in a smaller scaffold (scaffold 109), in the vicinity of a transposable element. For strain PYCC 5836 of *K. huempii*, both the *STE3* and *STE20* genes were found in one single scaffold. However, no *MFA* genes were found in that or any other scaffold. In the *PR* scaffold of the *K. huempii* strain, a transposable element was also found (Figure 5.5f). Interestingly, a syntenic block composed by *STE20*, CED83249 and CED85376 is present in all the *PR* regions analyzed in species in the sub-clade formed by the genera *Mrakia*, *Tausonia*, *Udeniomyces* and *Krasilnikovozyma* (Figure 5.5c to 5.5f).

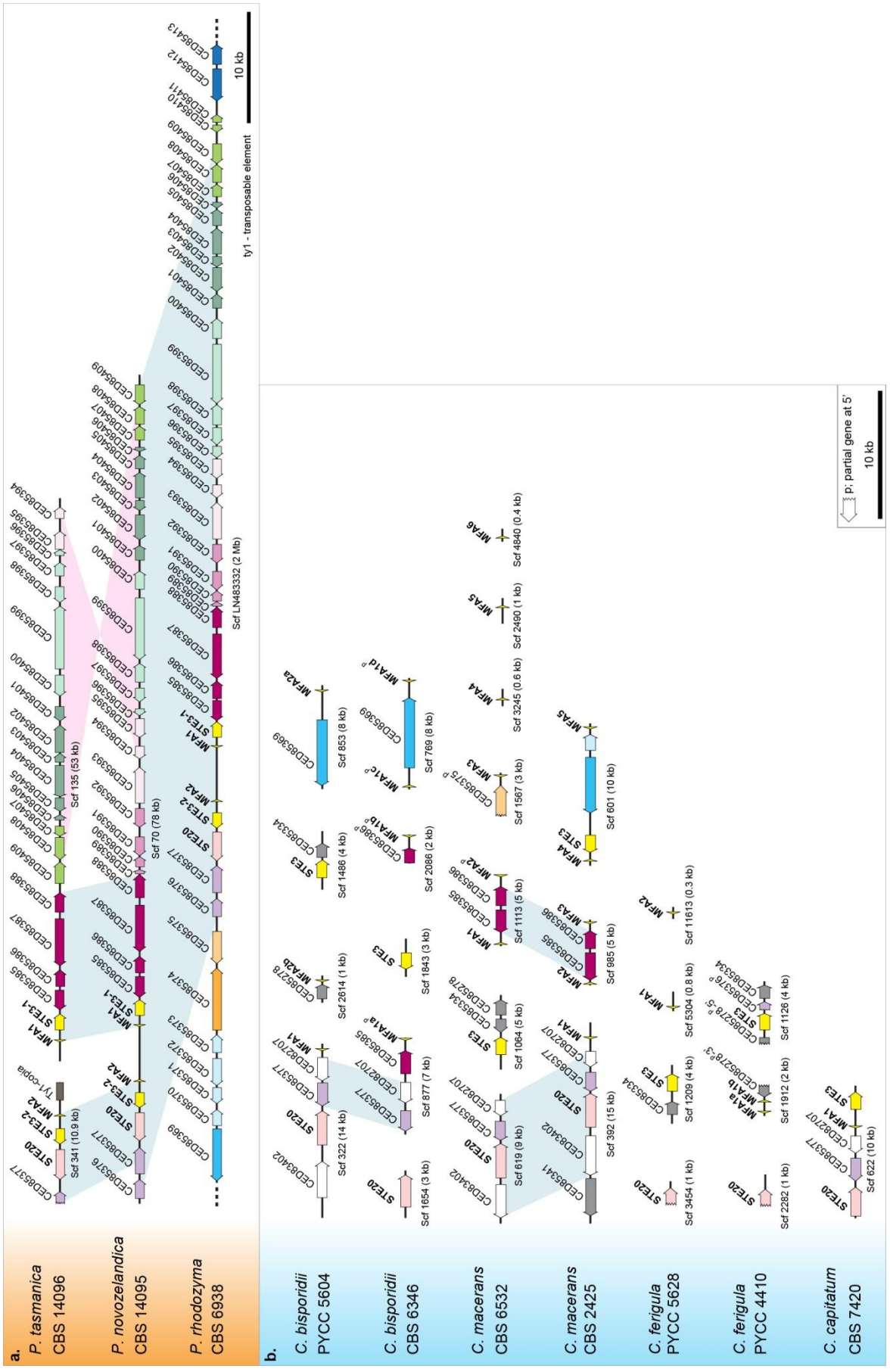


Figure 5.5. *PR* regions in members of the Cystofilobasidiales. Description is given in page 83.

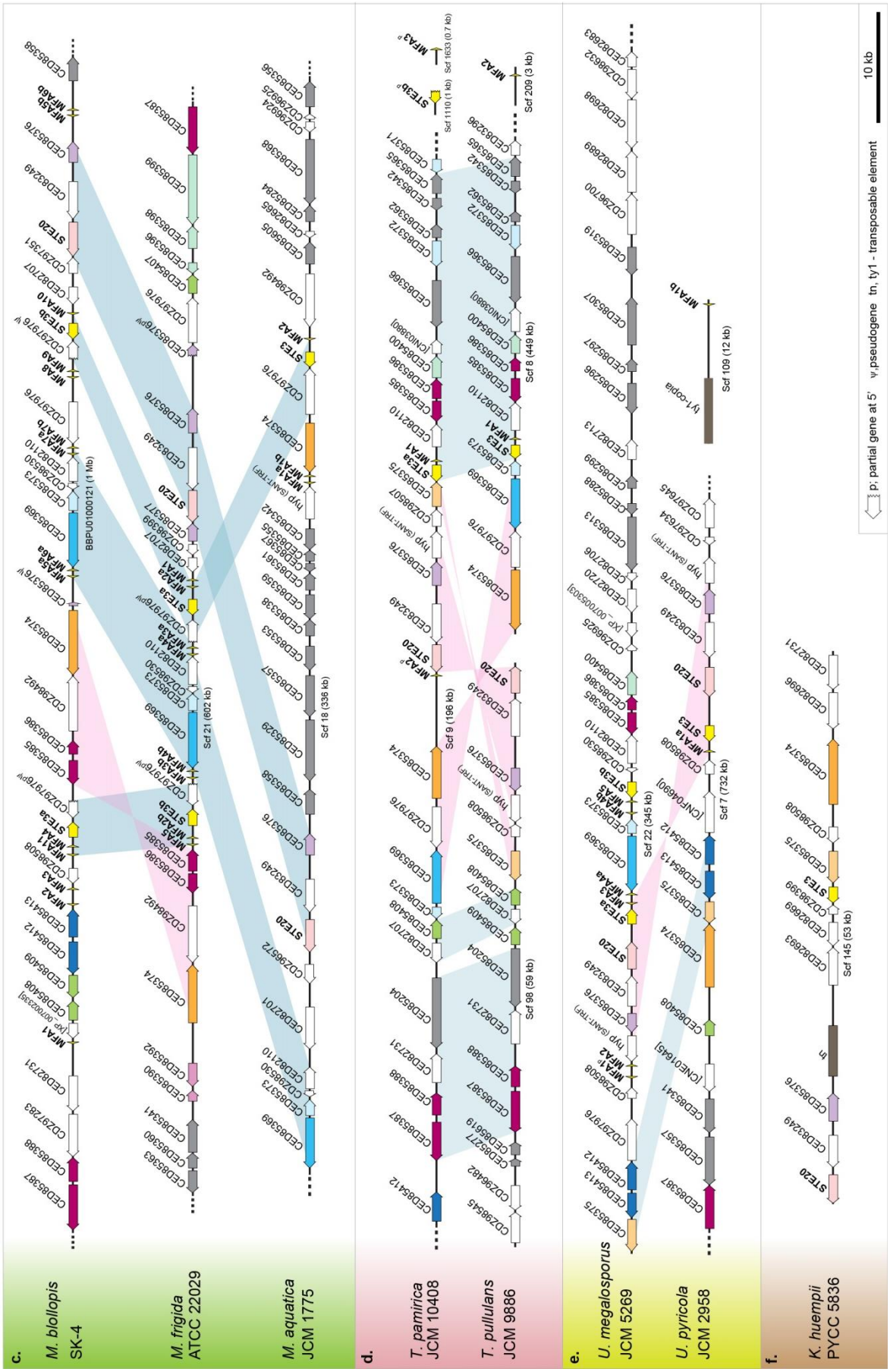


Figure 5.5. *PR* regions in members of the Cystofilobasidiales (continued). Description is given in page 83.

Figure 5.5. *PR* regions in members of the Cystofilobasidiales.
Scaffolds encoding *PR* genes for **(a)** *Phaffia* **(b)** *Cystofilobasidium* **(c)** *Mrakia* **(d)** *Tausonia* **(e)** *Udeniomyces* and **(f)** *K. huempii*. Strains are organized by genera delimitated by different colors (using the same color code as in Figure 5.1). Genes are depicted as arrows indicating the direction of transcription and are identified by the protein accession number of their putative orthologs in *P. rhodozyma* (CBS 6938) whenever possible. *MAT* genes and *MAT* related genes (*STE20*) are indicated in bold. Below each scaffold, its number and approximate size is indicated. Scaffolds ending with a dotted line indicate that the scaffold is only partially represented. Pheromone receptor genes and pheromone precursor genes are colored in yellow. The remaining genes are colored according to the organization found in *P. rhodozyma* (CBS 6938). Genes depicted in gray, are genes present in the *P. rhodozyma* (CBS 6938) *PR* scaffold but not present in the depicted region. Genes shown in white are genes that do not belong to the *PR* scaffold of *P. rhodozyma* CBS 6938. Next to each scale a key is given for specific features presented in the different panels. Blocks of synteny are indicated between the scaffolds in blue or light pink in case of an inversion.

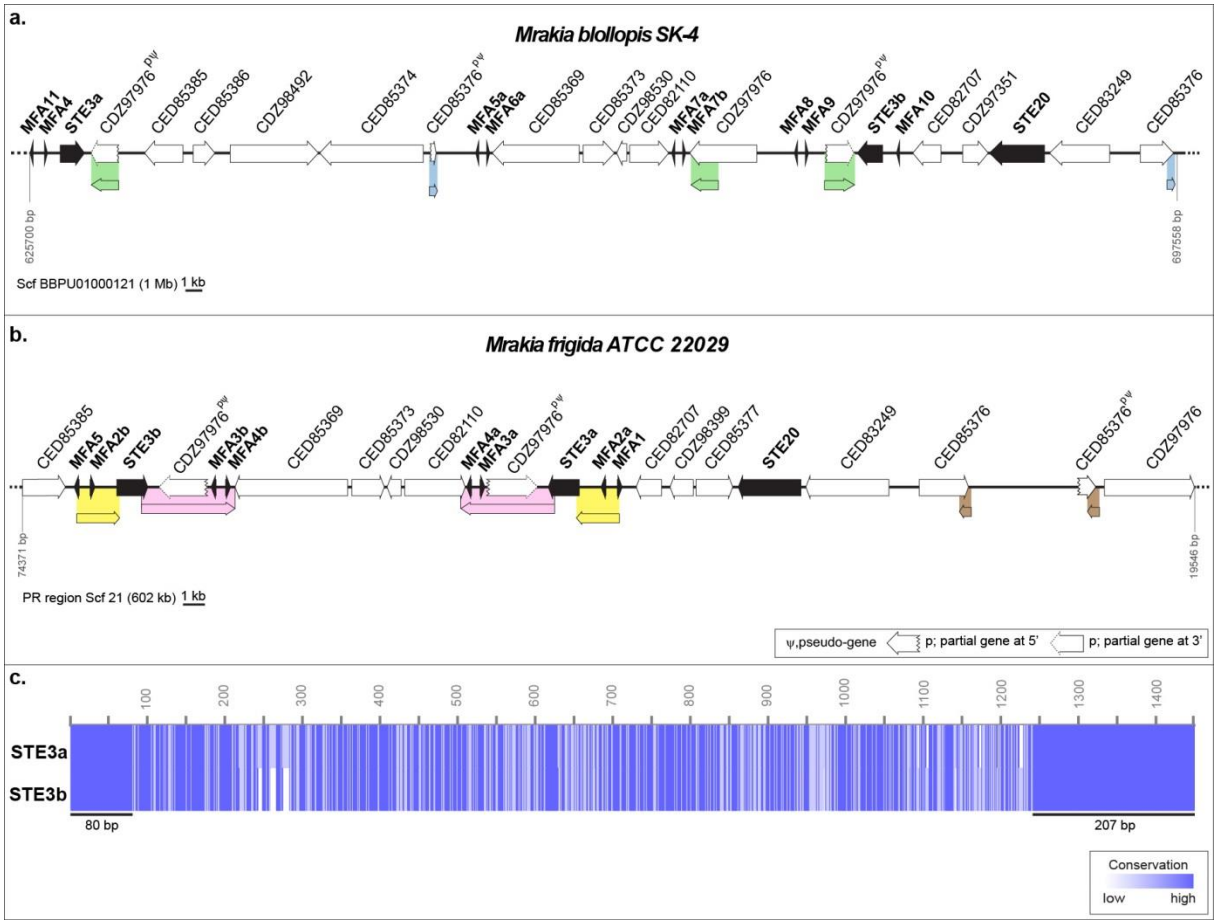


Figure 5.6. Repeat regions present in the *PR* scaffolds of *M. blollopis* and *M. frigida*.
(a) Detailed representation of part of the scaffold encoding the *PR* genes for strain SK-4 of *M. blollopis*. **(b)** Detailed representation of part of the scaffold encoding the *PR* genes for strain ATCC 22029 of *M. frigida*. Genes are depicted and identified as in previous figures. Colored in black are *STE3*, *MFA* and *STE20* genes, while the remaining genes are depicted in white. A key is given for specific features depicted in the first two panels. Blocks of color with adjacent arrows shown under each scaffold represent repeated regions. Each color identifies a particular repeat and the direction of the arrow indicates if it is a direct or inverted repeat. **(c)** Alignment of the two *STE3* alleles found for strain ATCC 22029 (*M. frigida*) indicating the identical regions at the beginning and at the end of the genes. Conservation is depicted by different shades of blue throughout the alignment, as indicated in the key.

5.3.6 The *HD* locus in Cystofilobasidiales

Similarly, to the analysis performed for the *PR* regions, the scaffolds containing the *HD* genes of the available Cystofilobasidiales genomes studied were compared. The *HD* scaffolds of the three *Phaffia* species show a high degree of synteny. In fact, *P. tasmanica* and *P. novazelandica* are completely syntenic. Comparing *P. novazelandica* or *P. tasmanica* with *P. rhodozyma* shows that synteny is interrupted by a single gene identified as CDZ96691, a hypothetical protein (Figure 5.7a). For *Cystofilobasidium* species*,* synteny also appears to be maintained in the scaffolds encoding the homeodomain transcription factor genes. Downstream of *HD2*, a gene identified as a putative ortholog of CDZ96734 (a Snf2 family DNA-dependent ATPase) of *P. rhodozyma* is present in all seven strains from the four *Cystofilobasidium* species studied (Figure 5.7b). In this small region, synteny is maintained not only across different mating types but also between different species. The high level of conservation of both gene content and overall synteny observed in the *HD* scaffolds of *Phaffia* and *Cystofilobasidium* species contrasts with what is observed for the *HD* scaffolds of *Mrakia*, *Tausonia* and *Udeniomyces* species (Figure 5.7c, d and e). For *M. blollopis* SK-4 and *M. frigida* ATCC 22029 only four genes, excluding the *HD1/HD2* pair, are shared between the two (Figure 5.7c), while the two strains of *M. blollopis* (SK-4 and Nwmf-AP1) are completely syntenic at the *HD* region (Figure V.3, Appendix V). For *Tausonia* species, no genes are shared between the two studied strains in the vicinity of the *HD* genes. The two pairs of *HD* genes found in *T. pamirica* differ considerably, but all four encoded proteins exhibit a putative homeodomain region (Figure 5.7d). In *T. pamirica* the two scaffolds where the *HD* pairs were found, harbor mostly genes not belonging to the *HD* scaffold of *P. rhodozyma*, with only one exception (CDZ96978 in grey) (Figure 5.7d). Between the *U. megalosporus* and *U. pyricola* strains some conservation of gene content and synteny is observed in the direct vicinity of the *HD* genes (Figure 5.7e). In *K. huempii*, the genes in the direct vicinity of the *HD1/HD2* pair are different from the ones found in the other Cystofilobasidiales species (Figure 5.7f).

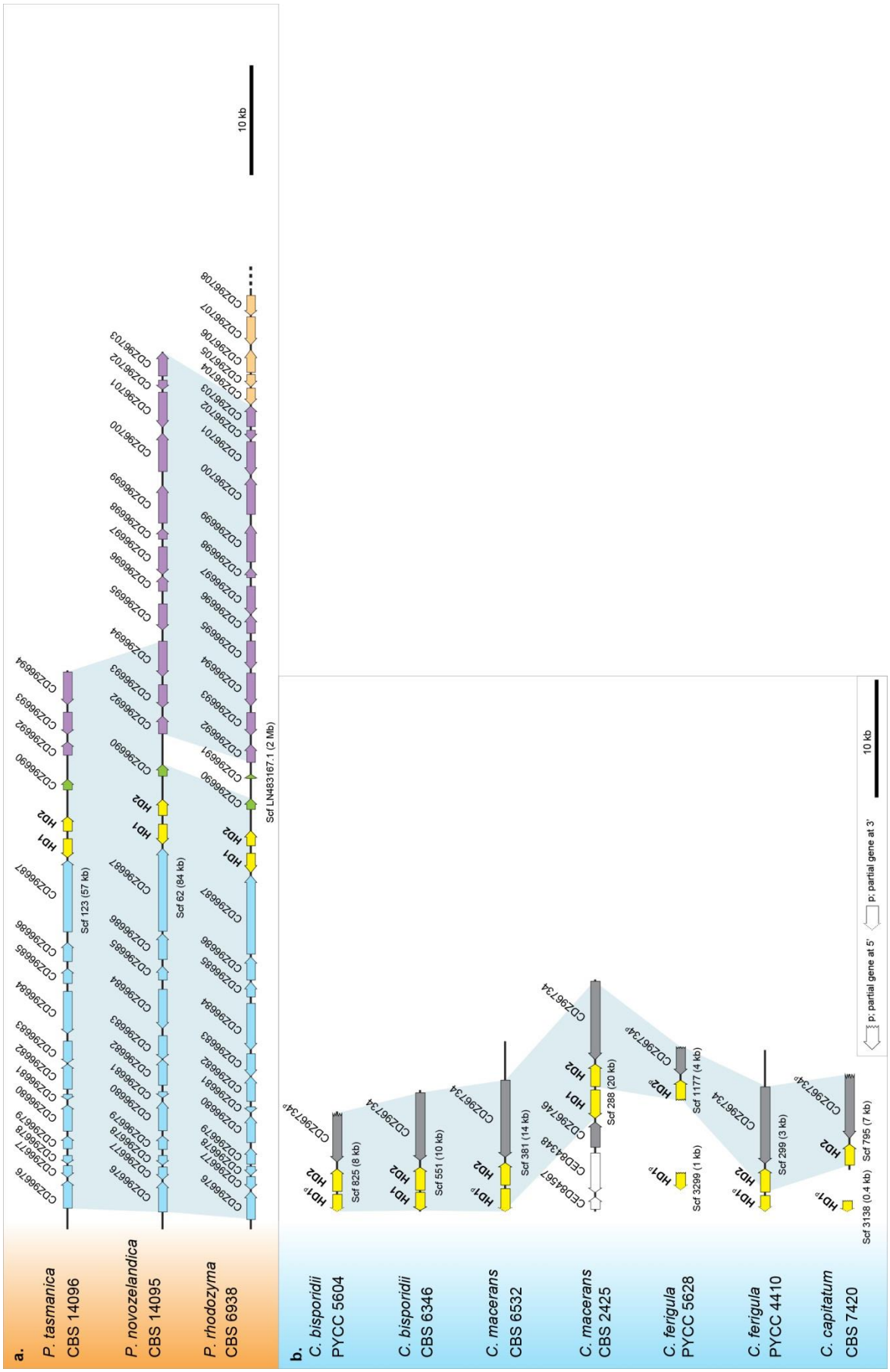


Figure 5.7. *HD* regions in members of the Cystofilobasidiales. Description is given in page 87.

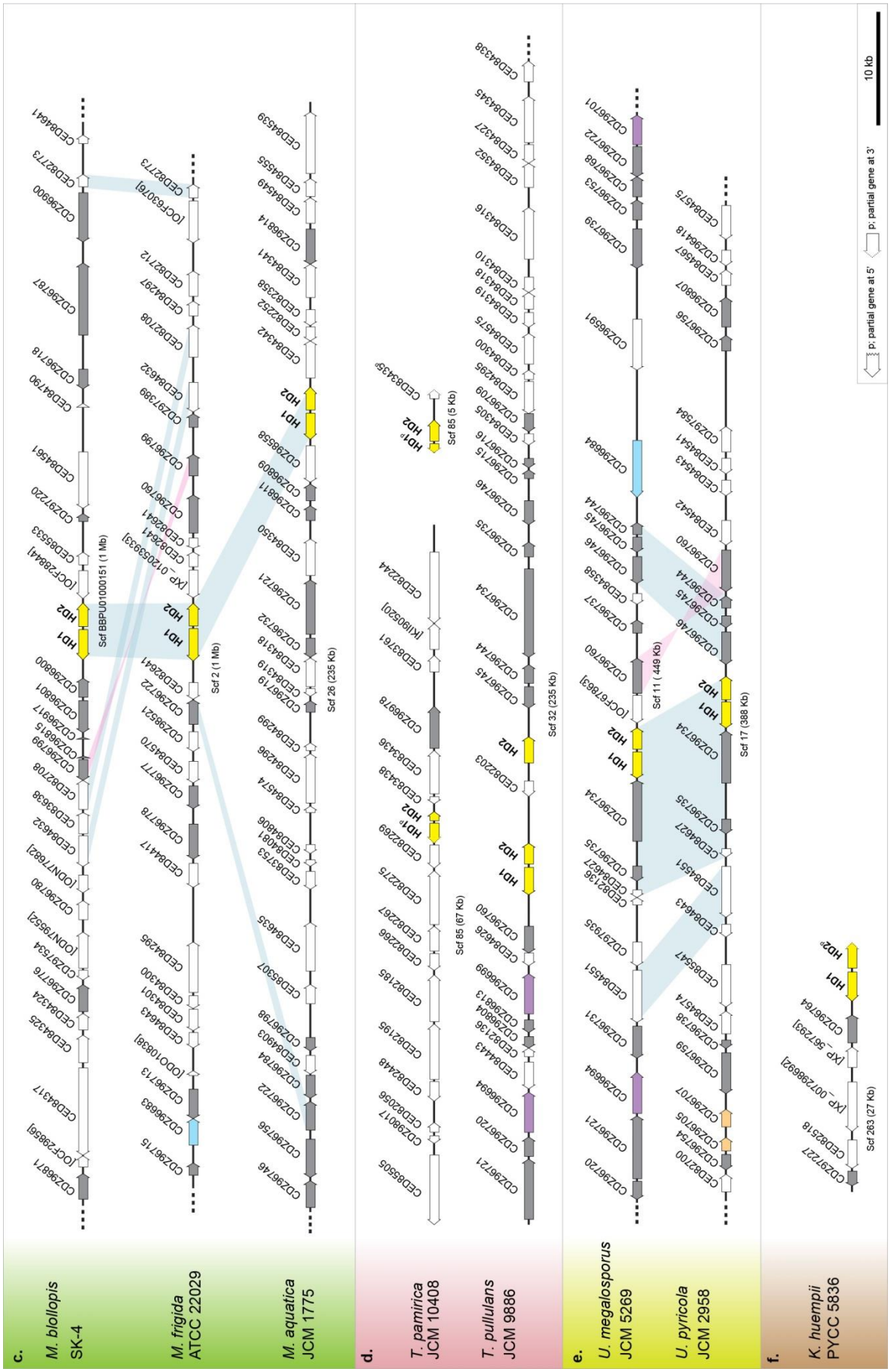


Figure 5.7. HD regions in members of the Cystofilobasidiales (continued). Description is given in page 87.

Figure 5.7. *HD* regions in members of the Cystofilobasidiales. Scaffolds encoding *HD* genes for **(a)** *Phaffia* **(b)** *Cystofilobasidium* **(c)** *Mrakia* **(d)** *Tausonia* **(e)** *Udeniomyces* and **(f)** *K. huempii*. Strains are organized by genera which are delimitated by different colors (using the same color code as in Figure 5.1). Genes are depicted as arrows indicating the direction of transcription and are identified by the protein accession number of their putative orthologs in *P. rhodozyma* (CBS 6938) whenever possible. Specific *MAT* genes i.e. *HD1* and *HD2* are indicated in bold. Below each scaffold represented, its number and approximate size is indicated. Scaffolds ending with dotted line indicate that the scaffold is only partially represented. Homeodomain transcription factor genes are colored in yellow. The remaining genes are colored according to the organization found in *P. rhodozyma* (CBS 6938). Genes depicted in gray, are genes present in *P. rhodozyma* (CBS 6938) *HD* scaffold but not present in the depicted region. Genes depicted in white, are genes that do not belong to the *HD* scaffold of *P. rhodozyma* (CBS 6938). Next to the each scale a key is given for specific features presented in the different panels. Blocks of synteny are indicated between the scaffolds in blue or light pink in case of an inversion.

5.4. Discussion

The results described in this chapter extend the initial analysis of *MAT* loci in *Phaffia rhodozyma* to other homothallic and heterothallic representatives of the Cystofilobasidiales and report some advances with respect to understanding the evolution of homothallism in basidiomycetes. To determine the structure and gene content of *MAT* loci, draft genomes of a set of 15 representative species encompassing most of the known phylogenetic diversity within Cystofilobasidiales were analyzed. For *Cystofilobasidium* this involved genome sequencing and *de novo* assembly of different mating types of three heterothallic species (*C. bisporidii*, *C. ferigula* and *C. macerans*). The results presented here suggest that in this taxonomic order, homothallism is not strictly linked to the presence in the genome of the same number and organization of *MAT* genes as observed in *P. rhodozyma*. Instead, a more complex situation was found where the number and organization of *MAT* genes underlying the homothallic sexual behavior seem to vary among the different genera. Hence, the results suggest that transitions to homothallism probably occurred several times independently in this lineage and reveal a particularly dynamic pattern of *MAT* gene evolution in this order.

5.4.1. *MAT* loci in *Phaffia*

All three *Phaffia* species show similar *PR* and *HD* regions. Each species has two *STE3* and *MFA* gene pairs (Table 5.2) and the *PR* region shares overall synteny, especially between *P. rhodozyma* and *P. novazelandica* (Figure 5.5a). An inverted region downstream of the *STE3-1* gene in *Phaffia tasmanica* likely reflects a normal degree of divergence between the species, which are expected to accumulate following speciation (Nosil & Feder 2012; Stukenbrock 2013). Phylogenetic inference showed that the two receptor alleles found in each *Phaffia* species appear to have diverged before speciation (Figure 5.1). Considering that in *P. rhodozyma* Ste3-1 is activated by Mfa2 and Ste3-2 is activated by Mfa1 (Chapter 3) (David-Palma *et al.* 2016), it is conceivable that these compatibilities also predate speciation, given the similarities observed between the *PR* regions of the three *Phaffia* species (Figure 5.5a). It is therefore also very likely that *P. novazelandica* and *P. tasmanica* display a pattern of compatibility between pheromones and receptors like that of *P. rhodozyma*. The genetic arrangement of the two *PR* clusters present in the *PR* locus of each *Phaffia* species is similar to other *PR* clusters present in many heterothallic basidiomycete species, with each functional *STE3* gene being in the vicinity of a *MFA* gene encoding a functional pheromone that nevertheless fails to activate its

neighboring receptor (Riquelme *et al.* 2005; Fraser *et al.* 2007). It seems likely that the genomic arrangement observed in the *PR* loci of all *Phaffia* species resulted from the fusion of two heterothallic loci, although other possible mechanisms, such as duplication followed by *de novo* evolution of compatibility of the *STE3/MFA* pairs through mutation cannot be entirely excluded. Synteny between the *HD* regions of the three *Phaffia* species is also conserved (Figure 5.7). For *P. rhodozyma,* it was suggested that the pair of *HD1/HD2* alleles is most likely compatible and that it may have evolved from a typical self-incompatible pair by mutation (Chapter 3) (David-Palma *et al.* 2016). Based on the overall similarity of the *PR* and *HD* regions and *MAT* genes of *P. tasmanica*, *P. novazelandica* and *P. rhodozyma* it seems reasonable to propose a common genetic basis for the homothallic behavior of the three species which, based on phylogenetic evidence (Figure 5.1), can be traced back to their most recent common ancestor (MRCA). This homothallic MRCA likely evolved from a heterothallic tetrapolar ancestor given the configuration of the *PR* locus and the fact that the two *MAT* loci appear to be unlinked in *P. rhodozyma* genomes (Sharma *et al.* 2015; Bellora *et al.* 2016). The evolutionary pathway leading from the tetrapolar heterothallic ancestor to the homothallic MRCA of all *Phaffia* species may have generated transitional states where only one of the two compatibility systems, controlled by the *PR* and *HD* loci respectively, would have been self-compatible, while the other remained self-sterile, as previously postulated by Fraser et al. (2007) for transitions from tetrapolar to bipolar heterothallic systems (Fraser *et al.* 2007).

5.4.2. *MAT* loci in *Mrakia*

Analysis of the *PR* and *HD* regions in homothallic species in *Mrakia* revealed different features. While the *PR* region shows some gene content conservation and even some synteny blocks (Figure 5.5c), in the *HD* region almost no common genes are present between *M. blollopis* and *M. frigida* (Figure5.7c). Two *STE3* alleles were found in *M. blollopis* and *M. frigida*, as in *Phaffia* species, but the number of predicted *MFA* genes is much higher raising the possibility that the two receptors may be activated by one, some or all the *MFA* gene products. The reason why a homothallic organism would maintain such degree of pheromone diversity and possible redundancy is obscure. For the heterothallic basidiomycete *Schizophyllum commune* it has been argued that functional redundancy of several *MFA* genes might be a selective advantage for the generation of multiple *PR* alleles within a population because it allows the possibility to generate a *STE3* gene with new specificity by subsequent mutations (Fowler *et al.* 2001; Kües 2015). However, it is hard to grasp its importance in a homothallic species. Pheromone receptors of *M. frigida* and *M. blollopis* do not display a clear trans-specific polymorphism as do the receptors in *Phaffia* (Figure 5.1). However, the topology lacks strong support for one of the branches in the *Mrakia* cluster (Figure 5.1 in green). In any case, allele diversification occurred independently in both *Phaffia* and *Mrakia* lineages (Figure 5.1). Interestingly, both *Mrakia* species showed several repeated regions within the *PR* scaffolds (Figure 5.6) suggesting that events such as gene duplication and further genomic rearrangements may have played a role in shaping these regions. Such mechanisms might explain for example the presence of several repeated *MFA* genes embedded in the *PR* region. These stretches of DNA with high similarity may also serve

as additional sources of repetitive DNA for illegitimate recombination events (Idnurm 2011b; Sun & Heitman 2016) adding to the complexity observed in the *PR* regions of the homothallic *Mrakia* lineages. The high level of sequence similarity observed at both 5'- and 3'-ends of the two *STE3* alleles of *M. frigida* (Figure 5.6) is possibly a consequence of intrachromosomal gene conversion, which in general causes homogenization of relatively short DNA regions. It has been shown that interchromosomal crossovers between *MAT* regions may occur, within the *MAT* loci itself by way of gene conversion as observed in *Cryptococcus deneoformans* or *Microbotryum violaceum* (Sun *et al.* 2012; Fontanillas *et al.* 2015; Idnurm *et al.* 2015; Sun & Heitman 2016).

For *M. blollopis* the availability of draft genomes from two distinct strains allowed comparison of the two sets of *STE3* alleles found in each strains, revealing that they are virtually identical and that synteny is maintained in the *PR* region similarly to what is observed when strains of *P. rhodozyma* are compared (Chapter 2) (David-Palma *et al.* 2016). Contrary to *P. rhodozyma*, *M. blollopis* presents a different number of *MFA* genes between strains and also one different mature pheromone. However the difference in the number of *MFA* genes between strains might be a consequence of the quality of the genome assembly presently available for the two strains, assembled into 167 and 1539 scaffolds, respectively for SK-4 (Tsuji *et al.* 2015) and Nwmf-AP1 (NCBI Assembly ASM81596v1), potentially hindering the identification of the complete set of *MFA* genes. Analysis of the *HD* regions of both strains (SK-4 and Nwmf-AP1) reveals them to be completely syntenic (Figure V.3, Appendix V). Comparison of the alleles in both *M. blollopis* strains showed that both Hd1 and Hd2 have hyper-polymorphic N-terminal regions, while displaying highly conserved homeodomain and C-terminal regions (Figure 5.4). Evidence from several studies indicates that in heterothallic species discrimination between self and non-self-interactions of the *HD1/HD2* alleles is conferred by the N-terminal region of both Hd1 and Hd2 proteins (Yee & Kronstad 1993; Kües *et al.* 1994; Asante-Owusu *et al.* 1996; Yue *et al.* 1997). Although with exceptions, such as in *Phanerochaete chrysosporium* (James *et al.* 2011), in which the most divergent regions of the Hd1 and Hd2 proteins are the C-terminal regions, in the vast majority of basidiomycetes this usually translates into a highly variable N-terminal region between different alleles of both *Hd1* and *Hd2* proteins, while the homeodomain and C-terminal regions display a more conserved pattern (Kämper *et al.* 1995; Badrane & May 1999; Hull *et al.* 2005; Coelho *et al.* 2011; Maia *et al.* 2015). Interestingly, although *M. blollopis* displays a homothallic behavior, the polymorphism pattern observed in both Hd1 and Hd2 (Figure 5.4) resembles that commonly observed in heterothallic basidiomycetes, instead of mimicking the observations in *P. rhodozyma*, in which the variable amino acid positions are evenly distributed throughout the proteins (Figure 2.3, Chapter 2).

Similarly, to *Phaffia* species, the MRCA of the homothallic lineages of *Mrakia* was possibly homothallic and could have evolved from a heterothallic ancestor. If so, the MRCA of the homothallic lineages of *Mrakia* may have obtained their genomic organization at the *PR* loci by recombination between two different mating types present in the heterothallic ancestor, similarly to what was hypothesized for *Phaffia* species, followed by subsequent rearrangements. Considering such a scenario the genomic organization at the *PR* locus observed in *M. aquatica*, could be representative of a possible heterothallic mating type, given that this species is not homothallic and has only one pheromone

receptor gene. On the other hand, it is conceivable that the MRCA of the homothallic lineages of *Mrakia* appeared due to a duplication event at the *PR* locus, followed not only by the generation of new compatibilities between receptor and pheromones, but also by genomic rearrangements. At the *HD* loci, the polymorphism pattern exhibited by both the Hd1 and Hd2 proteins appears more likely to have evolved in a heterothallic setting. It remains to be elucidated if these proteins are necessary for the completion of the homothallic sexual cycle observed in *M. blollopis*, as are the ones from *P. rhodozyma.* Additional genomic information regarding other strains from this species, and further study of its sexual cycle will be necessary to elucidate the molecular determinants responsible for the homothallism observed in *M. blollopis* and other homothallic *Mrakia* species.

5.4.3. *MAT* loci in *Cystofilobasidium*

Remarkably, another homothallic species studied, *C. capitatum* also differs from both *Phaffia* and *Mrakia* in what concerns the number of *MAT* genes at the *PR* region (Figure 5.5b). In this case, only one *STE3* gene and one *MFA* gene was found, leading to the hypothesis that the genes may form a compatible Ste3/Mfa pair. However, the pheromone precursor protein of *C. capitatum* does not display a site for N-terminal processing of the pheromone precursor similar to those observed in other Cystofilobasidiales. It has been reported that the C-termini of pheromones appear to be far more important for pheromone functionality than the N-termini (Kües 2015). Therefore, either the pheromone precursor has an atypical N-terminal processing site that still allows for the formation of a functional mature pheromone, or no mature pheromone is formed. Considering these two possibilities, the homothallic sexual cycle of *C. capitatum* may depend on (i) a self-compatible pair of pheromone and receptor or (ii) a constitutively active receptor that bypasses the need for a pheromone. It is possible that if compatibility exists between the extant Ste3/Mfa pair, that their compatibility arose from mutation. Examples are known of mutations in the pheromone that allow it to activate a receptor of the same *PR* sub-locus (Riquelme *et al.* 2005; Kües 2015). Similarly, examples of mutations in the receptors that permanently activate the receptor without the need for a cognate pheromone also have been reported (Olesnicky *et al.* 1999; James *et al.* 2011; Kües 2015), indicating therefore that either case would be possible in *C. capitatum*. Due to the existence of only one genome sequence for *C. capitatum* it was not possible to ascertain if its Hd1 and Hd2 proteins display a polymorphism pattern like any of the other homothallic species studied and therefore to put forward a hypothesis regarding their evolution or their role in the homothallic behavior displayed by this species.

Rearrangements within *MAT* are expected to underlie the transitions between heterothallic and homothallic sexual behaviors and this has been demonstrated for several species of the phylum Ascomycota (Yun *et al.* 1999; Inderbitzin *et al.* 2005; Gioti *et al.* 2012). In this lineage, studies have also shown a polyphyletic origin of homothallism (Inderbitzin *et al.* 2005; Nygren *et al.* 2011; Gioti *et al.* 2012). If homothallism evolved from heterothallism, this transition would require more change in the complex *MAT* system typical of the basidiomycetes, than in Ascomycetes. In Cystofilobasidiales a powerful approach to study transitions in sexual behavior over evolutionary time would consist in comparing the *MAT* regions of homothallic species with those of heterothallic species, especially of

strains belonging to different mating types. Unfortunately, the draft genomes obtained for the heterothallic species *C. bisporidii*, *C. ferigula* and *C. macerans* lack the level of completeness that would allow for a more extensive synteny analysis of the *MAT* regions (Table IV.1 on Appendix IV and Table V.1 on Appendix V). Nevertheless, the genomic data obtained for the different heterothallic species *Cystofilobasidium* was sufficient to show that pheromone receptors found in each species appear to have diverged from one another at different times, with *C. ferigula STE3* alleles being the earliest diverged and *C. macerans STE3* alleles being the most recently diverged (Figure 5.1 and Table 5.3). Specifically, while the *STE3* alleles from *C. ferigula* appear to have diverged prior to the emergence of the other *Cystofilobasidium* species studied, the *STE3* alleles from *C. macerans* and *C. bisporidii* seem to have diverged within each of the species themselves. This indicates that there were several events in which new *STE3* alleles appeared in a given species. By only analyzing two strains, for the heterothallic species, it is not possible to ascertain if more *STE3* alleles exist in these species. However that could be a possible scenario explaining the presence of more than one mature pheromone in each of the different mating types analyzed, like in the case of *Sporisorium reilianum*, which presents a triallelic *PR* locus, where in each mating type, two distinct *MFA* genes are encoded in the *PR* locus (Schirawski *et al.* 2005). The genomic data also allowed the discovery of shared mature pheromones in strains of distinct mating types of *C. macerans* (Figure 5.2). Taking together all data, it is possible that the number of mature pheromones is not related to the existence of more than two receptors per species, but rather to an idiosyncrasy common to several species in the Cystofilobasidiales (Table 5.4). The presence of identical pheromones in the two different mating types of *C. macerans* may suggest an intermediate situation between a heterothallic and homothallic system. Although the two *C. macerans* strains, CBS 6532 and CBS 2425, are from distinct mating types, *A1* and *A2* respectively, they co-exist in the same species with homothallic strains. According to Libkind et al. (2009), homothallic strains of this species exhibit much higher sexual competence than heterothallic strains. Additionally, CBS 6532 and CBS 2425, also show a distinctive polymorphism pattern when their *Hd1* and *Hd2* proteins are compared, displaying not only a highly variable N-terminal region but also a much less conserved homeodomain and C-terminal regions (Figure 5.4). The investigation of additional strains, both heterothallic and homothallic, will be required to fully understand the dynamics of *MAT* gene evolution, thereby shedding light on whether it may represent an intermediate situation between homothallism and heterothallism.

5.4.4. *MAT* loci in other members of the Cystofilobasidiales

In *Tausonia* and *Udeniomyces* species, for which no sexual state is known, the presence of this array of *MAT* genes, suggests that these organisms have the potential for sexual reproduction. Interestingly, for *T. pamirica* and *U. megalosporus* it is also apparent that the *STE3* alleles present in each species have diverged after speciation, similarly to what was observed for *C. macerans* or *C. bisporidii* in the *Cystofilobasidium* clade (Figure 5.1).

Krasilnikovozyma huempii (PYCC 5836) is the only species among those studied for which the set of *MAT* genes appears incomplete, given that no *MFA* genes were found in the draft genome. Although an initial report by Bab'eva et al. (2002) indicated that *Mrakia curviuscula,* as it was named then, was homothallic and presented sexual structures, namely teliospores (Bab'eva *et al.* 2002), this report was later refuted by suggestions that the structures previously identified as teliospores were chlamydospores or large vegetative cells (Kurtzman *et al.* 2011a). Further examination of this genus both at the genomic and the phenotypic level will be required to unequivocally determine if there is in fact sexual reproduction.

In general, the *PR* regions of all the Cystofilobasidiales studied show some conservation of gene content and between more closely related species even some synteny (Figure 5.5). In contrast, the genetic content of the neighboring regions of the *HD* genes found in the genera *Mrakia*, *Tausonia*, *Udeniomyces* and *Krasilnikovozyma* vary greatly in gene content (Figure 5.7). These observations differ from what has been found for the *HD* regions of several Agaricomycetes where conserved synteny near the *HD* locus was found between *Pleurotus djamor*, *Coprinopsis cinerea*, *Coprinellus disseminatus* and *Phanerochaete chysosporium* (James *et al.* 2011). Interestingly, only within the genera *Phaffia* and *Cystofilobasidium* is synteny present at the *HD* loci. These results may be indicative that while the *PR* loci of the different Cystofilobasidiales species have more genes in common than the *STE3* and *MFA* genes, the *HD* loci, in contrast may be restricted to the *HD* genes therefore displaying less conservation of gene content in the neighboring regions.

5.5. Conclusion

Within Cystofilobasidiales homothallism is widespread (Kurtzman *et al.* 2011a; Liu *et al.* 2015a; Liu *et al.* 2015b), being present in at least three distinct genera. The results indicate that transitions to homothallism probably occurred several times independently in this lineage and reveal a particularly dynamic pattern of *MAT* gene evolution, with high turnover of pheromone receptor alleles and exceptionally large numbers of mature pheromones encoded in the genomes of some of the species, meaning that compatibility between receptors and pheromones also had to evolve *de novo* several times. These assumptions imply a great plasticity of these *MAT* regions in the Cystofilobasidiales. Given that the two compatibility systems that characterize the Basidiomycetes, the *PR* an *HD* loci, may achieve self-compatibility independently at different times it is possible that this plasticity at the *MAT* regions allows for the existence of organisms with genetically transient states that while maintaining a heterothallic mating behavior are one step closer to homothallism.

The prevalence of homothallism within Cystofilobasidiales seems more likely to be due to a genetic background prone to the development of self-compatible *PR* or *HD* systems than, for example, to convergent evolution of all these genera towards homothallism due to similar ecological pressures. Although some of the homothallic species studied share certain common characteristics e.g. both *P. rhodozyma* and *C. capitatum* have been isolated from fruiting bodies of a *Cyttaria hariotii* in South America, or both *P. rhodozyma* and *M. frigida* are found in cold environments (Kurtzman *et al.* 2011a); no discernible evolutionary pressure appears common to these homothallic species. Nevertheless, the

homothallic behavior of these species, most certainly allows them some kind of ecological advantage, given its prevalence in this taxonomical order. Homothallism has been shown to encompass a great variety of genetic mechanisms that ultimately allow one single cell to reproduce sexually, taking advantage of most of the benefits allotted to this type of reproduction without the need for a compatible sexual partner (Lin & Heitman 2007; Roach *et al.* 2014; Wilson *et al.* 2015b). The homothallic species encompassed within the Cystofilobasidiales appear to follow this trend showing that somewhat different genetic makeups can lead to the same sexual behavior: homothallism.

CHAPTER 6

6. Concluding remarks and future perspectives

6.1. Concluding remarks and future perspectives

Sexual reproduction is a pervasive trait throughout eukaryotes even though it presents not only benefits but also costs, when compared with asexual reproduction (Maynard-Smith 1978). The main costs pertaining to sexual reproduction are (i) the need to find compatible partners in order to mate (ii) the conveyance of only 50% of the genetic content of each parental strain to the offspring and (iii) the possibility of breaking apart well adapted genomic configurations by undergoing sexual reproduction and hence recombination (Maynard-Smith 1978; Ni *et al.* 2011; Heitman *et al.* 2013; López-Villavicencio *et al.* 2013). Nevertheless, the benefits of sexual reproduction appear to outweigh its costs given its ubiquity among eukaryotes. Sexual reproduction allows the generation of *de novo* diversity in the progeny by mixing the genetic information of the parental strains through recombination. Whereas in the short term this enables organisms to cope with environmental challenges (Hamilton *et al.* 1990; Jokela *et al.* 2009; de Vienne *et al.* 2013; Heitman *et al.* 2013), in the long term the reshuffling of the parental genomic information provided by recombination avoids the irreversible accumulation of deleterious mutations in the genome, hence avoiding the so called Muller's "ratchet" (Muller 1964). However, the maintenance of the capability for sexual reproduction in homothallic organisms poses several questions regarding the balance between costs and benefits of sex. *Phaffia rhodozyma* reproduces mainly clonally, showing little evidence of recombination between different strains/populations (David-Palma *et al.* 2014). Nevertheless, it was demonstrated that not only does the species have *MAT* genes, as well as genes enabling meiosis (Chapter 2), but also that these are fundamental for sexual reproduction (Chapter 3). In the short term, *P. rhodozyma* homothallic sexual behavior, also designated as haploid selfing (Billiard *et al.* 2011), presents some advantages when compared to a heterothallic mating behavior, because it mitigates several of the costs associated with sexual reproduction (Lee *et al.* 2010; López-Villavicencio *et al.* 2013). Firstly, it precludes the need to find a compatible partner, especially when the basidia (and thus the basidiospores) are formed through pedogamy or endoreplication occurring in a single cell. Secondly, 100% of the parental genes are transmitted to the progeny, which in turn eliminates the cost of sex associated with breaking apart well-adapted genomic configurations. However questions remain as to how homothallic sexual cycles could insure the continuation of a homothallic species in the long term, given that parental strains are in many cases isogenic, representing an extreme limitation to recombination (Taylor *et al.* 2015). For long, homothallism was believed to ultimately lead to the extinction of the species, representing a so called "evolutionary dead end" (Bell 1982; Gioti *et al.* 2013b; López-Villavicencio *et al.* 2013), similarly to what was hypothesized for species reproducing only asexually (Roach & Heitman 2014). Some insights regarding how species with homothallic sexual cycles may escape this "evolutionary dead end", namely trough the generation of diversity *de novo*, or even the ability to avoid Muller's "ratchet", started to appear in the last few years, through studies conducted in the pathogenic basidiomycete, *Cryptococcus deneoformans* (Ni *et al.* 2013). *C. deneoformans*, is capable of heterothallic mating behavior between cells of distinct mating types (a and α) and additionally presents a homothallic mating behavior, named unisexual mating. Such homothallic behavior is characterized by the fusion of cells from the same mating type (either clones

or cells from distinct strains but with the same mating type) or endoreplication of the nucleus of a cell, followed by meiosis and the formation of basidiospores (Lin *et al.* 2005; Feretzaki & Heitman 2013a, b). It has been experimentally demonstrated that even when unisexual reproduction occurs between genetically identical cells (clones), a limited amount of genetic diversity may be added to the genotype of the progeny indicating that sexual reproduction not only serves to mix preexisting genetic diversity, but that the process itself creates some level of genetic diversity (Feretzaki & Heitman 2013b; Ni *et al.* 2013). This diversity *de novo* may be present in the resulting progeny as aneuploidy, chromosomal length polymorphisms, deletions or single nucleotide polymorphisms (Feretzaki & Heitman 2013b; Ni *et al.* 2013). Additionally, it was also demonstrated that unisexual reproduction reverses Muller's "ratchet" (Roach & Heitman 2014). The homothallic mating behavior of *P. rhodozyma* is distinct from unisexual reproduction in various ways, namely because no mating types were found in the species (Chapter 2) and because both *HD1* and *HD2* genes are necessary to the normal completion of the *P. rhodozyma* life cycle (Chapter 3), which differs from unisexual mating where none of the homeodomain transcription factors are required (Roach *et al.* 2014). However, there may be also some resemblances between the two homothallic cycles. Similarly, to what was observed for unisexual mating between clones in *C. deneoformans*, it is possible that the homothallic sexual cycle of *P. rhodozyma* contributes to increase diversity. Some of the progeny obtained from clonal unisexual mating in *C. deneoformans* presented chromosomal length polymorphisms (Ni *et al.* 2013). If this were to happen in *P. rhodozyma* it would explain why different studies have found different karyotypes for the same strain, like for strain ATCC 24229 which was reported to have seven chromosomes in one study (Nagy *et al.* 1994) and nine in another (Nagy *et al.* 1997). The same was reported for other strains in different studies, and it seems unlikely that technical limitations associated with PFGE, like co-migration events, could account for all the discrepancies (Nagy *et al.* 1994; Adrio *et al.* 1995; Cifuentes *et al.* 1997; Nagy *et al.* 1997), in which case the differences may conceivably have been caused by rounds of sexual reproduction in between karyotyping experiments. In *P. rhodozyma*, aneuploidy has also been inferred which can also be related to the homothallic life cycle of this species (Kucsera *et al.* 1998; Medwid 1998), potentially generating diversity *de novo* as in the unisexual reproducing *C. deneoformans* strains. Similarly to what was observed in the homothallic ascomycete *Aspergillus nidulans* (López-Villavicencio *et al.* 2013) and other homothallic fungi, *P. rhodozyma* can also outcross, even if at a low frequency (Kucsera *et al.* 1998; David-Palma *et al.* 2014). By doing so, *P. rhodozyma* can take advantage of the benefits of asexual reproduction, haploid selfing and outcrossing.

Like most scientific endeavors, this work answered many of the initial questions but at the same time became the starting point of a whole new set of interrogations. Although it was proven that both Hd1 and Hd2 proteins are necessary for the normal completion of the homothallic sexual behavior of *P. rhodozyma* (Chapter 3), further research will be needed to uncover the nature of the interaction between the homeodomain proteins, with a detailed dissection of the functional domains present in both proteins. Bacterial two-hybrid assays were conducted with the complete Hd1 and Hd2 proteins of strain CBS 6938, but no interaction was detected. Assays were also performed in an alternative

system, yeast two-hybrid assay, focusing the experiments on the N-terminal regions of the homeodomain proteins that are, for most basidiomycetes, the regions associated with heterodimerization and in this case no clear interaction was observed between the Hd1-Hd2 pairs (Chapter 3). Although alpha-helix motifs were predicted in the N-terminal regions of both Hd proteins, the presence of coiled-coils was only predicted in the C-terminal region of Hd1 proteins in all of *Phaffia* species. Given that such structures are usually associated with oligomerisation (Kües & Casselton 1992; Kües *et al.* 1992; Parry *et al.* 2008; Kües *et al.* 2011) and the mechanism of self-discrimination in most heterothallic basidiomycetes, it is possible that the interaction between the Hd proteins in this species involves the C-terminal domain, unlike the Hd proteins of other basidiomycetes. Further yeast two-hybrid assays focusing on the C-terminal regions of the transcription factor proteins may provide further insights into the interaction between these proteins. Additionally, sequencing of *HD* genes derived from additional *P. tasmanica* and *P. novazelandica* strains would be useful, in order to see if these species share the same pattern of sequence diversity observed in the homeodomain proteins of *P. rhodozyma.* The study of the interaction between homeodomain proteins in other homothallic species would also be extremely interesting considering the different patterns of polymorphism that were found in Hd proteins of different homothallic species, i.e. *P. rhodozyma* and *M. blollopis* (Chapter 5). Such a study could provide insights into the evolution of these genes and on how these self-compatible systems appeared.

The phylogenies of the pheromone receptors of the different species of the order Cystofilobasidiales showed a departure from the trans-specific polymorphism present in most of Tremellomycetes species, implying that the evolution of the *PR* loci and perhaps of the *MAT* regions in this group as a whole is different from what has been described so far for basidiomycetes (Chapter 5). Which selective forces could drive the creation of distinct receptors in so many genera? At the same time, although new Ste3 alleles appear to be created in several of the genera, their number does not appear to exceed two per strain (Chapter 5), which departs from what is observed in the Agaricomycetes (Kües 2015). Also intriguing is what could lead to the existence of so many different mature pheromones in the species within Cystofilobasidiales. In order to further explore the evolution of the *MAT* regions of these species more information is necessary, regarding for example the possible existence of additional Ste3 alleles within some of the species, like in *Cystofilobasidium* species and in *M. blollopis*. Sequencing of more *STE3* genes from strains of *C. macerans*, *C. bisporidii* and *C. ferigula* could provide further clues on the evolution of the *PR* regions. In fact, sequencing of *STE3* genes throughout the entire set of species within the Cystofilobasidiales would be interesting, considering the results obtained in the phylogenetic inferences performed in Chapter 2 and Chapter 5.

Moreover, to try to understand the direction of the transitions between heterothallic and homothallic species within the Cystofilobasidiales, it will be necessary to improve the quality of the genomic information of the heterothallic strains, so that a more extensive comparison may be performed between the homothallic *Phaffia* species and the heterothallic *Cystofilobasidium* species. Taking advantage of the genomic information already obtained in this work (Chapter 4 and 5) and enriching it with genomic information obtained using other available technologies, such as single molecule sequencing by Pacific Biosciences (PacBio) that delivers very long kb-sized reads, it will be possible to

obtain more complete assemblies of these genomes and, in doing so, shedding further light on extant *MAT* organization in *Phaffia* species, and on the molecular mechanism underlying transitions between heterothallism and homothallism.

The work presented in this thesis could prove useful at both fundamental and applied scientific fields. To begin with, it elucidated for the first time the genetic basis of primary homothallism in a basidiomycete, which to date remained largely understudied. Secondly, the comparison of the *MAT* regions of several species within the Cystofilobasidiales suggested the existence a great plasticity of these regions and can be seen as a stepping stone in understanding other genomic arrangements that also translate into a homothallic life cycle. In addition, this study also contributed with additional genomic information of several species within Cystofilobasidiales, which allowed for a robust phylogenetic hypothesis of relationships between species and genera. Moreover, the draft genomes generated by this study include genomes of two new species of *Phaffia* that also produce astaxanthin. These new species increase the genomic diversity of the astaxanthin producing yeasts, which could prove a valuable resource for biotechnology. Given that *P. rhodozyma* has several biotechnological applications, the knowledge of the genes responsible for its sexual reproduction, may further open the range of strategies for *P. rhodozyma* exploitation, e.g. by the generation of different genetic mating types that may facilitate strain improvement through selective breeding strategies.

Phaffia rhodozyma and the Cystofilobasidiales order appear to be fertile ground to further explore sexual reproduction in the basidiomycetes at the same time providing new information about biotechnological relevant species.

REFERENCES

Aanen DK, Hoekstra RF (2007) Why Sex Is Good: On Fungi and Beyond. In: *Sex in Fungi*. American Society of Microbiology.

Adrio JL, López M, Casqueiro J, Fernández C, Veiga M (1995) Electrophoretic karyotype of the astaxanthin-producing yeast *Phaffia rhodozyma*. *Current Genetics* **27**, 447-450.

Adrio JL, Veiga M (1995) Transformation of the astaxanthin-producing yeast *Phaffia rhodozyma*. *Biotechnology Techniques* **9**, 509-512.

Adrio JOS, Eacute, Luis*, et al.* (1993) Isolation of *Phaffia rhodozyma* auxotrophic mutants by enrichment methods. *The Journal of General and Applied Microbiology* **39**, 303-312.

Alby K, Schaefer D, Bennett RJ (2009) Homothallic and heterothallic mating in the opportunistic pathogen *Candida albicans*. *Nature* **460**, 890-893.

Alva V, Nam SZ, Soding J, Lupas AN (2016) The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res* **44**, W410-415.

Amselem J, Cuomo CA, van Kan JAL*, et al.* (2011) Genomic Analysis of the Necrotrophic Fungal Pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet* **7**, e1002230.

Andrewes AG, Phaff HJ, Starr MP (1976) Carotenoids of *Phaffia rhodozyma*, a red-pigmented fermenting yeast. *Phytochemistry* **15**, 1003-1007.

Asante-Owusu RN, Banham AH, Bohnert HU, Mellor EJ, Casselton LA (1996) Heterodimerization between two classes of homeodomain proteins in the mushroom *Coprinus cinereus* brings together potential DNA-binding and activation domains. *Gene* **172**, 25-31.

Astell CR, Ahlstrom-Jonasson L, Smith M*, et al.* (1981) The sequence of the DNAs coding for the mating-type loci of *Saccharomyces cerevisiae*. *Cell* **27**, 15-23.

Bab'eva IP, Lisichkina GA, Reshetova IS, Danilevich VN (2002) *Mrakia curviuscula* sp. nov.: A New Psychrophilic Yeast from Forest Substrates. *Microbiology* **71**, 449-454.

Badrane H, May G (1999) The divergence-homogenization duality in the evolution of the *b1* mating type gene of *Coprinus cinereus*. *Mol Biol Evol* **16**, 975-986.

Baldauf SL, Palmer JD (1993) Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc Natl Acad Sci U S A* **90**, 11558-11562.

Banham A, Asante-Owusu R, Gottgens B*, et al.* (1995) An N-terminal dimerization domain permits homeodomain proteins to choose compatible partners and Initiate sexual development in the mushroom *Coprinus cinereus*. *Plant Cell* **7**, 773 - 783.

Bardwell L (2004) A walk-through of the yeast mating pheromone response pathway. *Peptides* **25**, 1465-1476.

Barnett JAP, R. W.; Yarrow, D. (2000) *Yeasts: Characteristics and Identification*, 3rd Edition edn. Cambridge University Press Cambridge.

Battesti A, Bouveret E (2012) The bacterial two-hybrid system based on adenylate cyclase reconstitution in *Escherichia coli*. *Methods* **58**, 325-334.

Bell G (1982) *The Masterpiece of Nature :: The Evolution and Genetics of Sexuality* Springer US.

Bellora N, Moliné M, David-Palma M*, et al.* (2016) Comparative genomics provides new insights into the diversity, physiology, and sexuality of the only industrially exploited tremellomycete: *Phaffia rhodozyma*. *BMC Genomics* **17**, 901.

Bennett RJ, Johnson AD (2003) Completion of a parasexual cycle in *Candida albicans* by induced chromosome loss in tetraploid strains. *EMBO J* **22**, 2505-2515.

Bennett RJ, Turgeon BG (2016) Fungal Sex: The Ascomycota. *Microbiology Spectrum* **4**.

Billiard S, Lopez-Villavicencio M, Devier B*, et al.* (2011) Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. *Biol Rev* **86**, 421 - 442.

Billiard S, Lopez-Villavicencio M, Hood ME, Giraud T (2012) Sex, outcrossing and mating types: unsolved questions in fungi and beyond. *J Evol Biol* **25**, 1020-1038.

Blakeslee AF (1904) Sexual reproduction in the Mucorineae. *Proc. Am. Acad. Arts Sci.* **40**, 205–319.

Boekhout T (1991) *A revision of ballistoconidia-forming yeasts and fungi* Centralbureau voor Schimmelcultures.

Boekhout T, Fonseca Á, Sampaio JP*, et al.* (2011) Discussion of Teleomorphic and Anamorphic Basidiomycetous Yeasts. In: *The Yeasts* (eds. Boekhout T, Kurtzman CP, Fell JW), pp. 1339-1372. Elsevier, London.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120.

Bölker M, Urban M, Kahmann R (1992) The a mating type locus of *U. maydis* specifies cell signaling components. *Cell* **68**, 441-450.

Breitenbach J, Visser H, Verdoes JC, van Ooyen AJ, Sandmann G (2011) Engineering of geranylgeranyl pyrophosphate synthase levels and physiological conditions for enhanced

carotenoid and astaxanthin synthesis in *Xanthophyllomyces dendrorhous*. *Biotechnol Lett* **33**, 755-761.
Butler G, Kenny C, Fagan A*, et al.* (2004) Evolution of the *MAT* locus and its Ho endonuclease in yeast species. *Proc Natl Acad Sci U S A* **101**, 1632-1637.
Butler G (2007) The evolution of *MAT*: The Ascomycetes. In: *Sex in Fungi* (ed. Heitman J), pp. 3-18. ASM Press, Washington.
Caldwell GA, Naider F, Becker JM (1995) Fungal lipopeptide mating pheromones: a model system for the study of protein prenylation. *Microbiological Reviews* **59**, 406-422.
Callac P, Jacobe de Haut I, Imbernon M*, et al.* (2003) A novel homothallic variety of *Agaricus bisporus* comprises rare tetrasporic isolates from Europe. *Mycologia* **95**, 222-231.
Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973.
Casselton L, Olesnicky N (1998) Molecular genetics of mating recognition in basidiomycete fungi. *Microbiol Mol Biol Rev* **62**, 55 - 70.
Casselton L (2002) Mate recognition in fungi. *Heredity* **88**, 142 - 147.
Casselton L, Kues U (2007) The origin of multiple mating types in the model mushrooms *Coprinopsis cinerea* and *Schizophyllum commune*. In: *Sex in fungi: Molecular determination and evolutionary implications*, pp. 283 - 300.
Casselton LA (2008) Fungal sex genes-searching for the ancestors. *BioEssays* **30**, 711-714.
Chen RE, Thorner J (2007) Function and regulation in MAPK signaling pathways: lessons learned from the yeast *Saccharomyces cerevisiae*. *Biochim Biophys Acta* **1773**, 1311-1340.
Chen Z, Li J (2004) Phylogenetics and Biogeography of *Alnus* (Betulaceae) Inferred from Sequences of Nuclear Ribosomal DNA *ITS* Region. *International Journal of Plant Sciences* **165**, 325-335.
Chen ZD, Manchester SR, Sun HY (1999) Phylogeny and evolution of the Betulaceae as inferred from DNA sequences, morphology, and paleobotany. *Am J Bot* **86**, 1168-1181.
Cifuentes V, Hermosilla G, Martinez C*, et al.* (1997) Genetics and electrophoretic karyotyping of wild-type and astaxanthin mutant strains of *Phaffia rhodozyma Antonie Van Leeuwenhoek* **72**, 111-117.
Coelho M, Rosa A, Rodrigues N, Fonseca A, Goncalves P (2008) Identification of mating type genes in the bipolar basidiomycetous yeast *Rhodosporidium toruloides:* first insight into the *MAT* locus structure of the Sporidiobolales. *Eukaryot Cell* **7**, 1053 - 1061.
Coelho M, Goncalves P, Sampaio J (2011) Evidence for maintenance of sex determinants but not of sexual stages in red yeasts, a group of early diverged basidiomycetes. *BMC Evolutionary Biology* **11**, 249.
Coelho MA, Sampaio JP, Goncalves P (2010) A deviation from the bipolar-tetrapolar mating paradigm in an early diverged basidiomycete. *PLoS Genet* **6**.
David-Palma M, Libkind D, Sampaio JP (2014) Global distribution, diversity hot spots and niche transitions of an astaxanthin-producing eukaryotic microbe. *Mol Ecol* **23**, 921-932.
David-Palma M, Sampaio JP, Gonçalves P (2016) Genetic Dissection of Sexual Reproduction in a Primary Homothallic Basidiomycete. *PLoS Genet* **12**, e1006110.
Davidson RC, Nichols CB, Cox GM, Perfect JR, Heitman J (2003) A MAP kinase cascade composed of cell type specific and non-specific elements controls mating and differentiation of the fungal pathogen *Cryptococcus neoformans*. *Mol Microbiol* **49**, 469-485.
de Vienne DM, Giraud T, Gouyon P-H (2013) Lineage Selection and the Maintenance of Sex. *PLoS ONE* **8**, e66906.
Devier B, Aguileta G, Hood M, Giraud T (2009) Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. *Genetics* **181**, 209 - 223.
Diaz-Valderrama JR, Aime MC (2016) The cacao pathogen *Moniliophthora roreri* (Marasmiaceae) possesses biallelic A and B mating loci but reproduces clonally. *Heredity (Edinb)* **116**, 491-501.
Dodge BO (1931) Heterothallism and Hypothetical Hormones in Neurospora. *Bulletin of the Torrey Botanical Club* **58**, 517-522.
Dominguez-Bocanegra AR, Torres-Munoz JA (2004) Astaxanthin hyperproduction by *Phaffia rhodozyma* (now *Xanthophyllomyces dendrorhous*) with raw coconut milk as sole source of energy. *Appl Microbiol Biotechnol* **66**, 249-252.
Drozdetskiy A, Cole C, Procter J, Barton GJ (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* **43**, W389-394.
Ene IV, Bennett RJ (2014) The Cryptic Sexual Strategies of Human Fungal Pathogens. *Nature reviews. Microbiology* **12**, 239-251.

Fell J, Blatt G (1999) Separation of strains of the yeasts *Xanthophyllomyces dendrorhous* and *Phaffia rhodozyma* based on rDNA *IGS* and *ITS* sequence analysis. *J Ind Microbiol Biotechnol* **23**.

Fell JW, Hunter IL, Tallman AS (1973) Marine basidiomycetous yeasts (*Rhodosporidium spp*. n.) with tetrapolar and multiple allelic bipolar mating systems. *Can J Microbiol* **19**, 643-657.

Feretzaki M, Heitman J (2013a) Genetic circuits that govern bisexual and unisexual reproduction in *Cryptococcus neoformans*. *PLoS Genet* **9**, e1003688.

Feretzaki M, Heitman J (2013b) Unisexual reproduction drives evolution of eukaryotic microbial pathogens. *PLoS Pathog* **9**, e1003674.

Findley K, Rodriguez-Carres M, Metin B*, et al.* (2009) Phylogeny and phenotypic characterization of pathogenic *Cryptococcus* species and closely related saprobic taxa in the Tremellales. *Eukaryot Cell* **8**, 353-361.

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic acids research* **39**, W29-W37.

Fontanillas E, Hood ME, Badouin H*, et al.* (2015) Degeneration of the nonrecombining regions in the mating-type chromosomes of the anther-smut fungi. *Mol Biol Evol* **32**, 928-943.

Fowler T, Mitton M, Vaillancourt L, Raper C (2001) Changes in mate recognition through alterations of pheromones and receptors in the multisexual mushroom fungus *Schizophyllum commune*. *Genetics* **158**, 1491 - 1503.

Fowler TJ, Mitton MF, Rees EI, Raper CA (2004) Crossing the boundary between the Balpha and Bbeta mating-type loci in *Schizophyllum commune*. *Fungal Genet Biol* **41**, 89-101.

Fraser J, Heitman J (2003) Fungal mating-type loci. *Curr Biol* **13**, R792 - 795.

Fraser J, Heitman J (2004) Evolution of fungal sex chromosomes. *Mol Microbiol* **51**, 299 - 306.

Fraser J, Heitman J (2005) Chromosomal sex-determining regions in animals, plants and fungi. *Curr Opin Genet Dev* **15**, 645 - 651.

Fraser JA, Diezmann S, Subaran RL*, et al.* (2004) Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLoS Biol* **2**, e384.

Fraser JA, Hsueh YP, Findley KM, Heitman J (2007) Evolution of the mating-type locus: the basidiomycetes. In: *Sex in fungi: Molecular determination and evolutionary implications.* (eds. Heitman J, Kronstad JW, Taylor JW, Casselton LA), pp. 19-34. ASM Press, Washington DC.

Freihorst D, Fowler TJ, Bartholomew K*, et al.* (2016) 13 The Mating-Type Genes of the Basidiomycetes. In: *Growth, Differentiation and Sexuality* (ed. Wendland J), pp. 329-349. Springer International Publishing, Cham.

Galagan JE, Calvo SE, Cuomo C*, et al.* (2005) Sequencing of Aspergillus nidulans and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**, 1105-1115.

Gassel S, Breitenbach J, Sandmann G (2014) Genetic engineering of the complete carotenoid pathway towards enhanced astaxanthin formation in *Xanthophyllomyces dendrorhous* starting from a high-yield mutant. *Appl Microbiol Biotechnol* **98**, 345-350.

Gioti A, Mushegian AA, Strandberg R, Stajich JE, Johannesson H (2012) Unidirectional evolutionary transitions in fungal mating systems and the role of transposable elements. *Mol Biol Evol* **29**, 3215-3226.

Gioti A, Nystedt B, Li W*, et al.* (2013a) Genomic insights into the atopic eczema-associated skin commensal yeast *Malassezia sympodialis*. *MBio* **4**, e00572-00512.

Gioti A, Stajich JE, Johannesson H (2013b) *Neurospora* and the dead-end hypothesis: Genomic consequences of selfing in the model genus. *Evolution* **67**, 3600-3616.

Glass NL, Smith ML (1994) Structure and function of a mating-type gene from the homothallic species *Neurospora africana*. *Molecular and General Genetics MGG* **244**, 401-409.

Golubev VI, Bab'eva IP, Blagodatskaia VM, Reshetova IS (1977) Taxonomic study of yeast organisms isolated from the sap of birches (*Betula verrucosa* Ehrh.). *Mikrobiologiia* **46**, 564-569.

Golubev WI (1995) Perfect state of *Rhodomyces dendrorhous* (*Phaffia rhodozyma*). *Yeast* **11**, 101-110.

Griffith GW, Hedger JN (1994a) The breeding biology of biotypes of the witches' broom pathogen of cocoa, *Crinipellis perniciosa*. *Heredity* **72**, 278-289.

Griffith GW, Hedger JN (1994b) Spatial distribution of mycelia of the liana (L-) biotype of the agaric *Crinipellis perniciosa* (Stahel) Singer in tropical forest. *New Phytologist* **127**, 243-259.

Guindon S, Dufayard J-F, Lefort V*, et al.* (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**, 307-321.

Halary S, Malik SB, Lildhar L*, et al.* (2011) Conserved meiotic machinery in *Glomus spp.*, a putatively ancient asexual fungal lineage. *Genome Biol Evol* **3**, 950-958.

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**, 95 - 98.

Hamilton WD, Axelrod R, Tanese R (1990) Sexual Reproduction as an Adaptation to Resist Parasites (A Review). *Proceedings of the National Academy of Sciences of the United States of America* **87**, 3566-3573.

Hanson SJ, Byrne KP, Wolfe KH (2014) Mating-type switching by chromosomal inversion in methylotrophic yeasts suggests an origin for the three-locus *Saccharomyces cerevisiae* system. *Proceedings of the National Academy of Sciences* **111**, E4851-E4858.

Hara KY, Morita T, Endo Y*, et al.* (2014a) Evaluation and screening of efficient promoters to improve astaxanthin production in *Xanthophyllomyces dendrorhous*. *Appl Microbiol Biotechnol* **98**, 6787-6793.

Hara KY, Morita T, Mochizuki M*, et al.* (2014b) Development of a multi-gene expression system in *Xanthophyllomyces dendrorhous*. *Microb Cell Fact* **13**, 175.

Heitman J (2006) Sexual Reproduction and the Evolution of Microbial Pathogens. *Current Biology* **16**, R711-R725.

Heitman J, Sun S, James TY (2013) Evolution of fungal sexual reproduction. *Mycologia* **105**.

Heitman J (2015) Evolution of sexual reproduction: A view from the fungal kingdom supports an evolutionary epoch with sex before sexes. *Fungal Biology Reviews* **29**, 108-117.

Hermosilla G, Martinez C, Retamales P, Leon R, Cifuentes V (2003) Genetic determination of ploidy level in *Xanthophyllomyces dendrorhous*. *Antonie Van Leeuwenhoek* **84**, 279-287.

Herskowitz I, Rine J, Strathern J (1992) Mating-type Determination and Mating-type Interconversion in *Saccharomyces cerevisiae*. In: *The Molecular and Cellular Biology of the Yeast Saccharomyces* (eds. Broach JR, Pringle JR, Jones EW), pp. 583-656. Cold Spring Harbor Laboratory Press, New York.

Hibbett D, Binder M, Bischoff J*, et al.* (2007) A higher-level phylogenetic classification of the Fungi. *Mycol Res* **111**, 509 - 547.

Hicks JB, Herskowitz I (1977) Interconversion of yeast mating types. II. Restoration of mating ability to sterile mutants in homothallic and heterothallic strains. *Genetics* **85**, 373-393.

Hicks JB, Strathern JN, Herskowitz I (1977) Interconversion of yeast mating types. III. Action of the homothallism (HO) gene in cells homozygous for the mating type locus. *Genetics* **85**, 395-405.

Hsueh Y-P, Xue C, Heitman J (2007) G protein signaling governing cell fate decisions involves opposing Gα subunits in *Cryptococcus neoformans*. *Molecular Biology of the Cell* **18**, 3237-3249.

Hsueh YP, Shen WC (2005) A homolog of Ste6, the a-factor transporter in *Saccharomyces cerevisiae*, is required for mating but not for monokaryotic fruiting in *Cryptococcus neoformans*. *Eukaryot Cell* **4**, 147-155.

Hull C, Heitman J (2002) Genetics of *Cryptococcus neoformans*. *Annu Rev Genet* **36**, 557 - 615.

Hull C, Boily M, Heitman J (2005) Sex-specific homeodomain proteins Sxi1alpha and Sxi2a coordinately regulate sexual development in *Cryptococcus neoformans*. *Eukaryot Cell* **4**, 526 - 535.

Idnurm A (2011a) Sex Determination in the First-Described Sexual Fungus. *Eukaryotic Cell* **10**, 1485-1491.

Idnurm A (2011b) Sex and speciation: the paradox that non-recombining DNA promotes recombination. *Fungal Biol Rev* **25**, 121-127.

Idnurm A, Hood ME, Johannesson H, Giraud T (2015) Contrasted patterns in mating-type chromosomes in fungi: Hotspots versus coldspots of recombination. *Fungal Biology Reviews*.

Inderbitzin P, Harkness J, Turgeon BG, Berbee ML (2005) Lateral transfer of mating system in *Stemphylium*. *Proc Natl Acad Sci U S A* **102**, 11390-11395.

James TY, Kauff F, Schoch CL*, et al.* (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**, 818-822.

James TY, Lee M, van Diepen LT (2011) A single mating-type locus composed of homeodomain genes promotes nuclear migration and heterokaryosis in the white-rot fungus *Phanerochaete chrysosporium*. *Eukaryot Cell* **10**, 249-261.

Jinno S, Hata K, Shimidzu N, Okita T (1998) Phaffiaol, a new antioxidant isolated from a yeast *Phaffia rhodozyma*. *J Antibiot* **51**.

Johnson EA, Lewis MJ (1979) Astaxanthin formation by the yeast *Phaffia rhodozyma*. *Journal of General Microbiology* **115**, 173-183.

Jokela J, Dybdahl MF, Lively CM (2009) The maintenance of sex, clonal dynamics, and host-parasite coevolution in a mixed population of sexual and asexual snails. *Am Nat* **174 Suppl 1**, S43-53.

Kämper J, Reichmann M, Romeis T, Bolker M, Kahmann R (1995) Multiallelic recognition: nonself-dependent dimerization of the *bE* and *bW* homeodomain proteins in *Ustilago maydis*. *Cell* **81**, 73-83.

Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772-780.

Katoh K, Standley DM (2014) MAFFT: iterative refinement and additional methods. *Methods Mol Biol* **1079**, 131-146.

Keeney S (2008) Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. *Genome dynamics and stability* **2**, 81-123.

Kerrigan RW, Royer JC, Baller LM*, et al.* (1993) Meiotic behavior and linkage relationships in the secondarily homothallic fungus *Agaricus bisporus*. *Genetics* **133**, 225-236.

Knapp M, Stöckler K, Havell D*, et al.* (2005) Relaxed Molecular Clock Provides Evidence for Long-Distance Dispersal of *Nothofagus* (Southern Beech). *PLoS Biol* **3**, e14.

Koltin Y, Stamberg J, Lemke PA (1972) Genetic structure and evolution of the incompatibility factors in higher fungi. *Bacteriological Reviews* **36**, 156-171.

Kronstad JW, Staben C (1997) Mating type in filamentous fungi. *Annu Rev Genet* **31**, 245-276.

Kucsera J, Pfeiffer I, Ferenczy L (1998) Homothallic life cycle in the diploid red yeast *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*). *Antonie Van Leeuwenhoek* **73**, 163-168.

Kucsera J, Pfeiffer I, Takeo K (2000) Biology of the red yeast *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*). *Mycoscience* **41**, 195-199.

Kues U, Navarro-Gonzalez M (2010) Mating-type orthologous genes in the primarily homothallic *Moniliophthora perniciosa*, the causal agent of Witches' Broom Disease in cacao. *J Basic Microbiol* **50**, 442-451.

Kües U, Casselton LA (1992) Homeodomains and regulation of sexual development in basidiomycetes. *Trends Genet* **8**, 154-155.

Kües U, Richardson W, Tymon A*, et al.* (1992) The combination of dissimilar alleles of the Aalpha and Abeta gene complexes, whose proteins contain homeo domain motifs, determines sexual development in the mushroom *Coprinus cinereus*. *Genes Dev* **6**, 568 - 577.

Kües U, Gottgens B, Stratmann R*, et al.* (1994) A chimeric homeodomain protein causes self-compatibility and constitutive sexual development in the mushroom *Coprinus cinereus*. *EMBO J* **13**, 4054-4059.

Kües U, James T, Heitman J (2011) Mating Type in Basidiomycetes: Unipolar, Bipolar, and Tetrapolar Patterns of Sexuality. In: *Evolution of Fungi and Fungal-Like Organisms* (eds. Pöggeler S, Wöstemeyer J), pp. 97-160. Springer Berlin Heidelberg.

Kües U (2015) From two to many: Multiple mating types in Basidiomycetes. *Fungal Biology Reviews* **29**, 126-166.

Kurtzman C, Fell JW, Boekhout T (2011a) *The Yeasts: A Taxonomic Study* Elsevier Science.

Kurtzman CP, Fell JW, Boekhout T, Robert V (2011b) Chapter 7 - Methods for Isolation, Phenotypic Characterization and Maintenance of Yeasts. In: *The Yeasts (Fifth Edition)*, pp. 87-110. Elsevier, London.

Kwon-Chung KJ (1975) A New Genus, Filobasidiella, the Perfect State of *Cryptococcus neoformans*. *Mycologia* **67**, 1197-1200.

Labarere J, Noel T (1992) Mating type switching in the tetrapolar basidiomycete *Agrocybe aegerita*. *Genetics* **131**, 307-319.

Ledetzky N, Osawa A, Iki K*, et al.* (2014) Multiple transformation with the *crtYB* gene of the limiting enzyme increased carotenoid synthesis and generated novel derivatives in *Xanthophyllomyces dendrorhous*. *Arch Biochem Biophys* **545**, 141-147.

Lee SC, Ni M, Li W, Shertz C, Heitman J (2010) The evolution of sex: a perspective from the fungal kingdom. *Microbiol Mol Biol Rev* **74**, 298 - 340.

Lemke PA (1969) A reevaluation of homothallism, heterothallism and the species concept in Sistotrema brinkmanni. *Mycologia* **60**, 57-76.

Lengeler KB, Fox DS, Fraser JA*, et al.* (2002) Mating-type locus of *Cryptococcus neoformans*: a step in the evolution of sex chromosomes. *Eukaryot Cell* **1**, 704-718.

Li L, Shen G, Zhang ZG*, et al.* (2007) Canonical heterotrimeric G proteins regulating mating and virulence of *Cryptococcus neoformans*. *Mol Biol Cell* **18**, 4201-4209.

Libkind D, Sommaruga R, Zagarese H, van Broock M (2005) Mycosporines in carotenogenic yeasts. *Systematic and Applied Microbiology* **28**, 749-754.

Libkind D, Ruffini A, Broock M, Alves L, Sampaio JP (2007) Biogeography, host specificity, and molecular phylogeny of the basidiomycetous yeast *Phaffia rhodozyma* and its sexual form, *Xanthophyllomyces dendrorhous*. *Appl Environ Microbiol* **73**.

Libkind D, Moline M, de Garcia V, Fontenla S, van Broock M (2008) Characterization of a novel South American population of the astaxanthin producing yeast Xanthophyllomyces dendrorhous (Phaffia rhodozyma). *J Ind Microbiol Biotechnol* **35**, 151-158.

Libkind D, Gadanho M, van Broock M, Sampaio JP (2009) *Cystofilobasidium lacus-mascardii* sp. nov., a basidiomycetous yeast species isolated from aquatic environments of the Patagonian Andes, and *Cystofilobasidium macerans* sp. nov., the sexual stage of *Cryptococcus macerans*. *International Journal of Systematic and Evolutionary Microbiology* **59**, 622-630.

Libkind D, Moline M, Broock M (2011a) Production of the UVB-absorbing compound mycosporine–glutaminol–glucoside by *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*). *FEMS Yeast Res* **11**.

Libkind D, Tognetti C, Ruffini A, Sampaio JP, Van Broock M (2011b) *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*) on stromata of *Cyttaria hariotii* in northwestern Patagonian *Nothofagus* forests. *Rev Argent Microbiol* **43**, 226-232.

Lin G, Bultman J, Johnson E, Fell J (2012) Genetic Manipulation of *Xanthophyllomyces dendrorhous* and *Phaffia rhodozyma*. In: *Microbial Carotenoids From Fungi* (ed. Barredo J-L), pp. 235-249. Humana Press.

Lin J-r, Hu J (2013) SeqNLS: Nuclear Localization Signal Prediction Based on Frequent Pattern Mining and Linear Motif Scoring. *PLoS ONE* **8**, e76864.

Lin X, Hull CM, Heitman J (2005) Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. *Nature* **434**, 1017-1021.

Lin X, Heitman J (2007) Mechanisms of homothallism in Fungi and transitions between heterothallism and homothallism. In: *Sex in fungi: Molecular determination and evolutionary implications.* (eds. Heitman J, Kronstad JW, Taylor JW, Casselton LA), pp. 35-57. ASM Press, Washington DC.

Lin X, Patel S, Litvintseva AP*, et al.* (2009) Diploids in the *Cryptococcus neoformans* Serotype A Population Homozygous for the *α* Mating Type Originate via Unisexual Mating. *PLoS Pathogens* **5**, e1000283.

Liu XZ, Wang QM, Goker M*, et al.* (2015a) Towards an integrated phylogenetic classification of the Tremellomycetes. *Stud Mycol* **81**, 85-147.

Liu XZ, Wang QM, Theelen B*, et al.* (2015b) Phylogeny of tremellomycetous yeasts and related dimorphic and filamentous basidiomycetes reconstructed from multiple gene sequence analyses. *Stud Mycol* **81**, 1-26.

López-Villavicencio M, Debets AJM, Slakhorst M, Giraud T, Schoustra SE (2013) Deleterious effects of recombination and possible nonrecombinatorial advantages of sex in a fungal model. *Journal of Evolutionary Biology* **26**, 1968-1978.

Loto I, Gutierrez MS, Barahona S*, et al.* (2012) Enhancement of carotenoid production by disrupting the C22-sterol desaturase gene (*CYP61*) in *Xanthophyllomyces dendrorhous*. *BMC Microbiol* **12**, 235.

Lukács G, Linka B, Nyilasi I (2006) *Phaffia rhodozyma* and *Xanthophyllomyces dendrorhous*: astaxanthin-producing yeasts of biotechnological importance. *Acta Alimentaria* **35**, 99-107.

Luo Y, Ullrich RC, Novotny CP (1994) Only one of the paired *Schizophyllum commune* A alpha mating-type, putative homeobox genes encodes a homeodomain essential for A alpha-regulated development. *Mol Gen Genet* **244**, 318-324.

Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. *Science* **252**, 1162-1164.

Maekawa H, Kaneko Y (2014) Inversion of the chromosomal region between two mating type loci switches the mating type in *Hansenula polymorpha*. *PLoS Genet* **10**, e1004796.

Maia TM, Lopes ST, Almeida JM*, et al.* (2015) Evolution of Mating Systems in Basidiomycetes and the Genetic Architecture Underlying Mating-Type Determination in the Yeast *Leucosporidium scottii*. *Genetics* **201**, 75-89.

Malik S-B, Pightling AW, Stefaniak LM, Schurko AM, Logsdon JM, Jr. (2008) An Expanded Inventory of Conserved Meiotic Genes Provides Evidence for Sex in *Trichomonas vaginalis*. *PLoS ONE* **3**, e2879.

Manchester SR, Xiang Q-Y, Xiang Q-P (2007) Curtisia (Cornales) from the Eocene of Europe and its phytogeographical significance. *Botanical Journal of the Linnean Society* **155**, 127-134.

Martinez C, Hermosilla G, Leon R, Pincheira G, Cifuentes V (1998) Genetic transformation of astaxanthin mutants of *Phaffia rhodozyma*. *Antonie Van Leeuwenhoek* **73**, 147-153.

Mata-Gomez LC, Montanez JC, Mendez-Zavala A, Aguilar CN (2014) Biotechnological production of carotenoids by yeasts: an overview. *Microb Cell Fact* **13**, 12.

May G, Shaw F, Badrane H, Vekemans X (1999) The signature of balancing selection: Fungal mating compatibility gene evolution. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 9172-9177.

Maynard-Smith J (1978) *The Evolution of Sex* Cambridge University Press.

Mayrhofer S, Weber JM, Poggeler S (2006) Pheromones and pheromone receptors are required for proper sexual development in the homothallic ascomycete *Sordaria macrospora*. *Genetics* **172**, 1521-1533.

Mccusker JH (2006) Saccharomyces cerevisiae: an Emerging and Model Pathogenic Fungus. In: *Molecular Principles of Fungal Pathogenesis*. American Society of Microbiology.

Medwid RD (1998) *Phaffia rhodozyma* is polyploid. *Journal of Industrial Microbiology and Biotechnology* **21**, 228-232.

Meinhardt LW, Rincones J, Bailey BA*, et al.* (2008) *Moniliophthora perniciosa*, the causal agent of witches' broom disease of cacao: what's new from this old foe? *Molecular Plant Pathology* **9**, 577-588.

Miao L, Chi S, Tang Y*, et al.* (2011) Astaxanthin biosynthesis is enhanced by high carotenogenic gene expression and decrease of fatty acids and ergosterol in a *Phaffia rhodozyma* mutant strain. *FEMS Yeast Res* **11**, 192-201.

Miller MW, Yoneyama M, Soneda M (1976) *Phaffia*, a New Yeast Genus in the Deuteromycotina (Blastomycetes). *International Journal of Systematic and Evolutionary Microbiology* **26**, 286-291.

Minh BQ, Nguyen MAT, von Haeseler A (2013) Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution* **30**, 1188-1195.

Mitchell A, Chang HY, Daugherty L*, et al.* (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* **43**, D213-221.

Moliné M, Libkind D, de Garcia V, Giraudo M (2014) Production of Pigments and Photo-Protective Compounds by Cold-Adapted Yeasts. In: *Cold-adapted Yeasts* (eds. Buzzini P, Margesin R), pp. 193-224. Springer Berlin Heidelberg.

Mondego JM, Carazzolle MF, Costa GG*, et al.* (2008) A genome survey of *Moniliophthora perniciosa* gives new insights into Witches' Broom Disease of cacao. *BMC Genomics* **9**, 548.

Morrow CA, Fraser JA (2009) Sexual reproduction and dimorphism in the pathogenic basidiomycetes. *FEMS Yeast Research* **9**, 161-177.

Mukai Y, Ohno-Yamashita Y, Oshima Y, Harashima S (1997) The role of cysteine residues in the homeodomain protein Mat alpha 2 in mating-type control of *Saccharomyces cerevisiae*. *Mol Gen Genet* **255**, 166-171.

Muller HJ (1964) The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **1**, 2-9.

Nagy Á, Garamszegi N, Vágvölgyi C, Ferenczy L (1994) Electrophoretic karyotypes of *Phaffia rhodozyma* strains. *FEMS Microbiology Letters* **123**, 315-318.

Nagy Á, Palágyi Z, Ferenczy L, Vágvölgyi C (1997) Radiation-induced chromosomal rearrangement as an aid to analysis of the genetic constitution of *Phaffia rhodozyma*. *FEMS Microbiology Letters* **152**, 249-254.

Nelson MD, Fitch DH (2011) Overlap extension PCR: an efficient method for transgene construction. *Methods Mol Biol* **772**, 459-470.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268-274.

Ni M, Feretzaki M, Sun S, Wang X, Heitman J (2011) Sex in Fungi. *Annual Review of Genetics* **45**, 405-430.

Ni M, Feretzaki M, Li W*, et al.* (2013) Unisexual and Heterosexual Meiotic Reproduction Generate Aneuploidy and Phenotypic Diversity *De Novo* in the Yeast *Cryptococcus neoformans*. *PLoS Biol* **11**, e1001653.

Nieuwenhuis BP, Immler S (2016) The evolution of mating-type switching for reproductive assurance. *BioEssays* **38**, 1141-1149.

Niklitschek M, Alcaíno J, Barahona S*, et al.* (2008) Genomic organization of the structural genes controlling the astaxanthin biosynthesis pathway of *Xanthophyllomyces dendrorhous*. *Biological Research* **41**, 93-108.

Niklitschek M, Baeza M, Fernández-Lobato M, Cifuentes V (2012) Generation of Astaxanthin Mutants in *Xanthophyllomyces dendrorhous* Using a Double Recombination Method Based on

Hygromycin Resistance. In: *Microbial Carotenoids From Fungi: Methods and Protocols* (ed. Barredo J-L), pp. 219-234. Humana Press, Totowa, NJ.
Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 332-342.
Nygren K, Strandberg R, Wallberg A*, et al.* (2011) A comprehensive phylogeny of *Neurospora* reveals a link between reproductive mode and molecular evolution in fungi. *Mol Phylogenet Evol* **59**, 649-663.
Nygren K, Strandberg R, Gioti A, Karlsson M, Johannesson H (2012) Deciphering the relationship between mating system and the molecular evolution of the pheromone and receptor genes in *Neurospora*. *Mol Biol Evol* **29**, 3827-3842.
O'Connell J, Schulz-Trieglaff O, Carlson E*, et al.* (2015) NxTrim: optimized trimming of Illumina mate pair reads. *Bioinformatics* **31**, 2035-2037.
Okonechnikov K, Golosova O, Fursov M, team tU (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**, 1166-1167.
Olesnicky N, Brown A, Dowell S, Casselton L (1999) A constitutively active G-protein-coupled receptor causes mating self-compatibility in the mushroom *Coprinus*. *EMBO J* **18**, 2756 - 2763.
Olesnicky N, Brown A, Honda Y*, et al.* (2000) Self-compatible B mutants in *Coprinus* with altered pheromone-receptor specificities. *Genetics* **156**, 1025 - 1033.
Paoletti M, Seymour FA, Alcocer MJ*, et al.* (2007) Mating type and the genetic basis of self-fertility in the model fungus *Aspergillus nidulans*. *Curr Biol* **17**, 1384-1389.
Parry DAD, Fraser RDB, Squire JM (2008) Fifty years of coiled-coils and α-helical bundles: A close relationship between sequence and structure. *Journal of Structural Biology* **163**, 258-269.
Peterson KR, Pfister DH (2010) Phylogeny of *Cyttaria* inferred from nuclear and mitochondrial sequence and morphological data. *Mycologia* **102**, 1398-1416.
Peterson KR, Pfister DH, Bell CD (2010) Cophylogeny and biogeography of the fungal parasite *Cyttaria* and its host *Nothofagus*, southern beech. *Mycologia* **102**, 1417-1425.
Phaff H, Miller M, Yoneyama M, Soneda M (1972) A comparative study of the yeast florae associated with trees on the Japanese islands and on the west coast of North America. In: *Proceedings of the 4th IFS: Fermentation Technology Today Meeting Society of Fermentation Technology, Osaka, Japan*.
Phaff HJM, M.W.; Yoneyama, M.; Soneda, M. (1972) A comparative study of the yeast florae associated with trees on the Japanese Island and on the West Coast of North America, 759-774.
Pontecorvo G, Roper JA, Hemmons LM, Macdonald KD, Bufton AW (1953) The genetics of *Aspergillus nidulans*. *Adv Genet* **5**, 141-238.
Pontes A, Rohl O, Carvalho C*, et al.* (2015) *Cystofilobasidium intermedium* sp. nov. and *Cystofilobasidium alribaticum* f.a. sp. nov. isolated from Mediterranean forest soils. *Int J Syst Evol Microbiol*.
Raju NB (1992) Functional heterothallism resulting from homokaryotic conidia and ascospores in *Neurospora tetrasperma*. *Mycological Research* **96**, 103-116.
Raju NB, Perkins DD (1994) Diverse programs of ascus development in pseudohomothallic species of *Neurospora*, *Gelasinospora*, and *Podospora*. *Developmental Genetics* **15**, 104-118.
Ramesh MA, Malik S-B, Logsdon Jr JM (2005) A Phylogenomic Inventory of Meiotic Genes: Evidence for Sex in *Giardia* and an Early Eukaryotic Origin of Meiosis. *Current Biology* **15**, 185-191.
Raper JR (1966) *Genetics of sexuality in higher fungi* Ronald Press Co.
Raudaskoski M, Kothe E (2010) Basidiomycete mating type genes and pheromone signaling. *Eukaryot Cell* **9**, 847 - 859.
Reeve ECR (2014) *Encyclopedia of Genetics* Taylor & Francis.
Regenfelder E, Spellig T, Hartmann A*, et al.* (1997) G proteins in *Ustilago maydis*: transmission of multiple signals? *EMBO J* **16**, 1934-1942.
Reynders MB, Rawlings DE, Harrison STL (1997) Demonstration of the Crabtree effect in *Phaffia rhodozyma* during continuous and fed-batch cultivation. *Biotechnology Letters* **19**, 549-552.
Riquelme M, Challen MP, Casselton LA, Brown AJ (2005) The origin of multiple B mating specificities in *Coprinus cinereus*. *Genetics* **170**, 1105-1119.
Roach KC, Feretzaki M, Sun S, Heitman J (2014) Unisexual reproduction. *Adv Genet* **85**, 255-305.
Roach KC, Heitman J (2014) Unisexual reproduction reverses Muller's ratchet. *Genetics* **198**, 1059-1069.
Robertson CI, McMahon Kende A, Toenjes K, Novotny CP, Ullrich RC (2002) Evidence for interaction of *Schizophyllum commune* Y mating-type proteins in vivo. *Genetics* **160**, 1461-1467.

Rodriguez-Carres M, Findley K, Sun S, Dietrich FS, Heitman J (2010) Morphological and genomic characterization of *Filobasidiella depauperata:* a homothallic sibling species of the pathogenic *Cryptococcus* species complex. *PLoS ONE* **5**, e9620.

Rodriguez-Saiz M, de la Fuente JL, Barredo JL (2010) *Xanthophyllomyces dendrorhous* for the industrial production of astaxanthin. *Appl Microbiol Biotechnol* **88**, 645-658.

Romeis T, Kämper J, Kahmann R (1997) Single-chain fusions of two unrelated homeodomain proteins trigger pathogenicity in *Ustilago maydis*. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 1230-1234.

Rubinstein L, Altamirano A, Santopietro LD, Baigorí M, de Figueroa LC (1996) Transformation of *Phaffia rhodozyma* by electroporation. *Biotechnology Techniques* **10**, 929-932.

Rusche LN, Kirchmaier AL, Rine J (2003) The establishment, inheritance, and function of silenced chromatin in *Saccharomyces cerevisiae*. *Annu Rev Biochem* **72**, 481-516.

Sambrook J, Russell D (2001) *Molecular Cloning: A Laboratory Manual*, 3 edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sampaio JP, Gadanho M, Bauer R (2001) Taxonomic studies on the genus *Cystofilobasidium*: description of *Cystofilobasidium ferigula* sp. nov. and clarification of the status of *Cystofilobasidium lari-marini*. *Int J Syst Evol Microbiol* **51**, 221-229.

Sandmann G (2015) Carotenoids of biotechnological importance. *Adv Biochem Eng Biotechnol* **148**, 449-467.

Scazzocchio C (2014) Fungal biology in the post-genomic era. *Fungal Biology and Biotechnology* **1**, 7.

Schirawski J, Heinze B, Wagenknecht M, Kahmann R (2005) Mating Type Loci of *Sporisorium reilianum*: Novel Pattern with Three a and Multiple b Specificities. *Eukaryotic Cell* **4**, 1317-1327.

Schlesinger R, Kahmann R, Kamper J (1997) The homeodomains of the heterodimeric *bE* and *bW* proteins of *Ustilago maydis* are both critical for function. *Mol Gen Genet* **254**, 514-519.

Schmidt I, Schewe H, Gassel S*, et al.* (2011) Biotechnological production of astaxanthin with *Phaffia rhodozyma/Xanthophyllomyces dendrorhous*. *Appl Microbiol Biotechnol* **89**, 555-571.

Schurko AM, Logsdon JM (2008) Using a meiosis detection toolkit to investigate ancient asexual "scandals" and the evolution of sex. *BioEssays* **30**, 579-589.

Scorzetti G, Fell JW, Fonseca A, Statzell-Tallman A (2002) Systematics of basidiomycetous yeasts: a comparison of large subunit *D1/D2* and internal transcribed spacer rDNA regions. *FEMS Yeast Res* **2**, 495-517.

Sharma R, Gassel S*, et al.* (2015) The genome of the basal agaricomycete *Xanthophyllomyces dendrorhous* provides insights into the organization of its acetyl-CoA derived pathways and the evolution of Agaricomycotina. *BMC Genomics* **16**, 233.

Slaninova I, Kucsera J, Svoboda A (1999) Topology of microtubules and actin in the life cycle of *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*). *Antonie Van Leeuwenhoek* **75**, 361-368.

Spit A, Hyland RH, Mellor EJC, Casselton LA (1998) A role for heterodimerization in nuclear localization of a homeodomain protein. *Proceedings of the National Academy of Sciences* **95**, 6228-6233.

Stamatakis A (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*.

Stanke M, Steinkamp R, Waack S, Morgenstern B (2004) AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* **32**, W309-312.

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637 - 644.

Stanton BC, Hull CM (2007) Mating-Type Locus Control of Cell Identity. In: *Sex in Fungi*. American Society of Microbiology.

Stanton BC, Giles SS, Kruzel EK*, et al.* (2009) Cognate Site Identifier analysis reveals novel binding properties of the Sex Inducer homeodomain proteins of *Cryptococcus neoformans*. *Mol Microbiol* **72**, 1334-1347.

Stukenbrock EH (2013) Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *New Phytologist* **199**, 895-907.

Sun S, Heitman J (2011) Is sex necessary? *BMC Biology* **9**, 56.

Sun S, Hsueh Y-P, Heitman J (2012) Gene Conversion Occurs within the Mating-Type Locus of *Cryptococcus neoformans* during Sexual Reproduction. *PLoS Genet* **8**, e1002810.

Sun S, Heitman J (2016) Running Hot and Cold: Recombination Around and Within Mating-Type Loci of Fungi and Other Eukaryotes. In: *Environmental and Microbial Relationships* (eds. Druzhinina IS, Kubicek CP), pp. 3-13. Springer International Publishing, Cham.

Tamura K, Peterson D, Peterson N*, et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**, 2731-2739.

Taylor JW, Hann-Soden C, Branco S, Sylvain I, Ellison CE (2015) Clonal reproduction in fungi. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 8901-8908.

Tsong AE, Tuch BB, Johnson AD (2007) Rewiring Transcriptional Circuitry: Mating-Type Regulation in *Saccharomyces cerevisiae* and *Candida albicans* as a Model for Evolution. In: *Sex in Fungi*. American Society of Microbiology.

Tsuchiya E, Fukui S (1978) Binding of rhodotorucine A, a lipopeptidyl mating hormone, to a cells of *Rhodosporidium toruloides* for induction of sexual differentiation. *Biochemical and Biophysical Research Communications* **85**, 473-479.

Tsuji M, Kudoh S, Hoshino T (2015) Draft Genome Sequence of Cryophilic Basidiomycetous Yeast *Mrakia blollopis* SK-4, Isolated from an Algal Mat of Naga-ike Lake in the Skarvsnes Ice-Free Area, East Antarctica. *Genome Announcements* **3**.

Tusnady GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849-850.

Ullrich RC (1973) Sexuality, Incompatibility, and Intersterility in the Biology of the *Sistotrema brinkmannii* Aggregate. *Mycologia* **65**, 1234-1249.

Ullrich RC, Raper JR (1975) Primary Homothallism-relation to Heterothallism in the Regulation of Sexual Morphogenesis in *Sistotrema*. *Genetics* **80**, 311-321.

van Diepen LT, Olson A, Ihrmark K, Stenlid J, James TY (2013) Extensive trans-specific polymorphism at the mating type locus of the root decay fungus *Heterobasidion*. *Mol Biol Evol* **30**, 2286-2301.

Vázquez M, Santos V, Parajó JC (1997) Effect of the carbon source on the carotenoid profiles of *Phaffia rhodozyma* strains. *Journal of Industrial Microbiology and Biotechnology* **19**, 263-268.

Verdoes JC, Krubasik KP, Sandmann G, van Ooyen AJ (1999) Isolation and functional characterisation of a novel type of carotenoid biosynthetic gene from *Xanthophyllomyces dendrorhous*. *Mol Gen Genet* **262**, 453-461.

Villeneuve AM, Hillers KJ (2001) Whence Meiosis? *Cell* **106**, 647-650.

Visser H, Sandmann G, Verdoes JC (2005) Xanthophylls in Fungi - Metabolic Engineering of the Astaxanthin Biosynthetic Pathway in *Xanthophyllomyces dendrorhous*. In: *Microbial Processes and Products* (ed. Barredo JL), pp. 257-272. Humana Press Inc. , New Jersey.

Wainright PO, Hinkle G, Sogin ML, Stickel SK (1993) Monophyletic origins of the metazoa: an evolutionary link with fungi. *Science* **260**, 340-342.

Wang CL, Shim WB, Shaw BD (2010) *Aspergillus nidulans* striatin (StrA) mediates sexual development and localizes to the endoplasmic reticulum. *Fungal Genet Biol* **47**, 789-799.

Weber RW, Davoli P, Anke H (2006) A microbial consortium involving the astaxanthin producer *Xanthophyllomyces dendrorhous* on freshly cut birch stumps in Germany. *Mycologist* **20**.

Wery J, Gutker D, Renniers AC, Verdoes JC, van Ooyen AJ (1997) High copy number integration into the ribosomal DNA of the yeast *Phaffia rhodozyma*. *Gene* **184**, 89-97.

Wery J, Verdoes JC, Ooyen AJJv (1998) Efficient transformation of the astaxanthin-producing yeast *Phaffia rhodozyma*. *Biotechnology Techniques* **12**, 399-405.

Whitehouse HLK (1949) Heterothallism and sex in the fungi. *Biological Reviews of the Cambridge Philosophical Society* **24**, 411-447.

Wik L, Karlsson M, Johannesson H (2008) The evolutionary trajectory of the mating-type (mat) genes in Neurosporarelates to reproductive behavior of taxa. *BMC Evolutionary Biology* **8**, 109.

Wilken PM, Steenkamp ET, Wingfield MJ, de Beer ZW, Wingfield BD (2014) DNA Loss at the *Ceratocystis fimbriata* Mating Locus Results in Self-Sterility. *PLoS ONE* **9**, e92180.

Williams GC (1975) *Sex and Evolution* Princeton University Press.

Wilson AM, Godlonton T, van der Nest MA*, et al.* (2015a) Unisexual reproduction in *Huntiella moniliformis*. *Fungal Genet Biol* **80**, 1-9.

Wilson AM, Wilken PM, van der Nest MA*, et al.* (2015b) Homothallism: an umbrella term for describing diverse sexual behaviours. *IMA Fungus* **6**, 207-214.

Wozniak A, Lozano C, Barahona S*, et al.* (2011) Differential carotenoid production and gene expression in *Xanthophyllomyces dendrorhous* grown in a nonfermentable carbon source. *FEMS Yeast Res* **11**, 252-262.

Wu G, Zhao H, Li C*, et al.* (2015) Genus-Wide Comparative Genomics of Malassezia Delineates Its Phylogeny, Physiology, and Niche Adaptation on Human Skin. *PLOS Genetics* **11**, e1005614.

Xin M-x, Zhou P-j (2007) *Mrakia psychrophila* sp. nov., a new species isolated from Antarctic soil. *Journal of Zhejiang University. Science. B* **8**, 260-265.

Xue C, Hsueh Y, Heitman J (2008) Magnificent seven: roles of G protein-coupled receptors in extracellular sensing in fungi. *FEMS Microbiol Rev* **32**, 1010 - 1032.

Yee AR, Kronstad JW (1993) Construction of chimeric alleles with altered specificity at the b incompatibility locus of *Ustilago maydis*. *Proc Natl Acad Sci U S A* **90**, 664-668.

Yee AR, Kronstad JW (1998) Dual sets of chimeric alleles identify specificity sequences for the *bE* and *bW* mating and pathogenicity genes of *Ustilago maydis*. *Mol Cell Biol* **18**, 221-232.

Yue C, Osier M, Novotny CP, Ullrich RC (1997) The specificity determinant of the Y mating-type proteins of *Schizophyllum commune* is also essential for Y-Z protein binding. *Genetics* **145**, 253-260.

Yun S-H, Berbee ML, Yoder OC, Turgeon BG (1999) Evolution of the fungal self-fertile reproductive life style from self-sterile ancestors. *Proceedings of the National Academy of Sciences* **96**, 5592-5597.

APPENDIX I

Additional information pertaining to Chapter 2

Table I.1. Complete list of *MAT* genes and *MAT* related genes searched in the draft genome of CBS 7918^T.

Available at: https://figshare.com/s/6d9f91dc5d38d25166ae

Table I.2. Complete list of meiotic genes searched in the draft genome of CBS 7918^T.

Available at: https://figshare.com/s/6d9f91dc5d38d25166ae

Table I.3. Amino acid identity (%) between different alleles of pheromone receptors from selected basidiomycetes. Proteins were aligned using ClustalW as implemented in Bioedit (Hall 1999).

Species	Mating	Receptor gene	Accession number / Genome Locus (coordinates)	% aa identity
Phaffia rhodozyma (CRUB 1149)	Homothallic	*STE3-1*	PRJNA307837/ NODE_186 (11928..13388)	49
		STE3-2	PRJNA307837/ NODE_198(2416..3807)	
Phaffia rhodozyma (CBS 6938)	Homothallic	*STE3-1*	CED85384.1	50
		STE3-2	CED85379.1	
Phaffia rhodozyma (CBS 7918)	Homothallic	*STE3-1*	PRJNA306035/ LSVH01000226.1 (11858..13318)	50
		STE3-2	PRJNA306035/ LSVH01000253.1 (7004..8395)	
Microbotryum violaceum	Heterothallic bipolar	*A1*	ABU62847.1	20
		A2	ABU62846.1	
Ustilago hordei	Heterothallic bipolar	*Pra1*	CAJ41875.1	23
		Pra2	AAD56044.1	
Cryptococcus flavescens	Heterothallic tetrapolar	*A2*	CDR19322.1	29
		A1	CDR19327.1	
Sporidiolobos salmonicolor	Heterothallic tetrapolar	*A2*	ADM24775.1	26
		A1	ADM24772.1	
Cryptococcus neoformans	Heterothallic bipolar	*STE3 a*	AAN75624.1	31
		STE3 α	AAN75724.1	
Cryptococcus gattii	Heterothallic bipolar	*STE3 a*	AEG78597.1	31
		STE3 α	AEG78622.1	
Sporisorium reilianum	Heterothallic tetrapolar	*Pra1*	CAI59749.1	Pra1 vs. Pra2: 23 Pra1 vs. Pra3: 28 Pra2 vs. Pra3: 24
		Pra2	CBQ71125.1	
		Pra3	ABW21687.1	

Table I.4. Accession numbers and information for all sequences obtained and used in chapter 2.

Strain	Gene Sequence	Accession number / Scaffold (coordinates)
Phaffia rhodozyma KBP 2604	*STE3-1 (partial)*	KU315740
Phaffia rhodozyma ATCC 24261	*STE3-1 (partial)*	KU315745
Phaffia rhodozyma NRRL Y-17434	*STE3-1 (partial)*	KU315746
Phaffia rhodozyma GY13L04	*STE3-1 (partial)*	KU315744
Phaffia rhodozyma CRUB 1151	*STE3-1 (partial)*	KU315742
Phaffia rhodozyma CRUB 0853	*STE3-1 (partial)*	KU315743
Phaffia rhodozyma CRUB 1490	*STE3-1 (partial)*	KU315741
Phaffia rhodozyma ATCC 24229	*STE3-1 (partial)*	KU315749
Phaffia rhodozyma ATCC 24201	*STE3-1 (partial)*	KU315750
Phaffia rhodozyma ZP 922	*STE3-1 (partial)*	KU315747
Phaffia rhodozyma ZP 869	*STE3-1 (partial)*	KU315748
Phaffia rhodozyma CBS 6938	*STE3-1*	CED85384.1
Phaffia rhodozyma CBS 7918	*STE3-1*	PRJNA306035/ LSVH01000226.1 (11858..13318)
Phaffia rhodozyma CRUB 1149	*STE3-1*	PRJNA307837/ NODE_186 (11928..13388)
Phaffia rhodozyma KBP 2604	*STE3-2 (partial)*	KU315756
Phaffia rhodozyma ATCC 24261	*STE3-2 (partial)*	KU315759
Phaffia rhodozyma NRRL Y-17434	*STE3-2 (partial)*	KU315754
Phaffia rhodozyma GY13L04	*STE3-2 (partial)*	KU315753
Phaffia rhodozyma CRUB 1151	*STE3-2 (partial)*	KU315755

Strain	Gene Sequence	Accession number / Scaffold (coordinates)
Phaffia rhodozyma CRUB 0853	*STE3-2 (partial)*	KU315757
Phaffia rhodozyma CRUB 1490	*STE3-2 (partial)*	KU315760
Phaffia rhodozyma ATCC 24229	*STE3-2 (partial)*	KU315758
Phaffia rhodozyma ATCC 24201	*STE3-2 (partial)*	KU315761
Phaffia rhodozyma ZP 922	*STE3-2 (partial)*	KU315752
Phaffia rhodozyma ZP 869	*STE3-2 (partial)*	KU315751
Phaffia rhodozyma CBS 6938	*STE3-2*	CED85379.1
Phaffia rhodozyma CBS 7918	*STE3-2*	PRJNA306035/ LSVH01000253.1 (7004..8395)
Phaffia rhodozyma CRUB 1149	*STE3-2*	PRJNA307837/ NODE_198(2416..3807)
Phaffia rhodozyma KBP 2604	*HD1 (partial)*	KU315766
Phaffia rhodozyma ATCC 24261	*HD1 (partial)*	KU315764
Phaffia rhodozyma NRRL Y-17434	*HD1 (partial)*	KU315767
Phaffia rhodozyma ZP 922	*HD1*	KU315769
Phaffia rhodozyma ZP 869	*HD1*	KU315768
Phaffia rhodozyma GY13L04	*HD1*	KU315765
Phaffia rhodozyma CRUB 1151	*HD1*	KU315770
Phaffia rhodozyma ATCC 24229	*HD1*	KU315763
Phaffia rhodozyma ATCC 24201	*HD1*	KU315762
Phaffia rhodozyma CBS 6938	*HD1*	CDZ96688.1
Phaffia rhodozyma CBS 7918	*HD1*	PRJNA306035/ LSVH01000060.1 (57836..59696)
Phaffia rhodozyma CRUB 1149	*HD1*	PRJNA307837/
Phaffia rhodozyma ZP 922	*HD2*	KU315773
Phaffia rhodozyma ZP 869	*HD2*	KU315772
Phaffia rhodozyma GY13L04	*HD2*	KU315775
Phaffia rhodozyma CRUB 1490	*HD2*	KU315780
Phaffia rhodozyma CRUB 0853	*HD2*	KU315779
Phaffia rhodozyma ATCC 24229	*HD2*	KU315778
Phaffia rhodozyma ATCC 24201	*HD2*	KU315777
Phaffia rhodozyma KBP 2604	*HD2*	KU315776
Phaffia rhodozyma ATCC 24261	*HD2*	KU315771
Phaffia rhodozyma NRRL Y-17434	*HD2*	KU315774
Phaffia rhodozyma CBS 6938	*HD2*	CDZ96689.1
Phaffia rhodozyma CBS 7918	*HD2*	PRJNA306035/ LSVH01000060.1 (55622..57058)
Phaffia rhodozyma CRUB 1149	*HD2*	PRJNA307837/
Cryptococcus deneoformans JEC21	*STE3*	XP_570116.1
Cryptococcus deneoformans JEC20	*STE3*	AAN75624.1
Cryptococcus neoformans H99	*STE3*	XP_012049557.1
Cryptococcus neoformans 125.91	*STE3*	AAN75156.1
Cryptococcus gattii WM276	*STE3*	XP_003196044.1
Cryptococcus gattii E566	*STE3*	AAV28758.1
Cryptococcus flavescens CF05-CBS8359	*STE3*	CDR19326.1
Cryptococcus flavescens CF01-CBS4918	*STE3*	CDR19282.1
Kwoniella magrovensis CBS8507	*STE3*	CCM73226.1
Kwoniella magrovensis CBS10435	*STE3*	ASQD01000019.1
Kwoniella heveanensis CBS569	*STE3*	ACZ81463.1
Kwoniella heveanensis BCC8398	*STE3*	ASQB01000005
Dioszegia cryoxerica	*STE3*	ANT03-071JGI-262391
Dioszegia cryoxerica	*STE3*	ANT03-071JGI-351885
Tremella fuciformis tr26	*STE3*	LBGW01000351
Tremella mesentarica ATCC24925	*STE3*	ADO17672.1
Sporidiobolus salmonicolor CBS490	*STE3*	ADM24775.1
Sporidiobolus salmonicolor CBS483	*STE3*	ADM24772.1
Leucosporidium scottii CBS5931	*STE3*	CRX79175.1
Leucosporidium scottii CBS5930	*STE3*	CRX79175.1
Saccharomyces cerevisiae S288c	*STE3*	NP_012743.1
Phaffia rhodozyma CBS 7918	*MFA1*	PRJNA306035/ LSVH01000226 (14268..14399)
Phaffia rhodozyma CBS 6938	*MFA1*	PRJEB6925/ LN483332.1 (1417935..1418060)
Phaffia rhodozyma CRUB 1149	*MFA1*	PRJNA307837/ NODE_186 (14322..14453)
Phaffia rhodozyma CBS 7918	*MFA2*	PRJNA306035/ LSVH01000253 (9521..9646)
Phaffia rhodozyma CBS 6938	*MFA2*	PRJEB6925/ LN483332.1 (1423341..1423472)
Phaffia rhodozyma CRUB 1149	*MFA2*	PRJNA307837/ NODE_198(1166..1291)

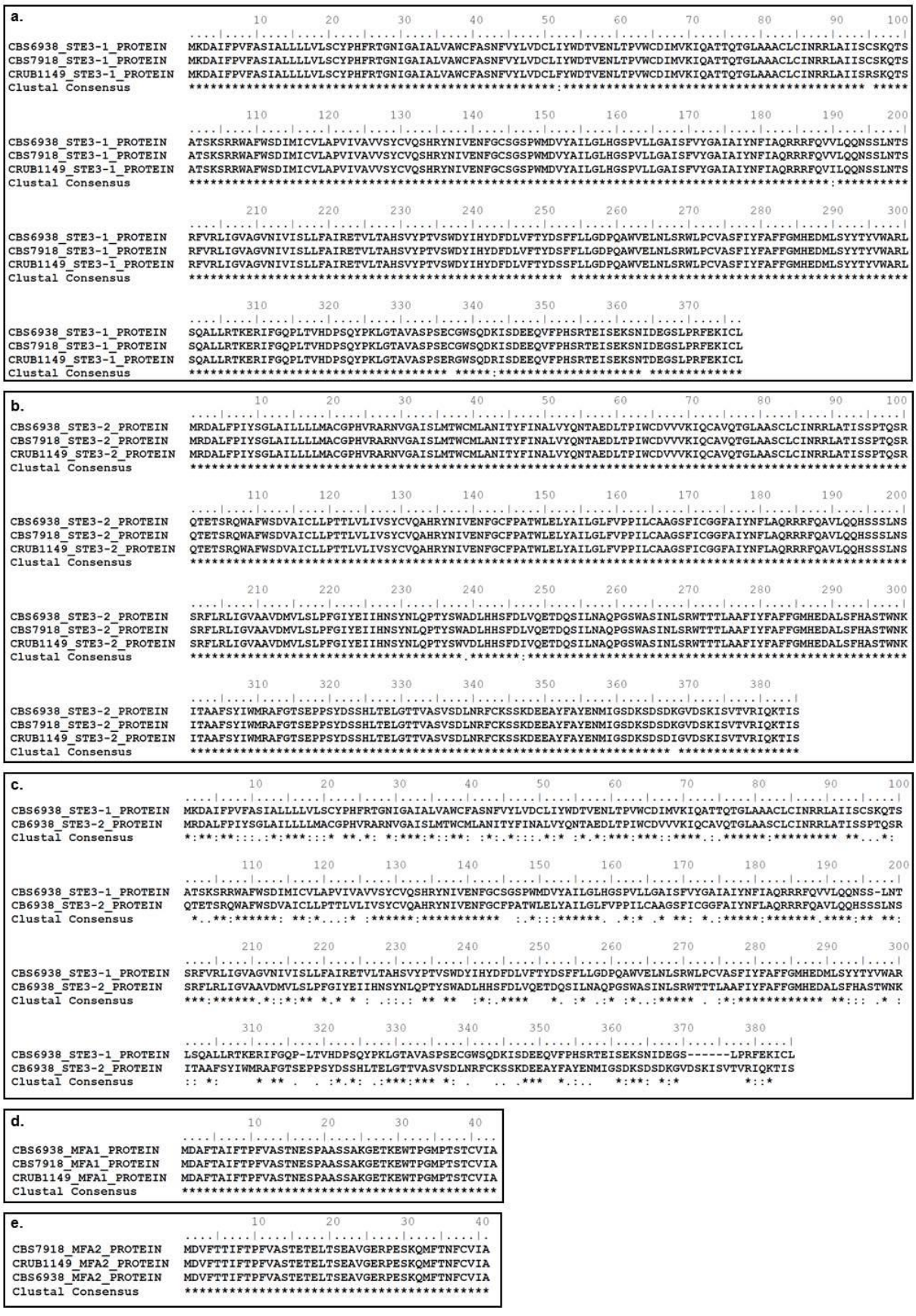


Figure I.1. *MAT* protein alignments. (a-e) Alignments of Ste3-1, Ste3-2, Ste3-1 vs Ste3-2, Mfa1and Mfa2 proteins from strains CBS 7918^{T}, CBS 6938 and CRUB 1149.

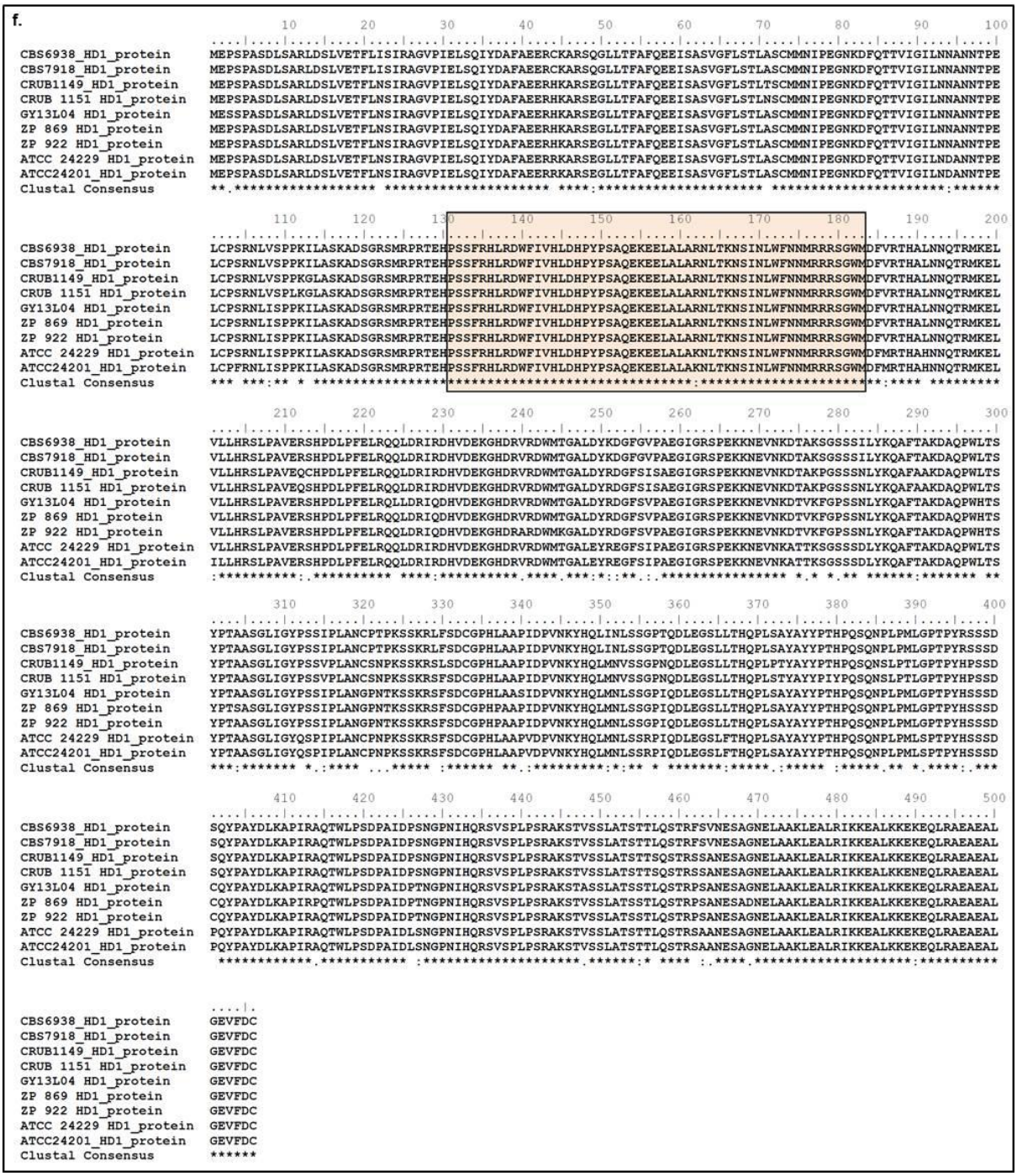


Figure I.1. *MAT* protein alignments (continued). (f) Alignment of Hd1 proteins from multiple strains of distinct *P. rhodozyma* populations. Box in light orange indicates homeodomain region.

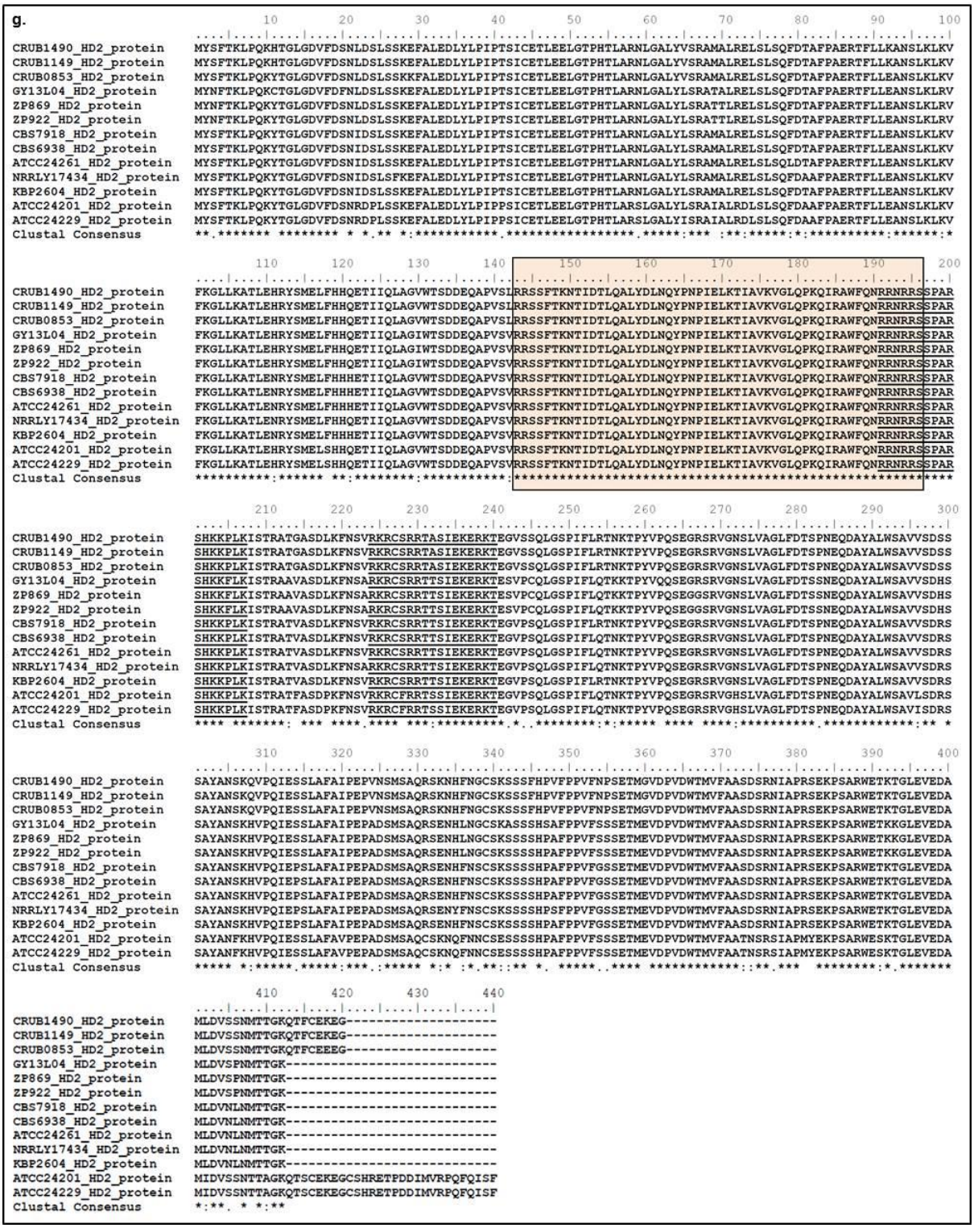


Figure I.1. *MAT* protein alignments (continued). (g) Alignment of Hd2 proteins from multiple strains of distinct *P. rhodozyma* populations. Box in light orange indicates homeodomain region and underlined regions indicate NLS. Alignments were performed using ClustalW as implemented in Bioedit (Hall 1999).

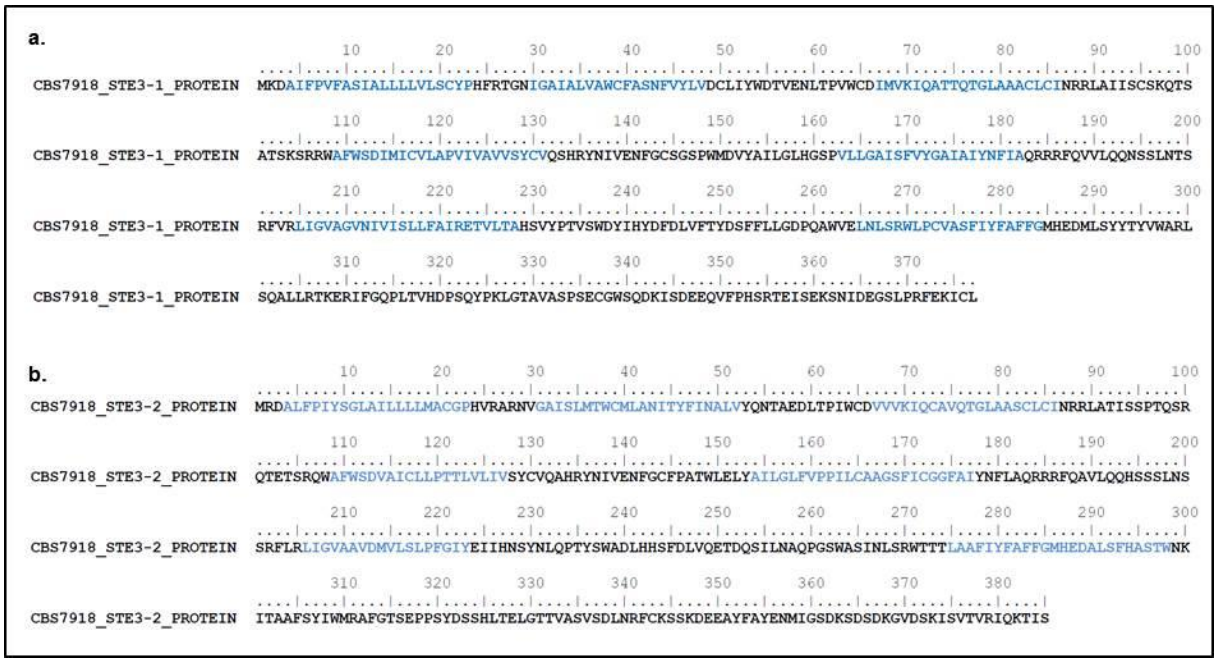


Figure I.2. Predicted pheromone receptors proteins Ste3-1 and Ste3-2 of strain CBS 7918^{T}. Pheromone receptor predicted proteins with the transmembrane domains highlighted in blue.

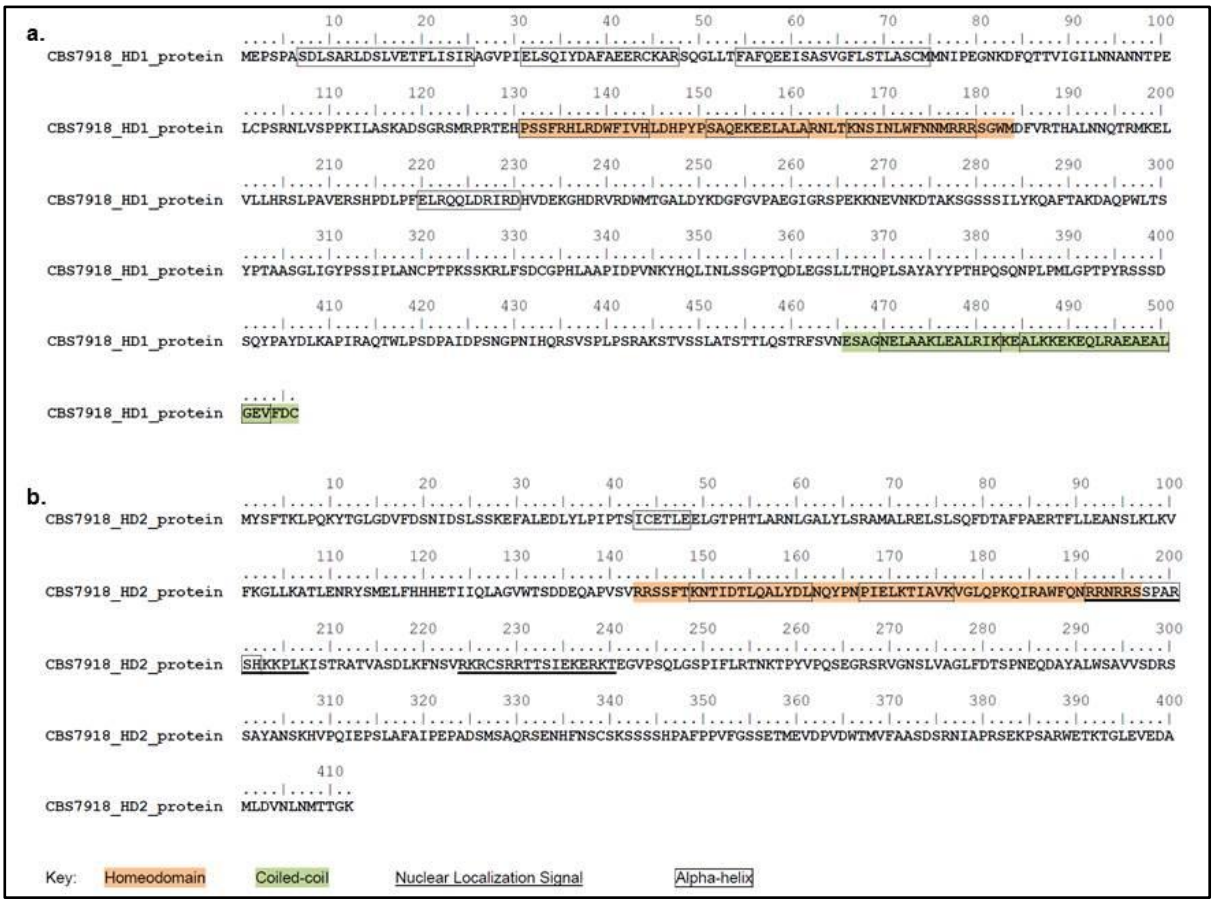


Figure I.3. Predicted Hd1 and Hd2 proteins of strain CBS 7918^{T}. Notable motifs of each of the proteins are highlighted as depicted in the figures key.

APPENDIX II

Additional information pertaining to Chapter 3

Table II.1. List of strains used in bacterial and yeast-two hybrid assays.

Strains	Relevant features
DH5α	*Escherichia coli laboratory strain (Gibco-BRL, Carlsbad, CA, USA)*
BTH101	*Escherichia coli reporter strain for BACTH (Bacterial Adenylate Cyclase Two-Hybrid System Kit) from EUROMEDEX (Cat. No. EUK001)*
Y2HGold	*Saccharomyces cerevisiae host strain for Matchmaker Gold Yeast Two-Hybrid System from Clontech (Cat. No. 630489); MATa, trp1-901, leu2-3, 112,ura3-52, his3-200, gal4Δ, gal80Δ,LYS2 : : GAL1UAS–Gal1TATA–His3, GAL2UAS–Gal2TATA–Ade2, URA3 : : MEL1UAS–Mel1TATA, AUR1-C, MEL1*
Y187	*Saccharomyces cerevisiae host strain for Matchmaker Gold Yeast Two-Hybrid System from Clontech (Cat. No. 630489); MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, gal80Δ, met–,URA3 : : GAL1UAS–Gal1TATA–LacZ,MEL1*

Table II.2. Primers and plasmids used for the construction of all gene deletion fragments except for gene *STE3-2*.

		Primers use to construct gene deletion fragments						Primers used for mutant confirmation		
Gene or region to knockout	**Backbone plasmid**	**Primers upstream from gene (5'-3')**	**Amplified fragment (bp)**	**Primers downstream from gene (5'-3')**	**Amplified fragment (bp)**	**Nested primers (5'-3')**	**Amplified fragment (bp)**	**Primers (5'-3')**	**Amplified fragment (bp)**	**Description**
Ste3-1	pPR2TN	MP066 (Kpn I) - GCGTCGGTACCGATGGAATGAGAGAG MP067 (EcoR I) - CTGGCGAATTCTGCTGTCTTTTAAGTTC	848	MP064 (Not I) - AGTGAGCGGCCGCATGATGCGCGCGACG MP065 (Hind III) - GTTGCAAGCTTCAATCTTAGCTGCGCTGG	1128	MP068 - GCGCTGGACAACACGGAGC MP069 - CAGTTTTTCATTCAATCCGAACGG	3869	MP029 - CGGTGGGCGTTCTGGTCGGAC MP030 - ACTCTGATGGCGAAGCAACGGC	836	Amplification of the *Ste3-1* gene (partial)
								MP067 - CTGGCGAATTCTGCTGTCTTTTAAGTTC MP092 - TTCCACCATGATATTCGGCAAGCAGG	2220	Upstream flanking region of *Ste3-1* gene and partial resistance cassette
								MP091 - AAGATGGATTGCACGCAGGTTCTCC MP065 - GTTGCAAGCTTCAATCTTAGCTGCGCTGG	2500	Downstream flanking region of *Ste3-1* gene and partial resistance cassette
Hd1	pBS-HYG	MP072 (Cla I) - AGATGATCGATTTTTCTATCTGGTGC MP073 (Kpn I) - ATGATGGTACCCGTAAGCTGTAGTTTC	775	MP070 (Sac I) - CTGTCGAGCTCAAACCGCTCTCAACC MP071 (Not I) - GGTTTGCGGCCGCTAGATTTCCCTTTTG	913	MP074 - AACGAAAGATGAAGACACCAG MP075 - GTTTCACCAGAACATATCAGG	3508	MP083 - AAAACCTCTCCAAGCGCC MP084 - TCTACATCGGTCTCTTCC MP154- ATGAGATCGTTCAACAGC (paired with MP086)	1002 2580	Amplification of the *Hd1* gene (partial)
								MP060 - GTGGAAACCGACGCCCCAGC MP073 - ATGATGGTACCCGTAAGCTGTAGTTTC	1500	Upstream flanking region of *Hd1* gene and partial resistance cassette
								MP070 - CTGTCGAGCTCAAACCGCTCTCAACC MP063 - TTTGCCCTCGGACGAGTGCTGG	2932	Downstream flanking region of *Hd1* gene and partial resistance cassette
Hd2	pPR2TN	MP076 (EcoR I) - GATGGGAATTCTTTCTATCTGGTGCAC MP073 (Kpn I) - ATGATGGTACCCGTAAGCTGTAGTTTC	774	MP077 (Not I) - GACGAGCGGCCGCTAAACCTCTTGTGAG MP078 (Hind III) - ACTCCAAGCTTTCTCTATGGTTGAGAG	629	MP079 - TTTCTATCTGGTGCACAACTTC MP080 - TATGGTTGAGAGTAATGAGC	3305	MP085 - TGTACAGTTTCACGAAGC MP086 - GTCTTCCGTTCTTTCTCG MP155 - ACATTTCATTAAGGCTGG (paired with MP085)	916 1595	Amplification of the *Hd2* gene (partial)
								MP076 - GATGGGAATTCTTTCTATCTGGTGCAC MP048 - GTACAGTCGCACAACCGACACGC	2254	Upstream flanking region of *Hd2* gene and partial resistance cassette
								MP037 - GCGTGTCGGTTGTGCGACTGTAC MP078 - ACTCCAAGCTTTCTCTATGGTTGAGAG	850	Downstream flanking region of *Hd2* gene and partial resistance cassette
	pJET1.2 + Zeo	MP156 (Bsm I) - TTATTGAATGCTTTCTATCTGGTGCAC MP157 (Cla I) - AAAAAATCGATCGTAAGCTGTAGTTTC	776	MP158 (Pst I) - TTATACTGCAGTCTCTATGGTTGAGAGTAATGAGC MP077 (Not I) - GACGAGCGGCCGCTAAACCTCTTGTGAG	630	MP079 - TTTCTATCTGGTGCACAACTTC MP080 - TATGGTTGAGAGTAATGAGC	3305	MP085 - TGTACAGTTTCACGAAGC MP086 - GTCTTCCGTTCTTTCTCG	916	Amplification of the *Hd2* gene (partial)
								MP049 - AGATAGAAACCCAACACTCGC MP114 - ATAAAAGATCTATGGCCAAGTTGACC	1745	Upstream flanking region of *Hd2* gene and partial resistance cassette
								MP115 - ATAATTCTAGATCAGTCCTGCTCCTC MP159 - TTATCTCGGTCAACTGCCGG	1446	Downstream flanking region of *Hd2* gene and partial resistance cassette
Ste3-1/mfa1	pPR2TN	MP087 (Kpn I) - TCATTGGTACCGTGGTAAGTCTGCCTG MP088 (Eco RI) - TCACCGAATTCCACTTGCGCGCTTGC	1058	MP064 (Not I) - AGTGAGCGGCCGCATGATGCGCGCGACG MP065 (Hind III) - GTTGCAAGCTTCAATCTTAGCTGCGCTGG	1126	MP068 - GCGCTGGACAACACGGAGC MP089 - GCTTGCATTCATACATACCTCAGC	4591	MP029 - CGGTGGGCGTTCTGGTCGGAC MP030 - ACTCTGATGGCGAAGCAACGGC	836	Amplification of the *Ste3-1* gene (partial)
								MP100 - TCCATCCTCAACTGATTGC MP101 - AGTGTAAAGAAGTCTCGG	357	Amplification of the *Mfa1* (complete) region
								MP091 - AAGATGGATTGCACGCAGGTTCTCC MP065 - GTTGCAAGCTTCAATCTTAGCTGCGCTGG	2500	Downstream flanking region of *Ste3-1* gene and partial resistance cassette
								MP092 - TTCCACCATGATATTCGGCAAGCAGG MP088 - TCACCGAATTCCACTTGCGCGCTTGC	2430	Downstream flanking region of *Mfa1* gene and partial resistance cassette
Ste3-2/mfa2	pPR2TN	MP093 (EcoR I) - TTTCAGAATTCCACCACATTCTCCCG MP094 (Kpn I) - AGAATTGGTACCCTCGACGGCTTTAGC	824	MP095 - AATATAAGCGGCCGCTACCCTGATACCTTG MP096 -TTCTGAAGCTTAAGTTCTGATTATCTAGAACCATCC	848	MP097 - CTCTACATAGAACTAGAATGGTCC MP098 - GAGTCCTTGATGTTGACGTAACC	3224	MP035 - TTATGCATCAACCGGCGTCTGGC MP036 - GGACACAGAGGCAACRGTAGTTCC	929	Amplification of the *Ste3-2* gene (partial)
								MP102 - TATATCATCTCTCGACCC MP103 - TTCATCTTGTCAGACAGC	611	Amplification of the *Mfa2* (complete) region
								MP091 - AAGATGGATTGCACGCAGGTTCTCC MP099 - ACATGCTGTAGGCTTATCGATAGCG	2400	Downstream flanking region of *Ste3-2* gene and partial resistance cassette
								MP093 - TTTCAGAATTCCACCACATTCTCCCG MP092 - TTCCACCATGATATTCGGCAAGCAGG	2000	Downstream flanking region of *Mfa2* gene and partial resistance cassette

Table II.2. (continued).

		Primers use to construct gene deletion fragments						Primers used for mutant confirmation		
Gene or region to knockout	**Backbone plasmid**	**Primers upstream from gene (5'-3')**	**Amplified fragment (bp)**	**Primers downstream from gene (5'-3')**	**Amplified fragment (bp)**	**Nested primers (5'-3')**	**Amplified fragment (bp)**	**Primers (5'-3')**	**Amplified fragment (bp)**	**Description**
Spo11	pBS-HYG	MP126 (Not I) - TATTTGCGGCCGCGTTGAATGTTTCAGG MP127 (Sac I) - TATTTGAGCTCGCTTTGTCAACCTATGCG	732	MP124 (Kpn I) - TATTTGGTACCTCAAGGAAACATGCG MP125 (Cla I) - TTAATATCGATGTCATTCCCCGAAGC	565	MP128 - AAGGAAACATGCGGGACG MP129 - GCTTTGTCAACCTATGCG	4000	MP130 - TTACTTCCATTCCTTCGGC MP131 - TGACTTGTCCAGTATCTCC	2750	Amplification of the *Spo11* gene (partial)
								MP063 - TTTGCCCTCGGACGAGTGCTGG MP133 - TTGCCGGAGGTGGTTGGATGG	2293	Upstream flanking region of *Spo11* gene and partial resistance cassette
								MP132 - TCGTTCCTCAGCCCTAAGACATGC MP060 - GTGGAAACCGACGCCCCAGC	1344	Downstream flanking region of *Spo11* gene and partial resistance cassette
Mfa1	pJET1.2 + Zeo	MP171 (Pst I) - TTTTTCTGCAGACTAGGTACATCACTCG MP 172 (Not I) - TTTATGCGGCCGCTCTTTACACTTTTGC	1165	MP173 (Cla I) - ATTTTATCGATAGGCAATCAGTTGAGG MP174 (Pme I) - TATTAGTTTAAACATAAGGCAGTCGACG	1024	MP175 - ACATCACTCGGTCAGATCC MP176 - ATAAGGCAGTCGACGAGG	3363	MP100 - TCCATCCTCAACTGATTGC MP101 - AGTGTAAAGAAGTCTCGG	357	Amplification of the *Mfa1* (complete) region
								MP171 - TTTTTCTGCAGACTAGGTACATCACTCG MP115 - ATAATTCTAGATCAGTCCTGCTCCTC	1937	Downstream flanking region of *Mfa1* gene and partial resistance cassette
								MP114 - ATAAAAGATCTATGGCCAAGTTGACC MP174 - TATTAGTTTAAACATAAGGCAGTCGACG	1794	Upstream flanking region of *Mfa1* gene and partial resistance cassette
Mfa2	pJET1.2 + Zeo	MP162 (Pst I) - TTTTTCTGCAGGATCAGAGGAGTACGATCG MP163 (Not I) - TTTTTGCGGCCGCATGGGATCTTCAGGTAG	998	MP166 (Cla I) - TTTTTATCGATGTATATACGAGATCACC MP167 (Bsm I) - TTTTTGAATGCAGAACGACTGGAAAGAG	1014	MP168 - ATCAGAGGAGTACGATCG MP170 - AGAACGACTGGAAAGAGC	3193	MP102 - TATATCATCTCTCGACCC MP103 - TTCATCTTGTCAGACAGC	611	Amplification of the *Mfa2* (complete) region
								MP162 - TTTTTCTGCAGGATCAGAGGAGTACGATCG MP115 - ATAATTCTAGATCAGTCCTGCTCCTC	1770	Upstream flanking region of *Mfa2* gene and partial resistance cassette
								MP114 - ATAAAAGATCTATGGCCAAGTTGACC MP167 - TTTTTGAATGCAGAACGACTGGAAAGAG	1784	Downstream flanking region of *Mfa2* gene and partial resistance cassette

Table II.3. Primers and plasmid used for the construction of the STE3-2 deletion fragment by overlap extension PCR.

Gene or region to knockout	**Plasmid**	**Primers upstream from gene (5'-3')**	**Amplified fragment (bp)**	**Primers for HYG resistance cassette (5'-3')**	**Amplified fragment (bp)**	**Primers downstream from gene (5'-3')**	**Amplified fragment (bp)**	**Nested primers (5'-3')**	**Amplified fragment (bp)**	**Primers (5'-3') used for mutant confirmation and amplification description**	**Amplified fragment (bp)**
Ste3-2	pJet1.2 resistance cassette storage	MP039 - TGTCAGGGCAAAAAGAGAAAAACAAGGGG MP040 - AAGAGCTTGTGTCGGATGAACTGTCGGCTGA TGAGCCGATGATCAGAGGAGTACGATCGAAAGAGAAGAT (Amplified from CBS 6938 gDNA)	1165	MP041 - TCGATATATAATCTTCTCTTTCGATCGTACTCCTC TGATCATCGGCTCATCAGCCGACAGTTCATCCGAC MP042 - TATACTACAACATTACCGAAGGTATCAAGGTATC AGGGTAATCATGAGAGATGACGGAGATGATGGTGATG (Amplified from pBS-HYG plasmid)	1899	MP043 - TCTGTTGACCATCACCATCATCTCCGTCATCTCT CATGATTACCCTGATACCTTGATACCTTCGGTAATG MP044 - AACGAGCAAGGAGACAAACAACAGGTCGAG (Amplified from CBS 6938 gDNA)	609	MP045 - GAGAAAAACAAGGGGAGATTTGGG MP046 - ACAGGTCGAGGGCAGCAAGGG	3480	MP035 - TTATGCATCAACCGGCGTCTGGC MP036 - GGACACAGAGGCAACRGTAGTTCC (Amplification of the *Ste3-2* gene (partial))	929
										MP060 - GTGGAAACCGACGCCCCAGC MP061 - GAGGGCTGTTGACGATAGGG (Downstream flanking region of *Ste3-2* gene and partial resistance cassette)	1156
										MP063 - TTTGCCCTCGGACGAGTGCTGG MP039 - TGTCAGGGCAAAAAGAGAAAAACAAGGGG (Upstream flanking region of *Ste3-2* gene and partial resistance cassette)	2542

Table II.4. Primers and plasmid used for the construction of the Zeocin resistance cassette.

Resistance cassette	Backbone Plasmid	Primers for promoter (*Ptef*) (5'-3')	Amplified fragment (bp)	Primers for *Sh ble* gene (5'-3')	Amplified fragment (bp)	Primers for terminator (*Tgpd*) (5'-3')	Amplified fragment (bp)	Nested primers (5'-3')	Amplified fragment (bp)	Other information
Zeocin	pJET1.2	MP112 (Not I) - ATAATGCGGCCGCATCGGCTCATCAGC MP113 (Bgl II) - ATGCGAGATCTGGTGAAGCTGTTCGAG (Amplified from CBS 6938 gDNA)	398	MP114 (Bgl II) - ATAAAAGATCTATGGCCAAGTTGACC MP115 (Xba I) - ATAATTCTAGATCAGTCCTGCTCCTC (Amplified from pcDNA-3.1/ZEO plasmid from Invitrogen)	374	MP116 (Xba I) - ATAATTCTAGAAGGCCTACGGTTCTCTCC MP117 (Cla I) - GACTGATCGATATCATGAGAGATGACGG (Amplified from CBS 6938 gDNA)	396	MP118 - ATCGGCTCATCAGCCGACAGTTCATC MP119 - ATCATGAGAGATGACGGAGATGATGG	1175	Primers MP118/MP119 were used to sequence the complete resistance cassette

Table II.5. Primers and plasmid used for the construction of the complementation plasmids.

Complemented gene	Backbone Plasmid	Primers (5'-3')	Amplified fragment (bp)	Description of amplified region	Complementation plasmid	Linearization of complementation plasmid	Mutant strain transformed with complementation plasmid	Primers used for mutant confirmation (5'-3')	Amplified fragment (bp)	Description
MFA2	pUC18+rDNA+ZEO	MP134 (Pst I) – ATTATCTGCAGACCTCGAGACCCAGCTCCG MP135 (BamH I) - ATTATGGATCCAGTGGTCGCCAAAGCTGACG	1122	MFA gene plus 500bp upstream and downstream of the gene	pUC18+rDNA+ZEO+MP134/MP135	Cla I	CBS 6938_ste3-2/mfa2Δ	MP102 - TATATCATCTCTCGACCC MP103 - TTCATCTTGTCAGACAGC	611	Amplification of the *MFA2* (complete) region
STE3-1	pUC18+rDNA+ZEO	MP136 (Pst I) - TATTACTGCAGTGTATCTCAAGTCTTGCC MP137 (BamHI) - TATTAGGATCCTTTCCGAGATTGATGGTCG	1994	STE3-1 gene plus 300bp upstream and downstream of the gene	pUC18+rDNA+ZEO+MP136/MP137		CBS 6938_ste3-1/ste3-2Δ	MP029 - CGGTGGGCGTTCTGGTCGGAC MP030 - ACTCTGATGGCGAAGCAACGGC	836	Amplification of the *STE3-1* gene (partial)
HD1	pUC18+rDNA+ZEO	MP138 (Pst I) - AAAAACTGCAGGAAAGATAGTGAAACTGG MP139 (BamH I) - AATTTGGATCCTTTCATCGGGCTCAAGC	2459	HD1 gene plus 300bp upstream and downstream of the gene	pUC18+rDNA+ZEO+MP138/MP139		CBS 6938_hd1/hd2Δ	MP083 - AAAACCTCTCCAAGCGCC MP084 - TCTACATCGGTCTCTTCC	1002	Amplification of the *HD1* gene (partial)
							CBS 6938_hd1Δ	MP083 - AAAACCTCTCCAAGCGCC MP084 - TCTACATCGGTCTCTTCC	1002	Amplification of the *HD1* gene (partial)

Table II.6. Sporulation data pertaining to plot (e) from Figure 3.2, plot (c) from Figure 3.3 and plot (a) from Figure 3.4.

	Assay	No. basidia per plate 1	No. basidia per plate 2	No. basidia per plate 3	Mean no. of basidia per plate	Mean no. of basidia per plate (trunc.)	Mean	Standard deviation
CBS 6938	1	836	1035	501	790.7	791	905.0	139.1
	2	995	981	1203	1059.7	1060		
	3	621	969	1003	864.3	864		
ste3-1Δ	1	1012	571	620	734.3	734	814.0	92.9
	2	743	851	783	792.3	792		
	3	943	854	950	915.7	916		
ste3-2Δ	1	1231	1153	918	1100.7	1101	1142.7	113.3
	2	1052	920	1842	1271.3	1271		
	3	820	998	1351	1056.3	1056		
hd2Δ	1	8	23	15	15.3	15	13.7	2.3
	2	4	12	16	10.7	11		
	3	20	13	13	15.3	15		
spo11Δ	1	587	387	720	564.7	565	649.3	77.6
	2	503	881	610	664.7	665		
	3	532	689	932	717.7	718		

Table II.7. Sporulation data pertaining to plot (f) from Figure 3.2. Additional sporulation data was also included for complementation mutant *hd1Δhd2Δ+HD1.*

	Assay	No. basidia per plate 1	No. basidia per plate 2	No. basidia per plate 3	Mean no. of basidia per plate	Mean no. of basidia per plate (trunc.)	Mean	Standard deviation
CBS 6938	1	904	832	861	865.7	866	897.3	36.2
	2	828	903	937	889.3	889		
	3	809	1004	998	937	937		
ste3-1Δ /mfa2Δ	1	854	743	792	796.3	796	881.7	76.5
	2	950	986	893	943	943		
	3	991	701	1027	906.3	906		
ste3-2Δ /mfa1Δ	1	725	1136	1021	960.7	961	1004.3	38.9
	2	1118	1002	928	1016	1016		
	3	915	1203	989	1035.7	1036		
hd1Δhd2Δ+HD1	1	39	32	22	30.7	31	43.7	11.4
	2	42	41	59	47.3	47		
	3	53	60	47	53.3	53		

Table II.8. Data from crosses of double and triple mutant strains.

	Number of basidia per plate		
Strains	Assay 1	Assay 2	Assay 3
CBS 6938	7900*		
ste3-1Δmfa1Δ × *ste3-2Δmfa2Δ*	70	38	21
ste3-1Δmfa1Δhd1Δ × *ste3-2Δmfa2Δhd2Δ*	9	0	2

**estimated value*

Table II.9. List of plasmids used in the Bacterial Adenylate Cyclase Two-Hybrid System.

Plasmids	Relevant features
pUT18	*High copy number vector that encodes the T18 fragment that is fused downstream of the MCS. This vector allows creating in-frame fusions at the N-terminal end of T18. BACTH System Kit from EUROMEDEX*
pKNT25	*Low copy number vector that encodes the T25 fragment that is fused in frame downstream from a MCS. This allows creating in-frame fusions at the N-terminal end of T25. BACTH System Kit from EUROMEDEX*
pUT18C-ZIP	*The plasmids pKT25-zip and pUT18C-zip serve as positive controls for complementation in the BACTH System Kit from EUROMEDEX. When pKT25-zip and pUT18C-zip are co-transformed into BTH101 E. coli cells, they restore a Cya+ phenotype. Vector pKT25-zip has a leucine zipper genetically fused in frame to the T25 fragment, while pUT18C-zi has a leucine zipper genetically fused in frame to the T18 fragment.*
pKNT25-ZIP	
pEX-K4-HD1	*Eurofins in-house standard vector with HD1 synthetic gene*
pEX-K4-HD2	*Eurofins in-house standard vector with HD2 synthetic gene*
pKNT25+HD1	*Plasmid with a HD1 in frame fusion at the N-terminal end of T25 subunit*
pKNT25+HD2	*Plasmid with a HD2 in frame fusion at the N-terminal end of T25 subunit*
pUT18+HD1	*Plasmid with a HD1 in frame fusion at the N-terminal end of T18 subunit*
pUT18+HD2	*Plasmid with a HD2 in frame fusion at the N-terminal end of T18 subunit*

Table II.10. List of primers used in the Bacterial Adenylate Cyclase Two-Hybrid System.

Primers (5'-3')	Relevant information
MP177(Hind III) - ACGCCAAGCTTGATGGAACCGAGTCCG	*MP177-MP178 amplify a 1543 bp fragment containing HD1 synthesized gene present on plasmid pEX-K4-HD1, with restriction sites for Hind III and Pst I in the fragment extremities*
MP178 (Pst I) - TCGACCTGCAGGCGCAGTCGAATACTTC	
MP179 (Hind III) - ACGCCAAGCTTGATGTATAGCTTCACC	*MP179-MP180 amplify a 1261 bp fragment containing HD2 synthesized gene present on plasmid pEX-K4-HD2, with restriction sites for Hind III and Pst I in the fragment extremities.*
MP180 (Pst I) - TCGACCTGCAGGCTTTCCCCGTTGTCATG	
MP191 - ACACTTTATGCTTCCGGC	*Allow the amplification and sequencing of the constructions performed in plasmid pKNT25.* *Fragment MP191-MP192 amplified in:* *- empty pKNT25 plasmid: 176bp* *- pKNT25+HD1:1694bp* *- pKNT25+HD2:1412bp*
MP192 - ACCAGCCTGATGCGATTGC	
MP193 - TGGCACGACAGGTTTCCC	*Allow the amplification and sequencing of the constructions performed in plasmid pUT18.* *Fragment MP193-MP194 amplified in:* *- empty pUT18 plasmid: 284 bp* *- pUT18+HD1: 1802 bp* *- pUT18+HD2: 1520 bp*
MP194 - ACAAGTCGATGCGTTCGC	

Table II.11. List of plasmids used in the Matchmaker Gold Yeast Two-Hybrid System.

Plasmids	Relevant features
pGBKT7 DNA-BD	*plasmid encoding the yeast Gal4 DNA-binding domain*
pGADT7 AD	*plasmid encoding the yeast Gal4 activation domain*
pGBKT7-53	*Plasmid use for positive control of protein interaction*
pGADT7-T	*Plasmid use for positive control of protein interaction*
pGBKT7-Lam	*Plasmid use for negative control of protein interaction*
pGBKT7+MP181/MP182	*pGBKT7 plasmid containing HD1 gene (short version)[1] cloned in frame with the GAL4 DNA-BD*
pGBKT7+MP181/MP183	*pGBKT7 plasmid containing HD1 gene cloned in frame with the GAL4 DNA-BD*
pGBKT7+MP184/MP185	*pGBKT7 plasmid containing HD2 gene (short version)[1] cloned in frame with the GAL4 DNA-BD*
pGBKT7+MP184/MP186	*pGBKT7 plasmid containing HD2 gene cloned in frame with the GAL4 DNA-BD*
pGADT7+MP187/MP188	*pGADT7plasmid containing HD1 gene (short version)[1] cloned in frame with the GAL4 activation domain*
pGADT7+MP189/MP190	*pGADT7plasmid containing HD2 gene (short version)[1] cloned in frame with the GAL4 activation domain*

[1]Short version of the genes HD1 and HD2 comprise the complete N-terminal region and homeodomain region of each of the genes (first 183 amino acids of HD1 and the first 196 amino acids of HD2).

Table II.12. List of primers used in the Matchmaker Gold Yeast Two-Hybrid System.

Primers (5'-3')	Relevant information
MP181-CATCATGGAGGAGCAGAAGCTGATCTCAGAGGAGGACCTGATGGAACCGAGTCCGGCG	*Amplifies fragment comprising the complete N-terminal and homeodomain regions of the HD1 gene plus a tail of 40 bp at both 5'[1] and 3'[2] extremities (complementary to the pGBKT7 plasmid)*
MP182-TCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTACATCCAGCCGCTGCGACG	
MP183-TCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTATTAGCAGTCGAATACTTCGCCC	*With MP181 primer, amplifies fragment comprising the complete HD1 gene plus a tail of 40 bp at both 5'[1] and 3'[2] extremities (complementary to the pGBKT7 plasmid)*
MP184-CATCATGGAGGAGCAGAAGCTGATCTCAGAGGAGGACCTGATGTATAGCTTCACCAAGTTGCC	*Amplifies fragment comprising the complete N-terminal and homeodomain regions of the HD2 gene plus a tail of 40 bp at both 5'[1] and 3'[2] extremities (complementary to the pGBKT7 plasmid)*
MP185-TCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTAAGAGCGCCGATTACGGCG	
MP186-TCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTATTATTTCCCCGTTGTCATGTTCAGG	*With MP184 primer, amplifies fragment comprising the complete HD2 gene plus a tail of 40 bp at both 5'[1] and 3'[2] extremities (complementary to the pGBKT7 plasmid)*
MP187-CGCCGCCATGGAGTACCCATACGACGTACCAGATTACGCTATGGAACCGAGTCCGGCG	*Amplifies fragment comprising the complete N-terminal and homeodomain regions of the HD1 gene plus a tail of 40 bp at both 5'[3] and 3'[4] extremities (complementary to the pGADT7 plasmid)*
MP188-TTGAAGTGAACTTGCGGGGTTTTTCAGTATCTACGATTCACATCCAGCCGCTGCGACG	
MP189-CGCCGCCATGGAGTACCCATACGACGTACCAGATTACGCTATGTATAGCTTCACCAAGTTGCC	*Amplifies fragment comprising the complete N-terminal and homeodomain regions of the HD2 gene plus a tail of 40 bp at both 5'[3] and 3'[4] extremities (complementary to the pGADT7 plasmid)*
MP190-TTGAAGTGAACTTGCGGGGTTTTTCAGTATCTACGATTCAAGAGCGCCGATTACGGCG	

[1]5'-tail before MCS in pGBKT7 plasmid: CATCATGGAGGAGCAGAAGCTGATCTCAGAGGAGGACCTG
[2]3'-tail after MCS in pGBKT7 p plasmid: TCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTA
[3]5'-tail before MCS in pGADT7 plasmid: CGCCGCCATGGAGTACCCATACGACGTACCAGATTACGCT
[4]3'-tail after MCS in pGADT7 plasmid: TTGAAGTGAACTTGCGGGGTTTTTCAGTATCTACGATTCA

Table II.13. List of putative homeodomain proteins retrieved from CBS 6938 genome.
Available at: https://figshare.com/s/feb854851e12f59d2d1a

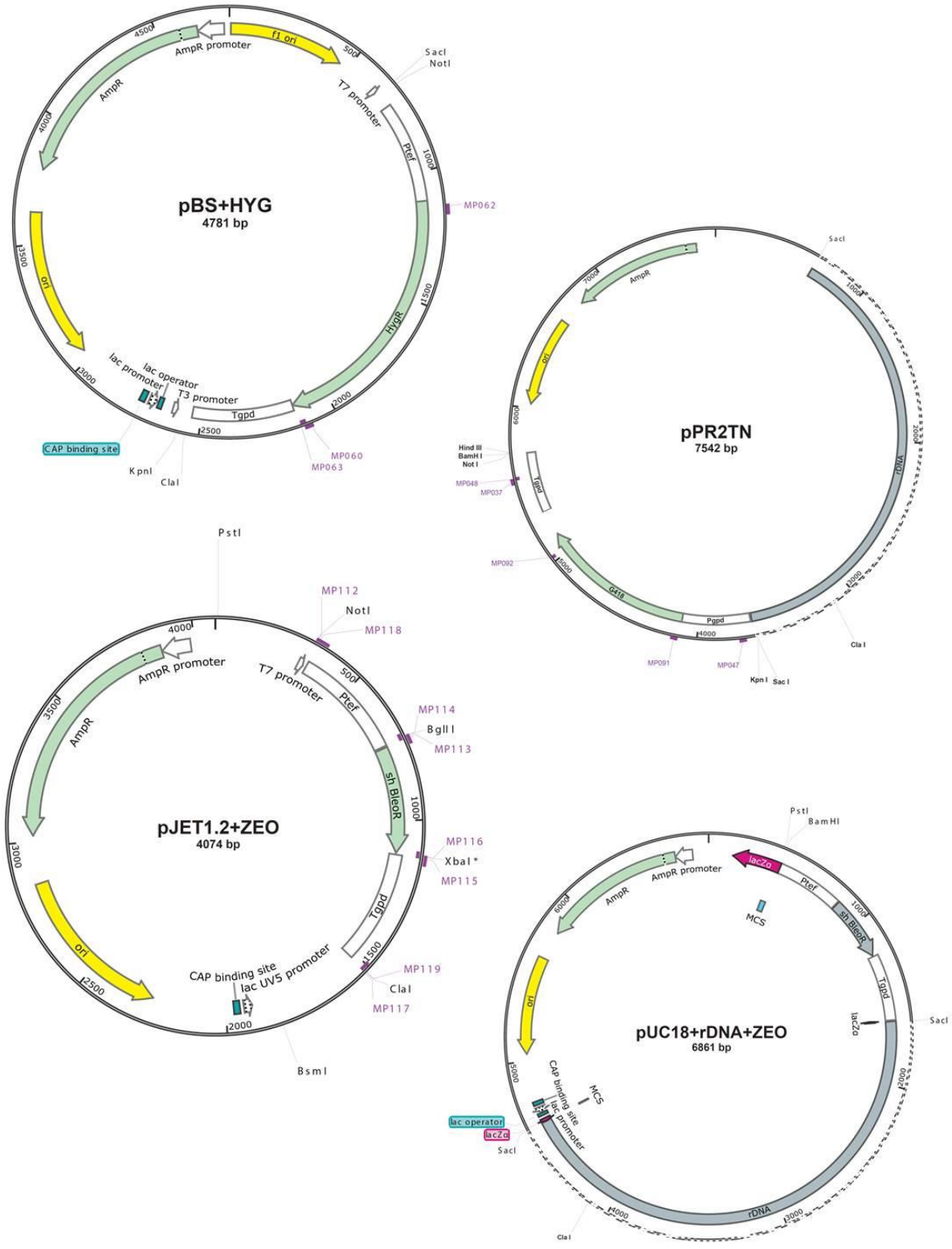
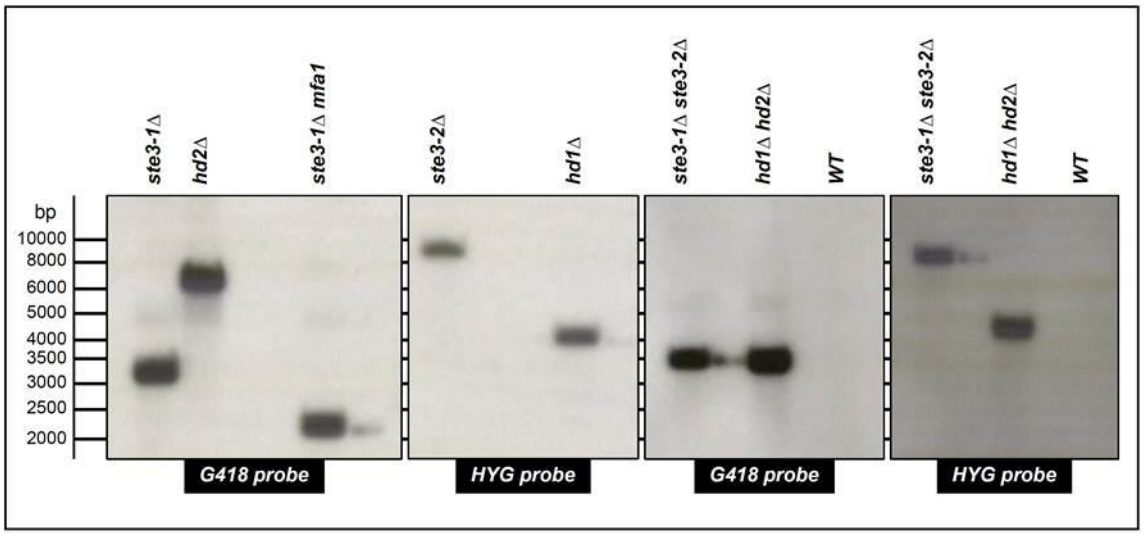


Figure II.1. Plasmids used and constructed in this work. Plasmid pBS-HYG and pPR2TN were used to construct deletion fragments used in this work. Plasmids pJET1.2 and pUC18 were used to generate pJET1.2+ZEO and pUC18+rDNA+ZEO plasmids. Depicted in each plasmid map are the primers (indicated as MP followed by their designation number) and enzyme restriction sites used for the construction of the deletion fragments used in this work.



Strains	Relevant genotypes	Expected fragment size / used probe
ste3-1Δ	*ste3-1Δ::G418*	3196 bp / G418
hd2Δ	*hd2Δ::G418*	6651 bp / G418
ste3-1Δ mfa1Δ	*ste3-1/mfa1Δ::G418*	2213 bp / G418
ste3-2Δ	*ste3-2Δ::HYG*	8337 bp / HYG
hd1Δ	*hd1Δ::HYG*	4332 bp / HYG
ste3-1Δ ste3-2Δ	*ste3-2Δ::HYG/ste3-1Δ::G418*	3196 bp / G418; 8337 bp / HYG
hd1Δ hd2Δ	*hd1Δ::HYG/hd2Δ::G418*	3189 bp / G418; 4332 bp / HYG
WT (CBS 6938)		No bands expected with any of the probes

Figure II.2. Southern blot analysis of selected deletion mutants. Autoradiographs of the results obtained for each mutant. Expected fragment sizes are shown.

HD1 (cloned into *Eurofins in-house standard vector pEX-K4-HD1)*

ATGGAACCGAGTCCGGCGAGTGACCTGTCAGCGCGTCTTGACAGTCTCGTGGAAACGTTCCTGATCAGCATTCGC
GCGGGAGTGCCGATCGAATTATCCCAGATCTACGATGCGTTCGCCGAAGAACGCTGCAAAGCGCGTTCTCAAGGA
CTGCTGACCTTCGCCTTTCAGGAAGAAATCTCCGCTTCAGTTGGCTTTCTCTCGACTCTCGCCTCGTGCATGATGAA
TATTCCGGAAGGCAACAAAGACTTTCAGACCACCGTCATTGGCATCTTAAACAATGCAAACAATACGCCAGAGCTGT
GTCCTAGCCGCAACTTAGTTTCCCCTCCGAAGATTCTGGCATCGAAAGCGGATTCGGGCCGCTCGATGCGTCCAC
GCACAGAGCATC**CCTCTTCGTTTCGGCATCTGCGCGATTGGTTCATTGTGCATCTGGATCATCCGTATCCTTCTGC**
TCAGGAGAAAGAGGAGCTTGCTCTGGCACGTAACCTGACCAAGAACTCGATCAACCTGTGGTTCAATAATATGC
GTCGTCGCAGCGGCTGGATGGATTTCGTACGCACACACGCGTTGAACAATCAGACCCGCATGAAAGAACTGGTGT
TGCTTCACCGCTCTCTGCCTGCCGTTGAACGTTCACATCCGGACCTTCCGTTTGAGCTGCGGCAGCAGCTCGATC
GCATTCGCGATCATGTGGATGAGAAAGGTCATGATCGCGTGCGTGATTGGATGACAGGTGCACTGGACTACAAAG
ACGGTTTTGGGGTCCCCGCCGAAGGAATTGGCCGTAGCCCCGAAAAGAAGAACGAAGTCAACAAAGATACCGCGA
AAAGCGGTTCCAGCTCCATCTTGTACAAACAAGCCTTTACTGCGAAAGATGCACAGCCGTGGCTGACCAGTTACCC
CACCGCCGCTAGCGGCTTGATTGGCTATCCGAGTTCAATCCCGCTGGCGAATTGTCCGACGCCGAAAAGTAGCAA
ACGGCTGTTTTCGGATTGTGGTCCGCACTTAGCTGCACCGATTGACCCGGTCAACAAATACCACCAGTTAATTAAC
CTCAGCTCAGGTCCTACCCAGGATCTGGAGGGGTCTTTGCTGACTCACCAACCATTGAGCGCGTATGCTTACTATC
CTACTCATCCGCAATCCCAGAATCCGTTACCAATGCTGGGTCCCACACCGTATCGTTCAAGCAGCGATTCGCAGTA
TCCAGCCTATGACCTGAAAGCACCGATTCGCGCGCAAACCTGGCTGCCGTCTGACCCTGCGATTGATCCGTCAAAT
GGGCCAAACATTCACCAACGTAGCGTTTCCCCATTACCAAGCCGTGCGAAAAGCACGGTATCCTCTCTGGCAACGA
GTACCACGCTCCAAAGCACGCGCTTTTCTGTGAATGAAAGTGCCGGTAATGAACTGGCGGCCAAACTGGAAGCCC
TTCGCATCAAGAAAGAAGCGCTGAAGAAAGAGAAAGAACAGTTGCGCGCCGAAGCTGAAGCACTGGGCGAAGTAT
TCGACTGCTAA

HD2 (cloned into *Eurofins in-house standard vector pEX-K4-HD2)*

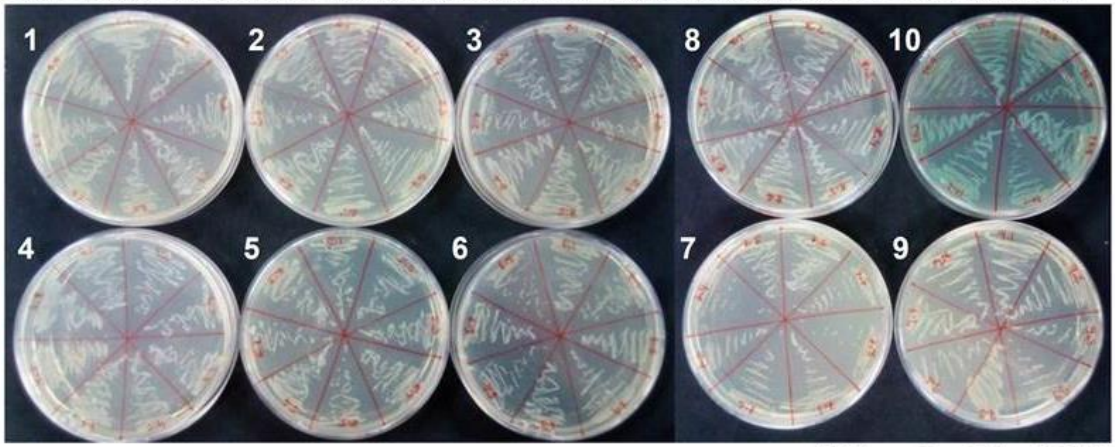
ATGTATAGCTTCACCAAGTTGCCGCAGAAATACACTGGTTTAGGAGATGTCTTTGATTCTAACATTGACAGTTTGAG
CAGCAAAGAGTTTGCTCTCGAAGATCTGTATCTGCCAATCCCGACGTCCATTTGCGAGACTTTGGAGGAACTGGGT
ACGCCGCATACCCTGGCCCGCAATTTAGGCGCCCTGTACCTGAGTCGTGCGATGGCGCTGCGCGAACTGAGTTTA
TCGCAGTTTGACACCGCATTCCCTGCCGAACGCACGTTTCTCTTGGAAGCCAACTCCCTCAAACTGAAAGTGTTTAA
AGGCCTGCTTAAAGCGACTCTGGAAAATCGCTACAGCATGGAACTGTTTCACCATCACGAAACCATCATTCAGCTG
GCCGGTGTCTGGACCT**CAGACGATGAACAAGCACCAGTATCCGTACGCCGTAGCTCATTCACGAAGAATACGAT**
TGACACCCTCCAAGCCCTGTATGACCTGAACCAATATCCGAACCCGATTGAGCTGAAAACCATTGCAGTTAAAGT
GGGCTTACAGCCGAAACAGATTCGTGCTTGGTTTCAGAATCGCCGTAATCGGCGCTCTAGCCCCGCACGGTCCC
ATAAGAAACCCCTTAAAATCTCTACCCGTGCTACTGTTGCGAGTGACCTGAAGTTCAACAGCGTCCGCAAACGTTGT
TCGCGCCGTACTACAAGCATCGAAAAGGAGCGCAAAACCGAGGGTGTGCCGTCACAACTGGGCTCACCGATCTTC
TTGCAGACCAACAAAACCCCATACGTTCCTCAGTCGGAAGGCCGCTCTCGCGTTGGTAATTCGTTAGTGGCAGGAC
TGTTCGATACATCGCCCAACGAACAGGATGCGTATGCGCTTTGGAGTGCCGTTGTGAGCGATCGTAGTAGCGCATA
TGCGAACAGTAAGCATGTGCCTCAGATCGAACCATCTCTGGCGTTTGCCATTCCGGAACCGGCGGATAGCATGTC
CGCGCAACGTTCGGAGAATCACTTCAATTCCTGCTCGAAATCGTCCAGTAGCCATCCGGCGTTTCCTCCGGTGTTT
GGGTCTAGCGAAACGATGGAAGTCGATCCGGTAGATTGGACCATGGTTTTCGCGGCTTCAGACTCTCGCAATATTG
CACCACGTAGCGAGAAACCGTCAGCTCGCTGGGAAACGAAAACAGGCCTTGAAGTCGAAGATGCCATGCTGGATG
TGAACCTGAACATGACAACGGGGAAATAA

Figure II.3. Synthetic coding sequence of the *HD1* and *HD2* genes. Parts of the sequences underline correspond to the synthetic cDNA sequence coding for the N-terminal regions of each protein, while parts of the sequences in bold correspond to the synthetic cDNA sequence coding for the homeodomain region.

a.

Plate number	Plasmids present in BTH101 E. coli	Testing
1	pKNT25+HD1 and pUT18+HD2	heterodimerization
2	pKNT25+HD2 and pUT18+HD1	heterodimerization
3	pKNT25+HD1 and pUT18+HD1	homodimerization
4	pKNT25+HD2 and pUT18+HD2	homodimerization
5	pKNT25 and pUT18	negative control
6	pKNT25+HD1 and pUT18	negative control
7	pKNT25+HD2 and pUT18	negative control
8	pKNT25 and pUT18+HD1	negative control
9	pKNT25 and pUT18+HD2	negative control
10	pKT25-zip and pUT18C-zip	positive control

b. **Luria-Bertani media supplemented with: 100µg/ml ampicillin, 50µg/ml kanamycin, 0,5mM IPTG and 40µg/ml X-gal**



c. **MacConkey media supplemented with: 1% Maltose, 100µg/ml ampicillin, 50µg/ml kanamycin and 0,5mM IPTG**

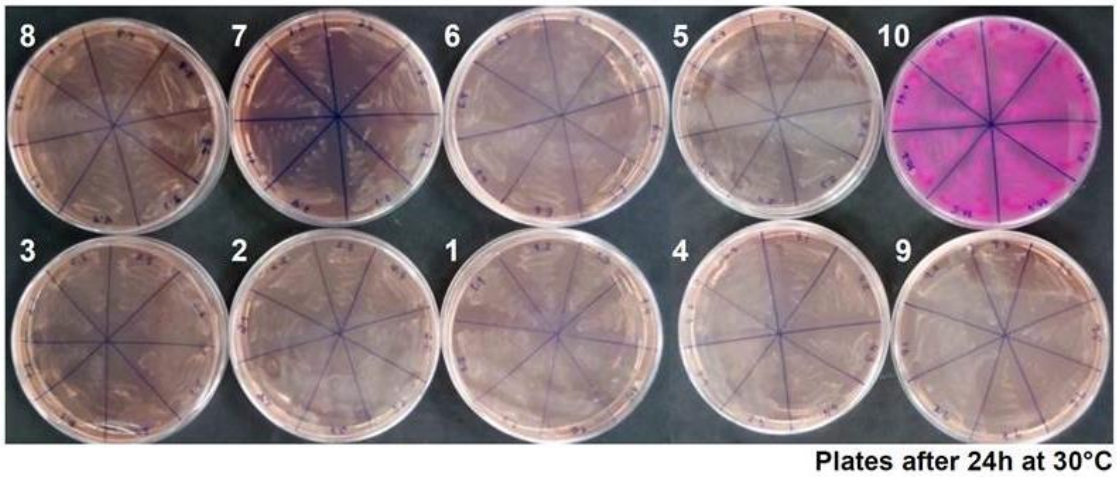


Figure II.4. Bacterial two-hybrid assays. *Escherichia coli* BHT101 transformants expressing various **(a)** combinations of fusion proteins, **(b)** eight replicates per plate growing on LB/ X-gal media and on **(c)** MacConkey/maltose medium plates.

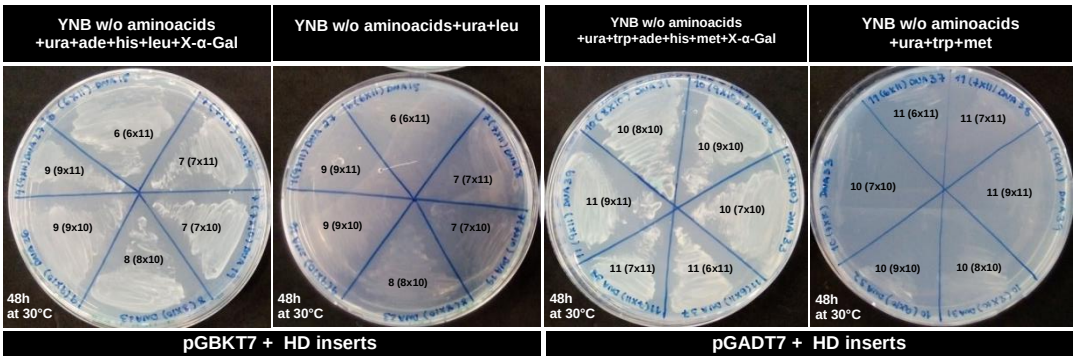
a.

Haploid id number	Plasmid in the haploid S. cerevisiae strain
1	pGBKT7 in Y2HGold cells
2	pGADT7 in Y187 cells
3	pGBKT7-53 in Y2HGold cells
4	pGADT7-T in Y187 cells
5	pGBKT7-Lam in Y2HGold cells
6	pGBKT7+MP181/182 in Y2HGold cells
7	pGBKT7+MP181/MP183 in Y2HGold cells
8	pGBKT7+MP184/MP185 in Y2HGold cells
9	pGBKT7+MP184/MP186 in Y2HGold cells
10	pGADT7 +MP187/MP188 in Y187 cells
11	pGADT7 +MP189/MP190 in Y187 cells

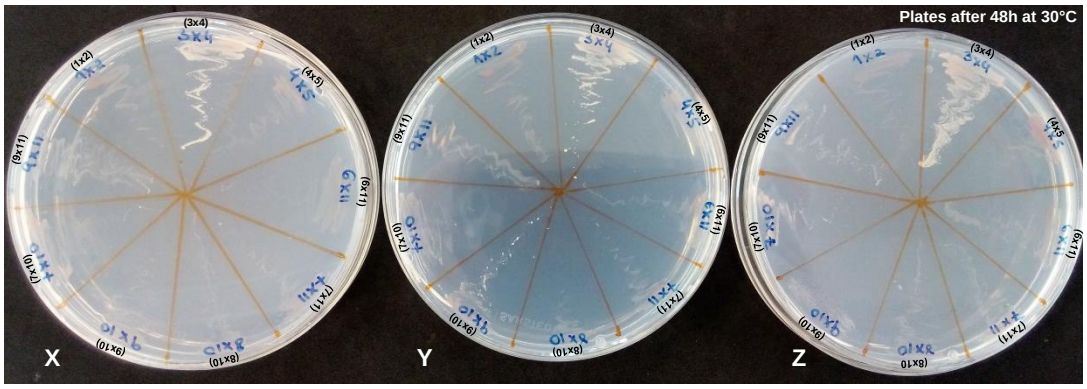
c.

Mating experiments using the haploid strains	Objective
1X2	Negative control with empty plasmids
3X4	Positive control for interaction
4X5	Negative control for interaction
6X11	Testing for heterodimerization hd1Xhd2
7X11	Testing for heterodimerization HD1Xhd2
8X10	Testing for heterodimerization hd2Xhd1
9X10	Testing for heterodimerization HD2Xhd1
7X10	Testing for homodimerization HD1Xhd1
9X11	Testing for homodimerization HD2Xhd2

b.



d. Yeast nitrogen base media (without amino acids) supplemented with: 0,02 mg/mL uracil, 2% agar and 2%



e Yeast nitrogen base media (without amino acids) supplemented with: 0,02 mg/ml uracil, 0,02 mg/ml histidine, 0,02 mg/ml adenine, 2% agar, 2% glucose and 40µg/ml X-α-Gal

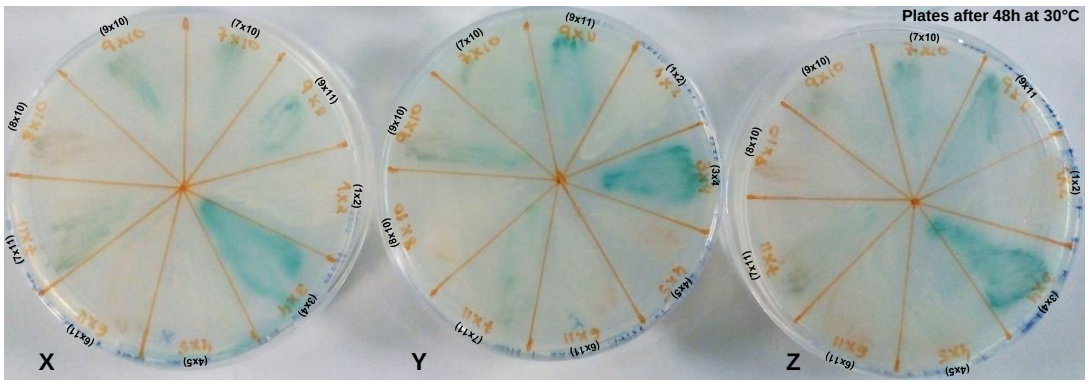


Figure II.5. Yeast two-hybrid assays. (a) List of haploid *S. cerevisiae* strains with the respective plasmids; **(b)** Strains with just the individual fusion proteins were unable to activate transcription of the any of the reporter

genes reporter genes. Identification of the haploids, like 6 (6x11), should be read as: haploid 6 used in the mating cross between haploid 6 and haploid 11; **(c)** Mating of the haploid strains with each other in different combinations; Plates with three independent diploid strains (X, Y and Z) expressing each of the combinations of fusion proteins growing on selective medium to assess adenine and histidine prototrophy **(d)** and indicator medium containing X-alpha GAL to assess expression of the *MEL1* reporter gene **(e)**.

APPENDIX III

Additional information pertaining to Chapter 4

III.1 Physiological characteristics of *P. novazelandica* and *P. tasmanica*

Table III.1. Physiological tests for three strains of each of the new *Phaffia* species. The different letters relate to the growth of each strain in each test: w = weak growth; s = slow growth; d or l= delayed growth. Results for *P. rhodozyma* were taken from bibliography (Miller *et al.* 1976; Barnett 2000).

				P. rhodozyma	*P. novazelandica*			*P. tasmanica*		
Type of test			**Test**	**CBS 7918^T**	**ZP 938**	**ZP 884**	**ZP 909**	**ZP 863**	**ZP 878**	**ZP 875**
Growth	Cabon sources	C1	D-Glucose	+	+	+	+	+	+	+
		C2	D-Galactose	-	-	-	-	-	ws	ws
		C3	L-Sorbose	-	l	-	-	-	-	-
		C4	D-Glucosamine	-	-	-	-	-	-	-
		C5	D-Ribose	d	-	-	-	-	-	-
		C6	D-Xylose	d	+	-	+	l	l	-
		C7	L-Arabinose	+	+	l	+	+	+	+
		C8	D-Arabinose	-	-	-	-	-	-	-
		C9	L-Rhamnose	-	-	-	-	-	-	-
		C10	Sucrose	+	+	+	+	+	+	+
		C11	Maltose	+	+	+	+	+	+	+
		C12	α, α-Trehalose	+	+	+	+	+	+	+
		C13	Methyl-α-D-Glucoside	d	+	s	+	+	s	s
		C14	Cellobiose	+	+	+	+	+	s	s
		C15	Salicin	+	+	+	+	+	+	+
		C17	Melibiose	-	+	s	+	+	+	+
		C18	Lactose	-	-	-	-	-	-	-
		C19	Raffinose	+	+	+	+	+	+	+
		C20	Melezitose	+	+	+	+	+	+	+
		C21	Inulin	-	w	w	w	w	w	ws
		C22	Starch	+	+	+	+	+	+	+
		C23	Glycerol	+	+	-	+	w	+	-
		C24	Erythritol	-	-	-	-	-	-	-
		C25	Ribitol	d	s	s	w	-	-	-
		C26	Xylitol	-	+	w	w	-	-	-
		C28	D-Glucitol	d	+	+	+	-	-	-
		C29	D-Mannitol	+	+	+	+	+	+	+
		C30	Galactitol	-	-	-	-	-	-	-
		C31	*myo*-Inositol	-	-	-	-	w	-	-
		C32	D-Glucono-1,5 lactone	+	+	+	+	+	+	+
		C35	D-Gluconate	+	+	+	+	w	w	ws
		C36	D-Glucuronate	d	+	+	+	w	w	ws
		C37	D-Galacturonate	d	+	+	+	-	+	-
		C38	DL-Lactate	d	-	-	-	-	-	-
		C39	Succinate	+	+	+	+	+	+	+
		C40	Citrate	+	+	+	+	+	+	+
		C41	Methanol	-	-	-	-	-	-	-
		C42	Ethanol	+	+	+	+	+	+	+
	Nitrogen sources	N1	Nitrate	-	-	-	-	-	-	-
		N2	Nitrite	-	-	-	-	-	-	-
		N3	Ethylamine	-	+	+	+	-	-	-
		N4	L-Lysine	w	+	+	+	+	+	w
		N5	Cadaverine	+	w	w	w	s	s	ws
		N6	Creatine	-	-	-	-	-	-	-
		N7	Creatinine	w, -	-	-	-	-	-	-

			P. rhodozyma	*P. novazelandica*			*P. tasmanica*		
Type of test		**Test**	**CBS 7918^T**	**ZP 938**	**ZP 884**	**ZP 909**	**ZP 863**	**ZP 878**	**ZP 875**
Fermentation	F1	D-Glucose	**w, d**	**s+**	**w**	**w**	**w**	**+**	**+**
	F2	D-Galactose	**-**	**-**	**-**	**-**	**-**	**-**	**-**
	F3	Maltose	**d,-**	**-**	**-**	**-**	**-**	**s**	**-**
	F4	Me α-D-Glucoside	**d,-**	**-**	**-**	**-**	**-**	**-**	**-**
	F5	Sucrose	**w, d**	**+**	**-**	**w**	**s+**	**+**	**+**
	F6	α, α-Trehalose	**d**	**-**	**-**	**-**	**-**	**-**	**-**
	F7	Melibiose	**-**	**-**	**-**	**-**	**-**	**w**	**-**
	F8	Lactose	**-**	**-**	**-**	**-**	**-**	**-**	**-**
	F9	Cellobiose	**d, -**	**-**	**-**	**-**	**-**	**-**	**-**
	F10	Melezitose	**w, d**	**-**	**-**	**-**	**s+**	**+**	**+**
	F11	Raffinose	**-**	**-**	**-**	**-**	**ws**	**w**	**w**
	F12	Inulina	**-**	**-**	**-**	**-**	**-**	**-**	**-**
	F13	Starch	**-**	**ws**	**-**	**-**	**ws**	**l**	**ws**
	F14	D-Xylose	**-**	**-**	**-**	**-**	**-**	**-**	**-**
Additional	V1	w/o vitamins	**-**	**-**	**-**	**-**	**-**	**-**	**-**
	O1	Cycloheximide 0.01%	**-**	**-**	**-**	**-**	**-**	**-**	**-**
	O2	Cycloheximide 0.1%	**-**	**-**	**-**	**-**	**-**	**-**	**-**
	M1	Starch formation	**+**	**+**	**+**	**+**	**+**	**+**	**+**
	M3	Urea Hydrolysis	**+**	**+**	**+**	**+**	**+**	**+**	**+**

Table III.2. Statistics for the genomes assembled for this work (Chapter 4). All statistics are based on contigs of size >= 500 bp.

Species (strain)	Total length (Mb)	No. Scaffolds	Largest scaffold (Kb)	GC%	N50	Average Coverage (x)
T. pullulans (JCM 9886)	24.0	535	825.7	58.5	240.005	86
T. pamirica (JCM 10408)	26. 3	2138	290.3	66.1	57.646	68
M. aquatica (JCM 1775)	21.1	340	830.8	56.5	265.958	96
U. megalosporus (JCM 5269)	21.5	409	1027.3	51.6	343.153	98
U. pyricola (JCM 2958)	26.1	275	1386.9	52.0	373.346	96
P. novazelandica (CBS 14095)	18.7	742	278.8	46.7	70.416	22
P. tasmanica (CBS 14096)	19.1	977	209.2	45.8	60.950	24
K. huempii (PYCC 5836)	23.2	653	319.2	47.7	78.279	13
C. capitatum (CBS 7420)	19.5	3492	63.6	58.6	11.773	21
C. ferigula (CBS 7201)	18.0	3980	83.2	63.8	9.761	30
C. macerans (CBS 6532)	20.4	4489	104.9	65.2	12.269	25
C. bisporidii (CBS 6347)	20.5	3974	59.7	63.8	10.033	25

III.2 Additional information pertaining to the species phylogeny and to the *MAT* genes found in of *P. novazelandica* and *P. tasmanica*

- Complete list of Rpa1, Rpa2, Rpb1, Rpb2, Rpc1 and Rpc2 predicted proteins
- Alignment used to infer the Cystofilobasidiales species phylogeny (Figure 4.6)
- Alignment used to infer the STE3 phylogeny (Figure 4.8)
- Alignment used to infer the MFA phylogeny (Figure 4.9)

Available at: https://figshare.com/s/a348cb3edc03d9757776

III.3 Additional information pertaining to the *MAT* and astaxanthin related genes found in of *P. novazelandica* and *P. tasmanica*

a.

```
                         10        20        30        40        50        60        70        80
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtE_CBS7918             MDYANILTAIPLEFTPQDDIVLLEPYHYLGKNPGKEIRSQLIEAFNYWLDVKKEDLEVIQNVVGMLHTASLLMDDVEDSS
crtE_ZP938               MDYANILTAVPLEFTPQDDIVLLEPYHYLGKNPGKEIRSQLIEAFNFWLDVKKEDLEVIQNVVGMLHTASLLMDDVEDSS
crtE_ZP875               MDYANILTAIPLEFTPQDDIVLLEPYHYLGKNPGKEIRSQLIEAFNFWLDVKKEDLEVIQNVVGMLHTASLLMDDVEDSS
Clustal Consensus        *********:************************************:********************************

                         90        100       110       120       130       140       150       160
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtE_CBS7918             VLRRGSPVAHLIYGIPQTINTANYVYFLAYQEIFKLRPTPIPMPVIPPSSASLQSSVSSASSSSSASSENGGTSTPNSQI
crtE_ZP938               VLRRGSPVAHLIYGIPQTINTANYVYFLAYQEIFKLRPTPTPTPVIPPSSASLQSSVSSASSSSSTSSDHGGTSTPNSQI
crtE_ZP875               VLRRGSPVAHLIYGIPQTINTANYVYFLAYQEIFKLRPTPTPIPVMPPSSASLQSSVSSASSFSSISSENGGTSTPNSQI
Clustal Consensus        ************************************ * **:**************** ** **::**********

                         170       180       190       200       210       220       230       240
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtE_CBS7918             PFSKDTYLDKVITDEMLSLHRGQGLELFWRDSLTCPSEEEYVKMVLGKTGGLFRIAVRLMMAKSECDIDFVQLVNLISIY
crtE_ZP938               PFSKDTYLDKVITDEMLSLHRGQGLELFWRDSLTCPSEEEYVKMVLGKTGGLFRIAVRLMMAKSECDIDFVQLVNLISIY
crtE_ZP875               PFSKDTYLDKVITDEMLSLHRGQGLELFWRDSLTCPSEEEYVKMVLGKTGGLFRIAVRLMMAKSECDIDFIQLVNLISIY
Clustal Consensus        ************************************************************************:*********

                         250       260       270       280       290       300       310       320
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtE_CBS7918             FQIRDDYMNLQSSEYAHNKNFAEDLTEGKFDTLQKKSTSPEILHHCVNYMRTETHSFEYTQEVLNTLSGALERELGRLQG
crtE_ZP938               FQIRDDYMNLQSSEYAHNKNFAEDLTEGKFNTLQKKSTSPEILHHCVNYMRTETHSFEYTRDVLNTLSGALERELDRLQG
crtE_ZP875               FQIRDDYMNLQSSEYAHNKNFAEDLTEGKFNTLQKKSTSPEILHHCVNYMRTETNSFEYTREVLNTLSGALERELNRLQK
Clustal Consensus        ******************************:**********************:*****::*************.***

                         330       340       350
                         ....|....|....|....|....|....|...
crtE_CBS7918             EFAEANSRMDLGDVDSEGRTGKNVKLEAILKKLADIPL
crtE_ZP938               EFAEFGSRMDIGDLDSEGKAGKNVKLETILKKLADIPL
crtE_ZP875               EFAEAGSRMDIGDEEAEEKAGKNVKLEAILKRLADIPL
Clustal Consensus        **** .****:** ::* ::*******:***:******
```

b.

```
                         10        20        30        40        50        60        70        80
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
idi_CBS7918              MSMPNIVPPAEVRTEGLSLEEYDEEQVRLMEERCILVNPDDVAYGEASKKTCHLMSNINAPKDLLHRAFSVFLFRPSDGA
idi_ZP938                MSMPNIVPPAEIRTDGLSLEEYDEEQVRLMEERCILVNPDDVAYGEASKKTCHLMSNINAPKDLLHRAFSVFLFRPSDGA
idi_ZP875                MSKPSVATPAEIRTDGLSLEEYDEEQVRLMEERCILVNPDDVAYGEASKKICHLMSNINAPKDLLHRAFSVFLFRPSDGA
Clustal Consensus        ** *.:..***:**:********************************** ****************************

                         90        100       110       120       130       140       150       160
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
idi_CBS7918              LLLQRRADEKITFPGMWTNTCCSHPLSIKGEVEEENQIGVRRAASRKLEHELGVPTSSTPPDSFTYLTRIHYLAPSDGLW
idi_ZP938                LLLQRRADEKITFPGMWTNTCCSHPLSIKGEVEEENQIGVRRAASRKLEHELGVPTSATPPDSFTYLTKIHYLAPSDGLW
idi_ZP875                LLLQRRADEKITFPGMWTNTCCSHPLSIKGEIEEENQMGVRRAASRKLEHELGVPTSSTPPDSFTYLTRIHYLAPSDGLW
Clustal Consensus        ********************************:*****:******************:**********:***********

                         170       180       190       200       210       220       230       240
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
idi_CBS7918              GEHEIDYILFSTTPTEHTGNPNEVSDTRYVTKPELQAMFEDESNSFTPWFKLIARDFLFGWWDQLLARRNEKGEVDAKSL
idi_ZP938                GEHEIDYILFSTTPTEHTGNPNEVSDTRYVTKPELQAMFEDESNSFTPWFKLIARDFLFGWWDQLLARRNEKGEVDAKSL
idi_ZP875                GEHEIDYILFSTTPTEHTGNPNEVSDTRYVTKSELQTMFEDESNSFTPWFKLIARDFLFGWWDELLARRNEKGEVDAKSL
Clustal Consensus        ********************************.***:**************************:****************

                         250
                         ....|....|.
idi_CBS7918              EDLSDNKVWKM
idi_ZP938                EDLSDDKVWKM
idi_ZP875                EYLSDDKVWKM
Clustal Consensus        * ***:*****
```

Figure III. 1. Alignment of the proteins from the biosynthetic route of astaxanthin, of the three *Phaffia* species.

c.

```
                          10        20        30        40        50        60        70        80
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtS_CBS7918     MFILVLLTGALGLAAFSWASIAFFSLYLAPRRSSLYNLQGPNHTNYFTGNFLDILSARTGEEHAKYREKYGSTLRFAGIA
crtS_ZP938       MSILVLLTGVLGLAAFSWAFIAFFSLYLAPRQSSLYNLQGPDHTNYFTGNFLDILSARTGEEHAKYREKYGSTLRFAGIA
crtS_ZP875       MSILILFTGALGLTAFFWASIAFFSLYLAPRRSSLYNLQGPDHTNFLTGNFLEILSARTGEEHAKYREKYGSTLRFAGIA
Clustal Consensus * **:*:**.***:** ** ***********:*********:***::*****:**************************

                          90       100       110       120       130       140       150       160
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtS_CBS7918     GAPVLNSTDPKVFNHVMKEAYDYPKPGMAARVLRIATGDGVVTAEGEAHKRHRRIMIPSLSAQAVKSMVPIFLEKGMELV
crtS_ZP938       GTPVLNSTDPKVFNHVMKEAYDYPKPGMAARVLRIATGDGVVTAEGEAHKRHRRIMIPSLSAQAVKSMVPIFLEKGMELV
crtS_ZP875       GMPVLNSTDPKVFNHVMKEAYDYPKPGMAARVLRIATGDGVVTAEGEAHKRHRRIMIPSLSAQAVKSMVPIFLEKGMELV
Clustal Consensus * ********************************************************************************

                          170       180       190       200       210       220       230       240
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtS_CBS7918     DKMMEDAAEKDMAVGESAGEKKATRLETEGVDVKDWVGRATLDVMALAGFDYKSDSLQNKTNELYVAFVGLTDGFAPTLD
crtS_ZP938       DKMMEDAAEKDMAVGDSAGEKKATRLETEGVDVKDWVGRATLDVMALAGFDYKSDSLQNKTNELYVAFVGLTDGFAPTLD
crtS_ZP875       DKMMEDAAEKDMTVGESIGEKKVTRLDVEGVDVKDWVGRATLDVMALAGFDYKSDSLQNKFNELYVAFVGLTDGFAPTFD
Clustal Consensus ************:**:* ****.***:.********************************* *****************:*

                          250       260       270       280       290       300       310       320
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtS_CBS7918     SFKAIMWDFVPYFRTMKRRHEIPLTQGLAVSRRVGIELMEQKKQAVLGSASDQAVDKKDVQGRDILSLLVRANIAANLPE
crtS_ZP938       SFMAIMWDFIPYFRTMKRKHEIPLTQGLAVSRRVGIELMEQKKQAVLDSSSDQAVDKKDVQGRDILSLLVRANIAANLPE
crtS_ZP875       SAMAIMWDFIPFFRTMKRRREIPLTQGLAVSRRVGIELMEQKKQSVLGSSSDQAVDKKDVEGRDILSLLVRANIAANLPE
Clustal Consensus *  ******:*:******:.:************************:**.*:**********:******************

                          330       340       350       360       370       380       390       400
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtS_CBS7918     SQKLSDEEVLAQISNLLFAGYETSSTVLTWMFHRLSEDKAVQDKLREEICQIDTDMPTLDELNALPYLEAFVKESLRLDP
crtS_ZP938       SQKLSDEEVLAQISNLLFAGYETSSTVLTWMFHRLSEDKAVQDRLREEICQINTDTPTLDELNALPYLEAFVKESLRLDP
crtS_ZP875       SQKLSDEEVLAQISNLLFAGYETSSTVLTWMFHRLSEDKTVQDKLREEICQIDTDMPTLDELNALPYLEAFVKESLRLDP
Clustal Consensus ***************************************:***:********:** ************************

                          410       420       430       440       450       460       470       480
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtS_CBS7918     PSPYANRECLKDEDFIPLAEPVIGRDGSVINEVRITKGTMVMLPLFNINRSKFIYGEDAEEFRPERWLEDVTDSLNSIEA
crtS_ZP938       PSPYANRECLKDEDFIPLAEPVIGRDGSVINEVRITKGTMVMLPLYNINRSKFIYGEDAEEFRPERWLDDATDSLNSIEA
crtS_ZP875       PSPYANRECLKDEDFIPLAEPVVGRDGSIINEVRVTKGTMVMLPLYNINRSKYIYGEDAEEFRPERWLEDTTESFNSIEA
Clustal Consensus ***********************:*****:*****:**********:******:***************:*.*:*:*****

                          490       500       510       520       530       540       550
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|..
crtS_CBS7918     PYGHQASFISGPRACFGWRFAVAEMKAFLFVTLRRVQFEPIISHPEYEHITLIISRPRIVGREKEGYQMRLQVKPVE
crtS_ZP938       PYGHQATFISGPRACFGWRFAVAEMKAFLFVTLRRVQFEPIISHPEYEHITLIISRPRIVGREKEGYQMRLQVKPVE
crtS_ZP875       PYGHQATFISGPRACFGWRFAVAEMKAFLFVTLRRVEFEPIISRPEYEHITLIISRPRIVGREKEGYQMRLKVKPVG
Clustal Consensus ******:*******************************:******:***************************:****
```

Figure III. 1. (continued)

d.

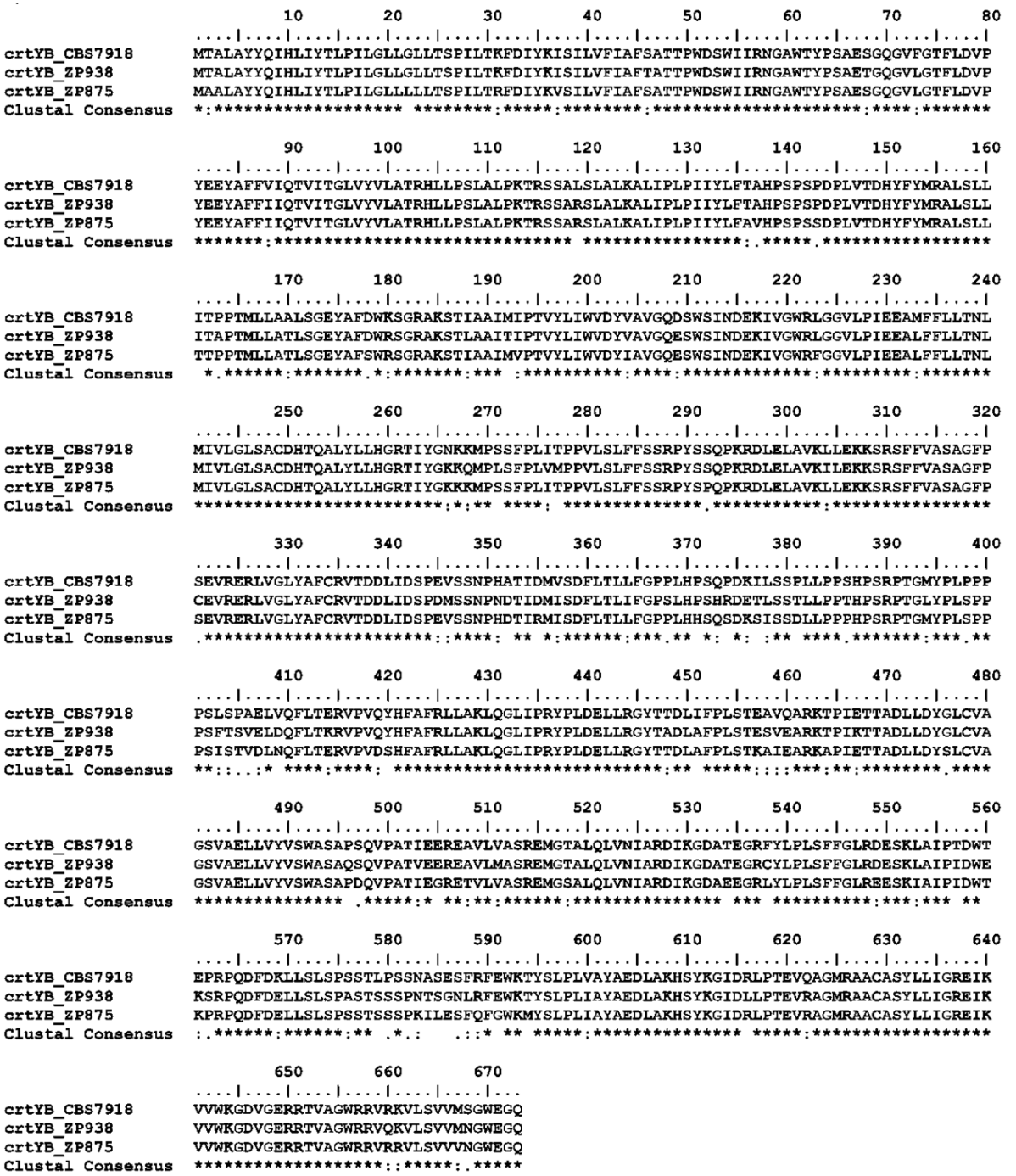


Figure III. 1. (continued)

e.

```
                          10        20        30        40        50        60        70        80
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtR_CBS7918              MATLSDLVILLLGALLALGFYNKDRLLGSSSSSASTTSGSSAATANGSKPTDSSGNGNAFKGDPRDFVARMKDQKKRLAV
crtR_ZP938                MATLSDLVILLLGALLALGFFYKDRLLGSSTSSTATASTSSAATTNGTKLAESNGNGHAFKGDPRDFVERMKDQKKRLAV
crtR_ZP875                MASLSDLVIILLGALLALGLYNKDRLLGS-KSSSATTTASTATTIDGTKSAESNGNGHIFKGDPRDFVARMIDQKKRLAV
Clustal Consensus         **:******:*********:: ******* .**::*:: *:*:* :*:* ::*.***: ********* ** ********

                          90        100       110       120       130       140       150       160
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtR_CBS7918              FYGSQTGTAEEYATRIAKEAKSRFGVSSLVCDIEEYDFEKLDQVPEDCAIVFCMATYGEGEPTDNAVQFIEMISQDDPEF
crtR_ZP938                FYGSQTGTAEEYATRIAKEAKSRFGVSSLVCDIEEYDFEKLDQVPEDCAIVFCMATYGEGEPTDNAIQFTEMINQDDPEF
crtR_ZP875                FYGSQTGTAEEYATRIAKEAKSRFGVSSLVCDIEEYDFEKIDQVPEDCAIVFCMATYGEGEPTDNAVQFIEMINQDDPEF
Clustal Consensus         ****************************************:************************:** ***.******

                          170       180       190       200       210       220       230       240
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtR_CBS7918              SEGSTLDGLKYVVFGLGNKTYEQYNVVGRQLDARLTALGATRVGERGEGDDDKSMEEDYLAWKDDMFAALATTLSFEEGA
crtR_ZP938                SNGSTLDGLKYVVFGLGNKTYEQYNVVGRQLDARLTALGATRIGERGEGDDDKSMEEDYLAWKDDMFAALAAKLSFEEGA
crtR_ZP875                SNGSTLDGLKYVVFGLGNKTYEQYNVIGRQLDSRLTSLGATRIGERGEGDDDKSMEEDYLAWKDDMFAALATTLSFEEGA
Clustal Consensus         *:************************:*****:***:*****:****************************:.*******

                          250       260       270       280       290       300       310       320
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtR_CBS7918              SGETPDFVVTEVPNHPIEKVFQGELSSRALLGSKGVHDAKNPYASPVLACRELFTGGDRNCIHLEFDITGSGITYQTGDH
crtR_ZP938                SGETPDFVVTEVPNHPIEKVFQGELSSRALFGSKGVHDAKNPYASPVLTCRELFTGGDRNCIHLEFDITGSGITYQTGDH
crtR_ZP875                SGETPDFVVTEVPSHPIEKVFQGELSSRALLGSKGVHDAKNPYASPVLACRELFTGGDRNCIHLEFDITGSGITYQTGDH
Clustal Consensus         *************.****************:*****************:******************************

                          330       340       350       360       370       380       390       400
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtR_CBS7918              VAVWPSNPDVEVERLLAVLGLTSPEKRRMIIQVVSLDPTLAKVPFPTPTTYDAVFRHYLDISAVASRQTLAVLAKYAPSE
crtR_ZP938                VAVWPSNPDVEVERLLAVLGLTLPEKRHMIIQVVSLDPTLAKVPFPTPTTYDAVFRHYLDVSAVASRQTLAVLAKYAPSE
crtR_ZP875                VAVWPSNPDVEVERLLAVLGLTSPEKRHMIIQVLSLDPTLAKVPFPTPTTYDAVFRHYLDVSAVASRQTLAILAKYAPSE
Clustal Consensus         ********************** ****:*****:**************************:**********:********

                          410       420       430       440       450       460       470       480
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtR_CBS7918              QAAEFLTRLGTDKQAYHTEVVGGHLRLAEVLQLAAGNDITVMPTAENTTVWNIPFDHVVSDVSRLQPRFYSISSSPKLHP
crtR_ZP938                EAAQFLTRLGTDKQAYHTEVVGGHLRLAEVLQLAAGNDIAVMPTAENTTVWNIPFDHVVSDISRLQPRFYSISSSPKLHP
crtR_ZP875                QAAEFLTRLGTDKQAYHTEVIGGHLRLAEVLQLAAGNDITVMPTAENTTMWNIPFDHIVSDVSRLQPRFYSISSSPKLYP
Clustal Consensus         :**:****************:*****************:*********:*******:***:****************:*

                          490       500       510       520       530       540       550       560
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtR_CBS7918              NSIHVTAVILKYESQATDRHPARWVFGLGTNYLLNVKQAANNETTPMISDGQDDVPEHVSAPKYTLEGPRGSYKHDDQLF
crtR_ZP938                NSIHVTAVILKYSSQATDGHPARWVFGLGTNYLLNVKQAANNETTPMISNGQDDVPEHVSSPKYTLEGPRGSYKHNDQLF
crtR_ZP875                NSIHVTAVILKYSSQASDRHPARWVFGLGTNYLLNVKQAANEETTPMISDGQDDIPEHVSAPKYTLEGPRGSYKHDDKLF
Clustal Consensus         ************.***:* **********************:*******:****:*****:**************:*:**

                          570       580       590       600       610       620       630       640
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtR_CBS7918              KVPIHVRRSTFRLPTSPKIPVIMIGPGTGVAPFRGFIQERIALARRSIAKNGPDALADWAPIYLFYGSRDEQDFLYAEEW
crtR_ZP938                KVPIHVRRSTFRLPTSPKIPVIMIGPGTGVAPFRGFIQERIALARRSIAKNGPDALADWAPIYLFYGSRDEQDFLYADEW
crtR_ZP875                KVPIHVRRSTFRLPTSPKIPVIMIGPGTGVAPFRGFIQERIALARRSIAKNGPDALADWAPIYLFYGSRDEQDFLYADEW
Clustal Consensus         ********************************************************************************:**

                          650       660       670       680       690       700       710       720
                          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
crtR_CBS7918              PAYEAELQGKFKIHVAFSRSGPRKPDGSKIYVQDLLWDQKEVIKSAIVEKRASVYICGDGRNMSKDVEQKLAAMLAESKN
crtR_ZP938                PAYEAELQGKFKIHVAFSRSGPRKPDGSKIYVQDLLWDQKEAIKSAIVEKRASVYICGDGRNMSKDVEQKLAAMLAESKN
crtR_ZP875                PAYEAELQGKFKIHVAFSRSGPRKPDGSKIYVQDLLWEQKEAIKSAIVEKRASVYICGDGRNMSKDVEQKLAAMLAESKN
Clustal Consensus         ************************************:***.****************************************

                          730       740
                          ....|....|....|....|.
crtR_CBS7918              GSAAVEGAAEVKSLKERSRLLMDVWS
crtR_ZP938                GSAAVEGAAELKSLKERSRLLMDVWS
crtR_ZP875                GSAAVEGAAEVKSLKERSRLLMDVWS
Clustal Consensus         **********:***************
```

Figure III. 1. (continued)

f.

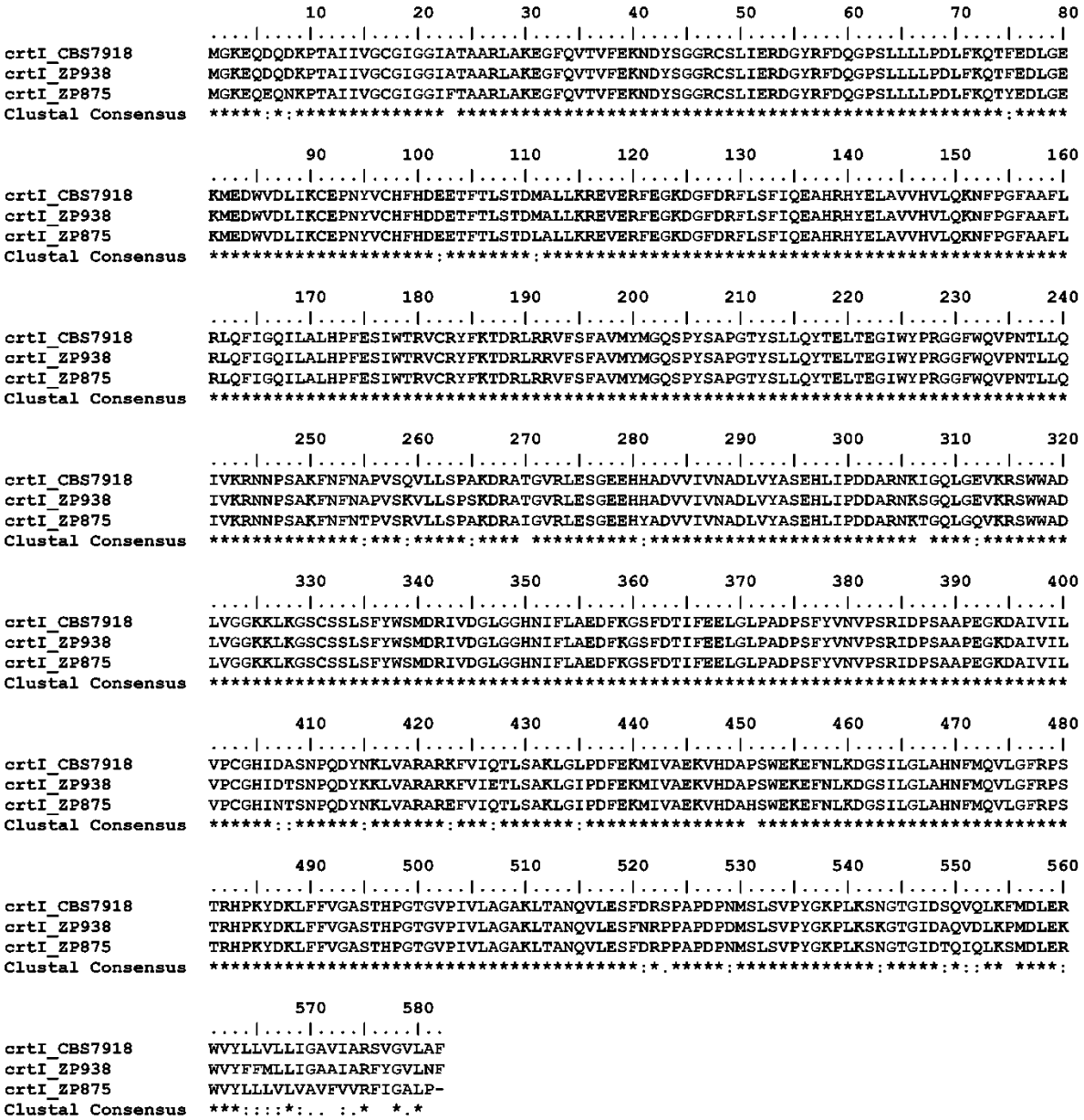


Figure III. 1. Alignment of the proteins from the biosynthetic route of astaxanthin, of the three *Phaffia* species.

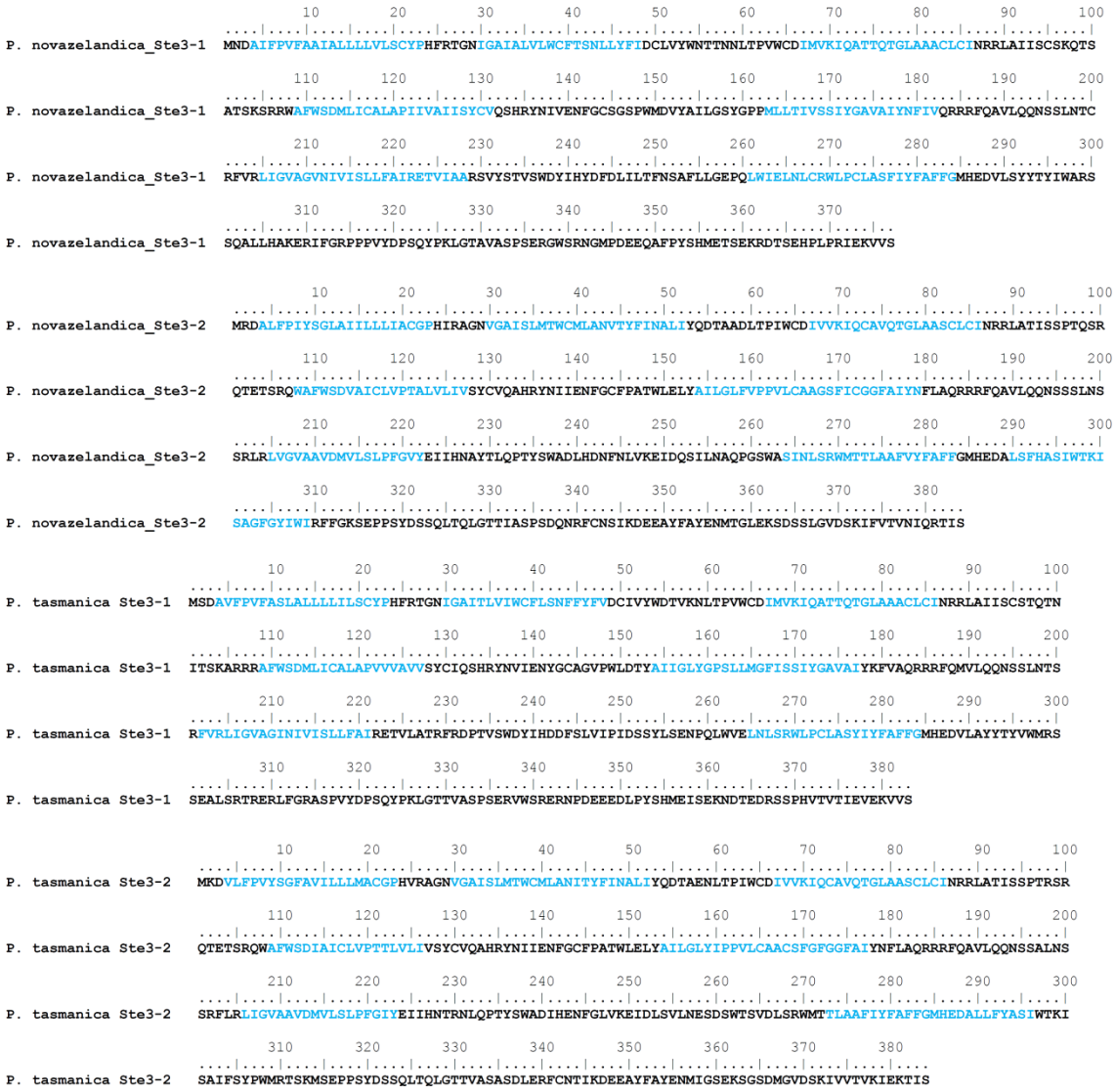


Figure III.2. Predicted proteins of the pheromone receptor genes of *P. novazelandica* and *P. tasmanica*. Highlighted in blue are the transmembrane domains of each sequence.

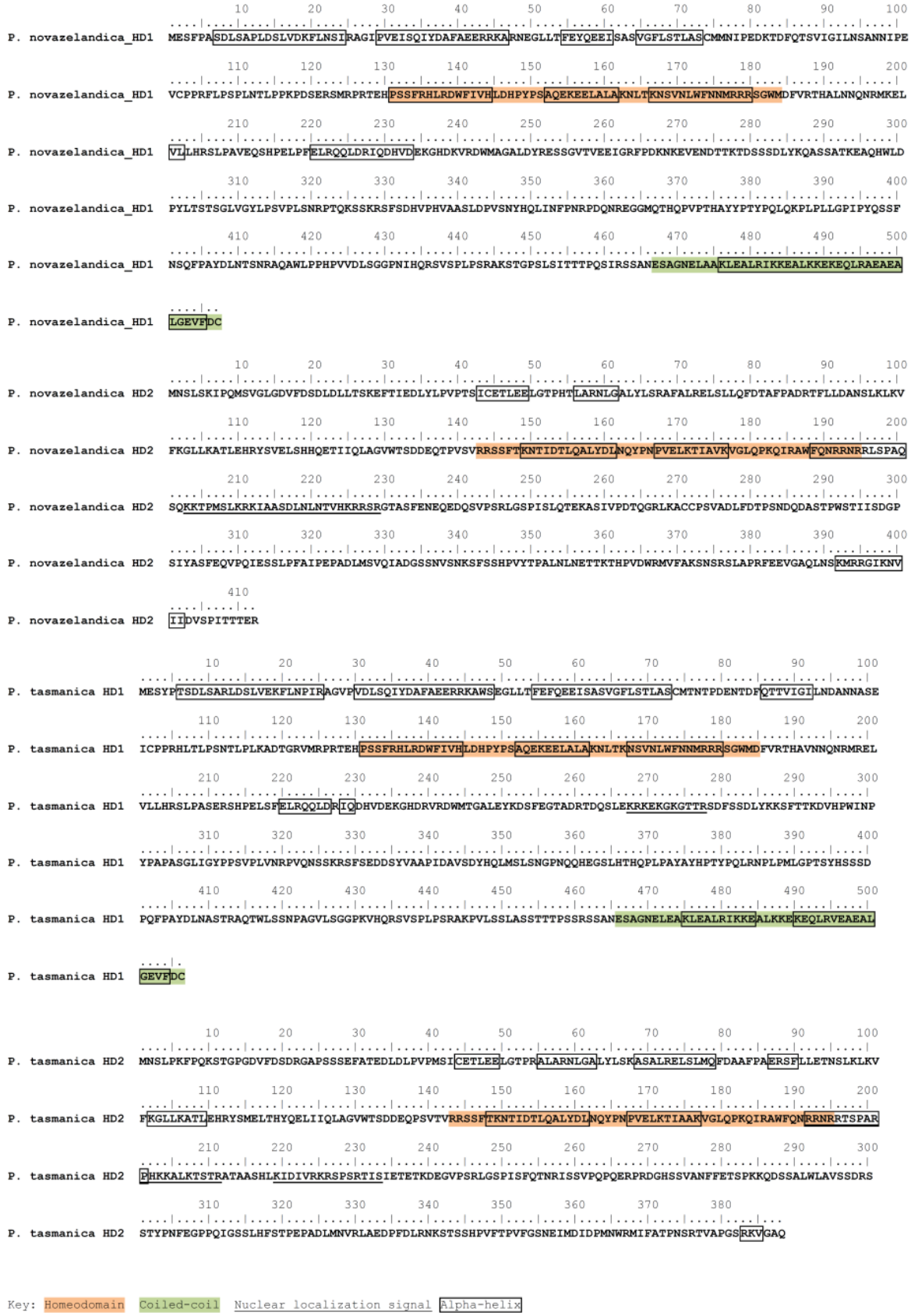


Figure III.3. Predicted proteins of the homeodomain transcription factor genes of *P. novazelandica* and *P. tasmanica*. Highlighted in the sequences are secondary features as decoded in the key.

APPENDIX IV

Additional information pertaining to Chapter 5

IV.1 Quality statistics for the draft genomes obtained for Chapter 5

Table IV.1. Statistics for the genomes assembled for this work (Chapter 5). All statistics are based on contigs of size ≥ 500 bp.

Species (strain)	Total length (Mb)	No. Scaffolds	Largest scaffold (Kb)	GC%	N50	Average Coverage (x)
C. ferigula (PYCC 5628)	20.9	8353	52.2	63.7	3.969	24
C. macerans (CBS 2425)	19.5	2496	175.5	61.3	19.764	17
C. bisporidii (CBS 6346)	20.0	4300	46.7	63.7	9.057	22

IV.2 Additional information pertaining to the *MAT* genes found in Cystofilobasidiales species

- Sequences of all pheromone receptor predicted proteins of the studied Cystofilobasidiales.
- Ste3 alignment used to infer the phylogenetic relationships depicted on Figure 5.1.
- Sequences of all Hd1 and Hd2 predicted proteins of the studied Cystofilobasidiales.
- Sequences of all *MFA* genes found in the studied Cystofilobasidiales.

Available at: https://figshare.com/s/d08c25d28e5369019921

IV.3. Taxonomic re-evaluation of strain Nwmf-AP1

Strain Nwmf-AP1 is identified in GenBank as belonging to species *M. frigida* however initial comparisons of the *MAT* genes found among all the *Mrakia* strains suggested it to be more closely related with *M. blollopis*. Consequently, sequences for some of the genes recently used in taxonomic clarifications within the Tremellomycetes class (Liu *et al.* 2015b), specifically the *ITS* region, Rpb1 and Rpb2 proteins were retrieved from the Nwmf-AP1 draft genome and compared with other *Mrakia* sequences available (Figure V.1). The results obtained indicate that strain is Nwmf-AP1 most likely a *M. blollopis* strain.

- *ITS* alignment used to infer the phylogenetic relationships depicted on Figure V.1.
- Rpb1 alignment used to infer the phylogenetic relationships depicted on Figure V.1.
- Rpb2 alignment used to infer the phylogenetic relationships depicted on Figure V.1.

Available at: https://figshare.com/s/d08c25d28e5369019921

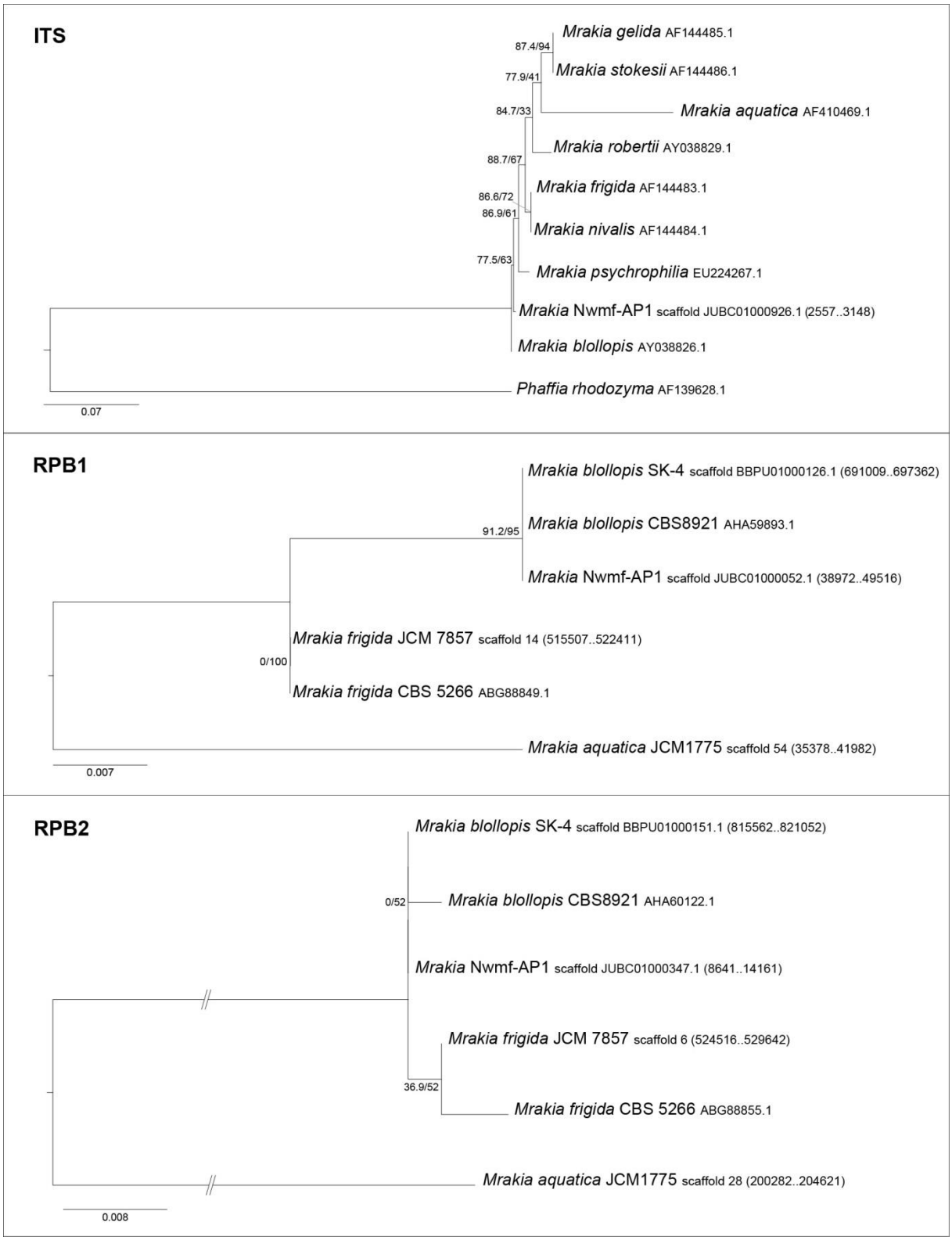


Figure IV.1. Phylogenetic inferences illustrating the likelihood of strain Nwmf-AP1 belonging to *Mrakia blollopis* species instead of *Mrakia frigida*. Phylogenies of *ITS* region, Rpb1 and Rpb2 proteins. Phylogenetic trees were inferred with IQ-TREE software (Nguyen *et al.* 2015) using ultrafast bootstrap (Minh *et al.* 2013) and SH-like aLRT branch test (Guindon *et al.* 2010) with 10000 replicates each. Best-fit models were chosen according to AICc; TIM2+G4, LG and VT+I for *ITS*, Rpb1 and Rpb2, respectively. Final datasets were composed by 640 nucleotides, 186 amino acids and 375 amino acids. Numbers in branches correspond to: SH-aLRT support (%) / ultrafast bootstrap support (%). Scales are in the number of nucleotide/amino-acid substitutions per site. Accession numbers or scaffold number and respective coordinates are next to each strains name.

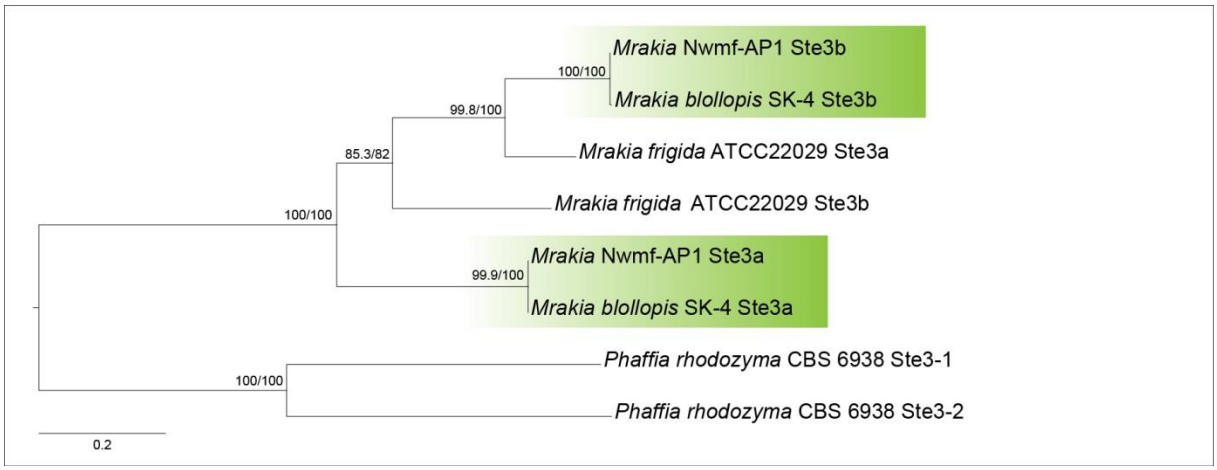


Figure IV.2 Pheromone receptor phylogeny. Phylogenetic tree inferred with IQ-TREE software (Nguyen *et al.* 2015) using ultrafast bootstrap (Minh *et al.* 2013) and SH-like aLRT branch test (Guindon *et al.* 2010) with 10000 replicates each. Best-fit model was chosen according to AICc (LG+F+G4). Final dataset composed by 390 amino acids. Numbers in branches correspond to: SH-aLRT support (%) / ultrafast bootstrap support (%). Scale is in the number of amino-acid substitutions per site. Ste3 alleles of strains SK-4 and Nwmf-AP1 are highlighted in green boxes.

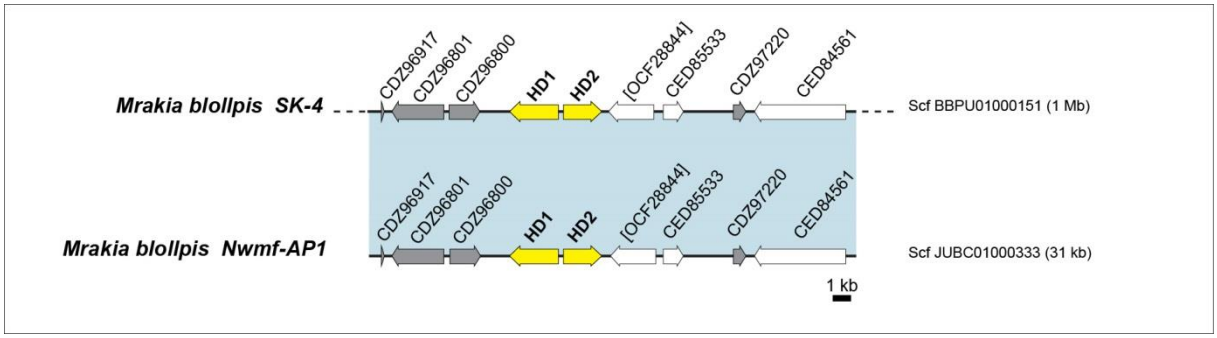


Figure IV.3 *HD* regions of *M. blollopis* strains. Scaffolds regions encoding *HD* genes in each of the draft genomes available for *M. blollopis* strains are presented. Genes are depicted as arrows indicating the direction of transcription and are identified by the protein accession number of their putative orthologs in *P. rhodozyma* (CBS 6938) whenever possible. On the right side of each scaffold, its number and approximate size is indicated. Scaffold ending with a dotted line indicates that it is only partially represented. Homeodomain transcription factor genes are colored in yellow. Genes depicted in gray, are genes present in the *P. rhodozyma* (CBS 6938) *HD* scaffold but not present in the depicted region (see Figure 5.7). Genes shown in white are genes that do not belong to the *HD* scaffold of *P. rhodozyma* (CBS 6938). Synteny is indicated between the scaffolds by a block in light blue.