OPEN

Differential patterns of postmigration HIV-1 infection acquisition among Portuguese immigrants of different geographical origins

Victor Figueiredo Pimentel[a], Marta Pingarilho[a], Giordano Sole[a], Daniela Alves[a], Mafalda Miranda[a], Isabel Diogo[b], Sandra Fernandes[b], Andrea Pineda-Pena[a], M. Rosário O. Martins[a], Ricardo Camacho[c], Perpétua Gomes[b,d], Ana B. Abecasis[a], on behalf of the Portuguese HIV-1 Resistance Study Group

Objective: To investigate the dynamics of phylogenetic transmission clusters involving immigrants of Portuguese Speaking Countries living in Portugal.

Design/methods: We included genomic sequences, sociodemographic and clinical data from 772 HIV migrants followed in Portugal between 2001 and 2017. To reconstruct HIV-1 transmission clusters, we applied phylogenetic inference from 16 454 patients: 772 migrants, 2973 Portuguese and 12 709 global controls linked to demographic and clinical data. Transmission clusters were defined using: clusters with SH greater than 90% (phylogenetic support), genetic distance less than 3.5% and clusters that included greater than 66% of patients from one specific geographic origin compared with the total of sequences within the cluster. Logistic regression was performed to assess factors associated with clustering.

Results: Three hundred and six (39.6%) of migrants were included in transmission clusters. This proportion differed substantially by region of origin [Brazil 54% vs. Portuguese Speaking African Countries (PALOPs) 36%, $P < 0.0001$] and HIV-1 infecting subtype (B 52%, 43% subtype G and 32% CRF02_AG, $P < 0.001$). Belonging to a transmission cluster was independently associated with treatment-naive patients, $CD4^+$ greater than 500, with recent calendar years of sampling, origin from PALOPs and with seroconversion. Among Brazilian migrants – mainly infected with subtype B – 40.6% were infected by Portuguese. Among migrants from PALOPs – mainly infected with subtypes G and CFR02_AG – the transmission occurred predominantly within the migrants' community (53 and 80%, respectively).

Conclusion: The acquisition of infection among immigrants living in Portugal differs according to the country of origin. These results can contribute to monitor the HIV epidemic and prevent new HIV infections among migrants.

AIDS 2022, **36**:997–1005

Keywords: HIV-1, migrants, molecular epidemiology, transmission clusters

[a]Global Health and Tropical Medicine (GHTM), Instituto de Higiene e Medicina Tropical/Universidade Nova de Lisboa (IHMT/UNL), [b]Laboratório de Biologia Molecular (LMCBM, SPC, CHLO-HEM), Lisboa, Portugal, [c]KU Leuven, Clinical and Epidemiological Virology, Department of Microbiology and Immunology, Rega Institute for Medical Research, Belgium, and [d]Centro de Investigação Interdisciplinar Egas Moniz (CiiEM), Instituto Superior de Ciências da Saúde Egas Moniz, Caparica, Portugal.

Correspondence to Victor Figueiredo Pimentel, Global Health and Tropical Medicine (GHTM), Institute of Hygiene and Tropical medicine (IHMT) – New University of Lisbon, Lisbon 1349-008, Portugal.

Tel: +351 213 652 600; e-mail: victor.pimentel@ihmt.unl.pt

Received: 2 September 2021; revised: 8 February 2022; accepted: 14 February 2022.

DOI:10.1097/QAD.0000000000003203

Introduction

Migrants from low-income and mid-income countries, including those from sub-Saharan Africa are highly affected by HIV/AIDS in the European Union/European Economic Area (EU/EEA) and a priority group for HIV prevention. In 2017, 41% of new HIV diagnoses in EU/EEA Member states were among migrants [1,2].

Portugal has an important HIV burden, with 559 913 people diagnosed and 973 new diagnoses in 2019 [3]. Migrants continue to bear an important burden of HIV infection in Portugal and account for 32% of HIV infections diagnosed among adults and adolescents [3]. In the past decade, there has been an increase in the circulation of non-B strains and circulating recombinant forms (CRFs) of HIV-1 in Europe and North America [4,5]. In Portugal and other European countries, HIV-1 subtype B is still predominant but the proportion of non-B-infected individuals has progressively increased [6,7].

Advances in molecular epidemiologic methods allowed to perform HIV transmission cluster inference through the identification of HIV-infected individuals with genetically related viruses [8–10]. As such, research that combines behavioral, clinical and molecular epidemiology to assess its relationship with HIV transmission cluster topology could provide empirical evidence for HIV transmission dynamics and rapidly growing clusters [11], undisclosed transmission risk factors, international transmission routes and antiretroviral treatment (ART) resistance transmission dynamics [12].

Knowledge concerning HIV-1 transmission cluster dynamics involving migrants is scarce. The present effort aims to study transmission cluster topology and correlate it with a risk factor, subtypes, therapy exposure and country of origin to better understand the HIV-1 transmission dynamics and risk factors in this vulnerable population.

In this study, we used HIV-1 pol sequences generated over a period of almost 20 years from individuals followed up in Egas Moniz Hospital in Portugal-Lisbon together with a large set of background control sequences in migrants from Portuguese Speaking countries [Brazil and Portuguese Speaking African Countries (PALOPs)] followed up in Portugal.

Our findings suggest the usage of this type of approach to inform public health prevention and surveillance strategies.

Methods

Study participants and design

The protocol was in accordance with the declaration of Helsinki and approved by the Ethical Committee of Centro Hospitalar de Lisboa Ocidental (CHLO) (108/CES-2014; 13 October 2014). The study population consisted of HIV-1-infected individuals ($n = 3745$) followed in Portugal, of those 2973 were from Portuguese and 772 were migrants (Brazil and PALOPs). All genomic sequences were generated as part of routine clinical resistance testing at the participating sites using standard (Sanger) sequencing procedures. Individuals who had at least one sequence available were included in this study. Duplicate sequences from the same patient were removed, keeping only the first sequence sampled from each patient. Clinical data, sex, age, together with HIV-1 viral load and $CD4^+$ cell count, were collected at diagnosis or after treatment initiation. Recent infections were defined based on the ambiguity rate of the genomic sequence. We assumed chronic infection as an ambiguity value higher than 0.45% and seroconverters' ambiguity value equal or below 0.45%.

HIV-1 sequences and transmission cluster identification

All samples included in this study were previously subtyped by Rega v3.0, as described in [13]. Samples that showed any signs of homologous recombination were removed from the study. After quality control, we built three separate datasets for the most prevalent subtypes: subtypes B, G and CRF02_AG. Using BLAST, we then selected 12 709 global control sequences from the Los Alamos database (http://hiv-web.lanl.gov) and 2973 from the Portuguese REGA HIV-1 database (PTHIVDB). For each subtype, sequences were aligned against the global background dataset selected as controls using Virulign [14] to verify the reliability of the alignments. The HIV type 1 K03455.1 (HXB2) pol nucleotide sequence (nt) positions (2251–3251) were included in the alignment to serve as a reference and edited manually using MEGA7 software, with a final sequence alignment length of 1000 bases. The codons associated with drug resistance mutations for surveillance were stripped from the sequences. Maximum-likelihood phylogenies were constructed in FastTree with the generalized time-reversible model. Statistical support of clades was assessed using the Shimodaira–Hasegawa-like test (SH-test). Putative transmission clusters were identified using ClusterPicker v1.332 [15] and defined as clades with (1) high branch support ($\geq$0.90 SH-test), (2) maximum pairwise genetic distance less than 3.5% between all sequences present in the cluster. For analysis of cluster size, we defined clusters with eight patients or more as large and clusters with less than eight patients as small.

Origin of infection based on transmission clusters

To investigate the postmigration transmission patterns across individuals with different nationalities, we categorized origin of the migrants' infection based on the following categories below:

(1) Infected by immigrants if the cluster contains at least 66% of the sequences composed by immigrants;
(2) Infected by Portuguese if the cluster contains at least 66% of the sequences composed by Portuguese;
(3) Unknown if half of the sequences are composed of immigrants and half by control sequences;
(4) Infected outside Portugal if the cluster contains at least 66% of the sequences composed of global sequences sampled outside Portugal.

Antiretroviral drug resistance analysis

Partial *pol* sequences from drug-naive patients, including PR/RT regions, were screened for drug-resistant mutations according to WHO guidelines using the Calibrated Population Resistance Tool (http://hivdb.stanford.edu/) according to the latest update 2009 [16]. In order to access Acquired Drug Resistance (ADR) in treated patients, the Genotypic Resistance Interpretation Algorithm of the HIVdb program (http://sierra2.stanford.edu/sierra/servlet/JSierra) was used.

Statistical analysis

Statistical analysis was performed using R version 3.4.1 (R Foundation, Vienna, Austria). Frequency tables were made for qualitative variables. In order to choose parametric or nonparametric tests, a normality test of Kolmogorov–Smirnov was used. Median and interquartile range (IQR) were used for asymmetric quantitative or ordinal variables, that were compared using Mann–Whitney *U* or Kruskal–Wallis test. Differences between proportions were performed with the chi-square test or Fisher's exact test, as appropriate. We used univariate and multivariate analyses to determine the independence of associations between exposed variables (Sample date, age, first $CD4^+$ cell count, subtypes, treatment status, chronic status and country of birth (Brazilian vs. PALOPs) and outcome (clustering).

Results

Epidemiological and clinical data

Of the 772 migrants' sequences analyzed, 633 (82%) belonged to non-B subtypes. The most prevalent HIV-1 variant was CRF02_AG ($n = 370$, 47.9%) followed by subtype G ($n = 263$, 34%) and subtype B ($n = 139$, 18%). Of those patients for whom epidemiological information was available, 51.03% (394/772) were male individuals vs. 48.18% (372/772) female individuals; 78.75% (608/772) were migrants from PALOPs (determined by place of birth) vs. 21.25% (164/772) migrants from Brazil. Five hundred and thirty-seven (69.56%) were drug-naive and 216 (27.98%) treated. The median age was 39 years (IQR 32–48). The median value for the first $CD4^+$ cell count was 268 cells/μl (IQR 124–409 cells/μl). The median value for the first viral load was 4.6 log copies/ml (IQR 3.95–5.27 log copies/ml) (Table 1).

Table 1. Summary of sociodemographic and clinical characteristics of migrants in Portugal, 2001–2017.

Characteristics	Overall ($n = 772$)	Brazil ($n = 164$)	PALOPs ($n = 608$)	*P* value
Age				
Median ± IQR	39 (32–48)	34 (28–41)	41 (33–50)	<0.0001
Gender (number, %)				
Male	394 (51.03)	117 (71.34)	277 (45.56)	<0.0001
Female	372 (48.18)	46 (28.05)	326 (53.62)	
Unknown	6 (0.78)	1 (0.6)	5 (0.82)	
Treatment status (number, %)				
Naive	537 (69.56)	131 (79.88)	406 (66.78)	<0.0001
Treated	216 (27.98)	30 (18.29)	186 (30.59)	
Missing data	19 (2.46)	3 (1.83)	16 (2.63)	
Risk group (number, %)				
MSM	46 (5.96)	30 (18.29)	16 (2.63)	<0.0001
HET	35 (4.53)	5 (3.05)	30 (4.93)	
Others	33 (4.27)	11 (6.71)	22 (3.62)	
Missing data	658 (85.23)	118 (71.95)	540 (88.81)	
Subtype (number, %)				
B	139 (18)	139 (84.76)	0	<0.0001
G	263 (34.07)	20 (12.19)	243 (39.97)	
CRF02_AG	370 (47.93)	5 (3.05)	365 (60.03)	
HIV-1 RNA level (log_{10} copies/ml)				
median ± IQR	4.6 (3.95–5.27)	4.7 (4.2–5.2)	4.5 (3.9–5.2)	0.109
First $CD4^+$ cell counts (cells/μl)				
median ± IQR	268 (124–409)	406 (262–561)	247 (119–363)	<0.0001

Quantitative variable values are expressed as median and IQR (25–75 percentiles) and quantitative variables are displayed as number of cases (percentage and 95% confidence Interval in parentheses). HIV-1 subtypes were determined based on the partial pol sequence. $CD4^+$, $CD4^+$ T-cell count; PALOPS, Portuguese speaking African countries; VL, viral load in log_{10}.

When patients were stratified according to origin country (Brazil vs. PALOPs), we found that Brazilian migrants were 7 years younger than migrants from PALOPs ($P < 0.001$) and were more likely to be male individuals 71.34% compared with PALOPs (45.56%) ($P < 0.001$). The median value of the first $CD4^+$ cell count was significantly higher for Brazilians than for PALOPs migrants ($P < 0.001$). The most prevalent subtype among migrants from Brazil was subtype B (84.76%), whereas for PALOPs, the most prevalent subtypes were CRF02_AG and subtype G (60.03 and 39.97%), respectively. Brazilian migrants were more likely to be MSM than PALOP (18.29 and 2.63%) ($P < 0.001$) (Table 1).

Inference of transmission clusters

On the basis of the PR + RT phylogenetic analysis, transmission clusters were defined as clades with a branch support value greater than 90% and a maximum genetic distance less than 3.5% for subtypes B, G and CRF02_AG.

Table 2. Demographic and clinical characteristics of the study population according to whether or not they belong to a molecular transmission cluster.

Characteristics	Cluster ($N=306$)	Noncluster ($N=466$)	P value
Age			
median ± IQR	37 (30–46)	41 (33–49)	0.034
Gender (number, %)			
Male	43 (38–48)	57 (52–62)	0.101
Female	37 (32–42)	63 (58–68)	
Treatment status			
Naive	44 (39–48)	56 (52–60)	<0.001
Treated	29 (23–36)	71 (64–77)	
Country of origin (number, %)			
Brazil	54 (46–62)	46 (38–54)	<0.0001
PALOPs	36 (32–40)	64 (60–68)	
Risk group (number, %)			
MSM	48 (33–63)	52 (37–67)	0.3899
HET	34 (20–52)	66 (48–80)	
Others	48 (31–66)	52 (34–69)	
Overall_TDR	37 (23–53)	63 (47–77)	0.4381
NRTI	1.7 (0.67–4.3)	3.6 (2.07–6.47)	0.1844
NNRTI	3.8 (2.04–7.15)	5.6 (3.58–8.92)	0.3351
PI	1.7 (0.67–4.3)	0.6 (0.18–2.41)	0.2233
Subtype (number, %)			
B ($n=139$)	52 (43–60)	48 (40–57)	<0.001
G	43 (37–50)	57 (50–63)	
CRF02_AG	32 (28–38)	68 (62–72)	
HIV-1 RNA level (log_{10} copies/ml)			
Median ± IQR	4.56 (3.94–5.13)	4.69 (4.0–5.33)	0.264
$CD4^+$ cell counts (cells/μl)			
Median ± IQR	348 (220–511)	227 (115–367)	<0.0001

HET, heterosexual; NNRTI, nonnucleotide reverse transcriptase inhibitors; NRTI, nucleotide reverse transcriptase inhibitors; PALOPS, Portuguese speaking African countries; PI, protease inhibitors.

We identified 180 transmission clusters containing 306 of the 772 patients (39.6%). The average cluster size was three, with a minimum of two (93 clusters), and a maximum of 33 (one cluster). Characteristics of the patients who were in or outside of clusters are presented in Table 2.

Comparison between patients in clusters (i.e. connected) vs. patients not in clusters (i.e. singletons) revealed that individuals who are linked in a cluster were significantly younger compared with those not in clusters (median age = 37 vs. 41 years, $P=0.034$).

Baseline $CD4^+$ T-cell count was higher in patients in clusters than in those not in the cluster (348 vs. 227) ($P<0.01$) (Table 2). Individuals infected with subtype B were also more likely to be in clusters than those infected with other subtypes G and CRF02_AG (52 vs. 43 vs. 32%; $P<0.01$). Migrants from Brazil were more frequently in clusters compared with migrants from PALOPs ($P<0.0001$).

Treated patients were less likely to be part of transmissions clusters compared with naive patients ($P<0.01$) (Table 2). Prevalence of TDRMs was similar within and outside of identified transmission clusters [6.4% (15/234) vs. 8.5% (26/299); $P=0.438$]. This TDR included four cases with resistance to protease inhibitors, four to NRTIs and nine to NNRTIs, with the predominant gender being male (60%) (Table 2).

Table 3. Characteristics of HIV-1 migrants in transmission clusters, stratified by cluster size, Portugal, 2001–2017 ($n=306$).

Characteristic	In large clusters ($n=32$)		In small clusters ($n=274$)	
	N	%	N	%
Infected within Portuguese clusters'	31	97[a]	7	2.5
Infected within clusters of migrants	1	3[a]	195	71%
Recent infections	19	59[a]	13	4.7
Drug resistance mutations (HIVdb)	1	3.1	15	5.4
Treatment status (naive)	30	94[a]	210	77
$CD4^+$ cell counts >500 (cells/μl)	11	34[a]	38	13.8

[a]Chi-square test of independence variables: P less than 0.05.

On the basis of our definitions of origin of infection, proportion of sequences in large clusters (defined as >8 tips) consisted more of naive, had migrants more likely to be in clusters with Portuguese sequences, contained a significantly higher proportion of recently infected individuals and consisted of significantly fewer patients with a high level of $TCD4^+$(>500) compared with small clusters (Table 3).

Patients in the large clusters were on average younger than patients in the smaller clusters (28 vs. 38 years old) and this difference was statistically significant ($P<0.001$). However, drug resistance mutations were identified more frequently in the small than in the large clusters, yet this was not statistically significant (Table 3).

We sought to determine, which demographic/risk characteristics were associated with clustering in order to identify subpopulations with higher rates of transmission. The results from univariable and multivariable logistic regression models for the association between the probabilities of belonging to a molecular transmission cluster were performed using a set of 772 patients for whom there was information for all the variables considered: country of origin, clustering status, sampling year, treatment status, age, $CD4^+$ cell count and chronic status. After adjustment, PALOPs had a lower odds of being in transmission clusters [OR = 0.60; 95% CI (0.36–0.98), $P=0.043$]. Patients with $CD4^+$ cell counts higher than 500 cells/μl had a higher odds of being in transmission clusters, compared with lower $CD4^+$ cell count. Patients with a recent calendar year of sampling [OR = 1.70; 95% CI (1.04–2.79, $P=0.033$] had a higher probability to be in clusters than patients with an older year of sampling. Seroconverter patients were more likely to cluster than chronic (OR = 1.59; 95% CI, 0.99–2.55; $P=0.053$) and lastly, being ART-

Table 4. Factors associated with the probability of belonging to a molecular transmission cluster: results from univariable and multivariable logistic regression models.

Raw			Adjusted		
Covariate	OR (95% CI)	*P* value	Covariate	OR (95% CI)	*P* value
Age at sample date			Age at sample date		
≥ 56[a]			≥ 56[a]		
18–24	1.18 (0.596–2.35)	0.629	18–24	–	
25–40	0.993 (0.61–1.60)	0.977	25–40	–	
41–55	0.469 (0.23–0.77)	**0.003**	41–55	–	
$CD4^+$ cell count (cells/μl)			$CD4^+$		
≤200[a]			≤200[a]		
201–500	2.58 (1.64–4.05)	**<0.001**	201–500	2.22 (1.38–3.55)	<0.001
≥500	5.61 (3.11–10.11)	**<0.001**	≥500	3.99 (2.14–7.45)	<0.001
Subtype			Subtype		
CRF_02AG[a]			CRF_02AG[a]		
B	2.23 (1.59–3.33)	**<0.0001**	B	–	
G	1.59 (1.15–2.21)	0.005	G	–	
Country		**<0.0001**	Country		
Brazil[a]			Brazil[a]		
PALOPs	0.46 (0.32–0.66)		PALOPs	0.60 (0.36–0.98	0.043
Treatment		**<0.0001**	Treatment		
Treated[a]			Treated[a]		
Naïve	1.82 (1.29–2.53)		Naive	1.53 (0.99–2.37)	0.052
Sampling year			Sampling year		
2001–2008[a]			2001–2008[a]		
2009–2017	1.90 (1.29–2.79)	**<0.001**	2009–2017	1.70 (1.04–2.79)	0.033
Chronic			Chronic		
Yes[a]			Yes[a]		
No	2.23 (1.62–3.07)	**<0.0001**	No	1.59 (0.99–2.55)	0.053

PALOPS, Portuguese speaking African countries.
[a]Reference classes.

naive at sampling were also associated with transmission clusters [OR = 1.53; 95% CI (0.99–2.37), $P = 0.052$; Table 4].

Among subtype B infections, 72 (52%) of the 139 Brazilian migrants were members of a transmission cluster. They were distributed over 54 clusters. The majority of these clusters (40.6%) predominantly included patients who originated from Portugal, indicating that these migrants may have been infected by Portuguese and most likely after migration to Portugal. On the other hand, 22% of these clusters included mainly patients originated from Brazil, indicating those patients most likely were infected by Brazilians in Brazil. For 26.4% of the clusters, there was the same proportion of sequences from Brazilian migrants and Portuguese. Finally, around 11% of migrants were infected by migrants (Fig. 1).

Among subtype G infections, 114 (43%) of the 263 migrants were in transmission clusters. Contrarily to subtype B, here 53% of sequences were in clusters belonging to migrant community based on the proportion of sequences from migrants in the clusters being higher than 66%, suggesting the transmission occurred intramigrants population. On the other hand, 19% of migrants clustered in clusters where the majority of patients were Portuguese suggesting a putative transmission from Portuguese to migrants. Unknown transmission corresponded to 19% of patients. Finally, 4% of migrants were clustering with sequences from abroad (infected somewhere outside of Portugal).

When we split migrants from PALOPs and Brazil, we observed quite different profiles, where 56% of sequences from PALOPs were infected by migrants, 20% infected by Portuguese, and 4% infected outside Portugal. On the other hand, 58% of migrants from Brazil harboring subtype G virus were infected by Portuguese, 17% infected by migrants and 8% infected outside Portugal.

Regarding CRF02_AG, our findings imply that transmissions occur at higher proportions (80%) between migrants, followed by 6.7% infected by Portuguese, 6.3% unknown transmission, and 5% infected outside Portugal. The same analysis was carried out splitting between PALOPs and Brazilians. For PALOPs: 80% infected by migrants, 11% unknown transmission, 5% infected outside Portugal, and 4% infected by Portuguese; For Brazilians: 60% infected by Portuguese, 40% represented cross transmissions between PALOPs and Brazilian migrants.

Discussion

Analysis of viral genetic sequences provides a route to uncovering transmission dynamics [17]. Particularly for

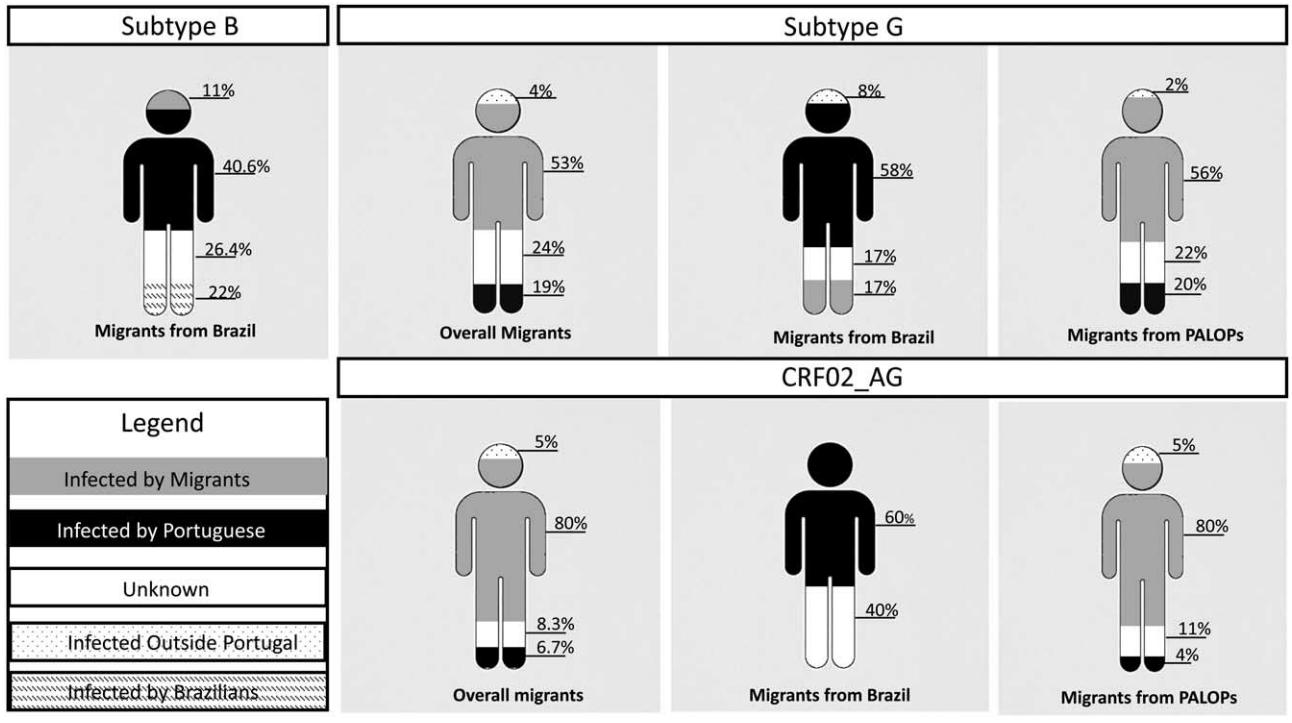

Fig. 1. Diagram of the reconstructed origin of HIV-1 transmission among immigrants followed up in Portugal, based on the phylogenetic clustering method. The bars displayed in different colors correspond to the putative origin of HIV infection for subtypes B (only Brazilian migrants), G, and CRF02_AG.

HIV, phylogenetic analysis in the context of its rapid rate of evolution allows identifying genetic clusters of densely connected subpopulations [18]. These clustering approaches can reveal transmission patterns otherwise hidden when using traditional epidemiological approaches (e.g. obscured transmission risk behaviors, like erroneously self-reported heterosexual males whose viruses cluster within clusters of MSM patients [19]).

Migrants mainly originated from sub-Saharan Africa, represent – after local diagnoses – the second largest group of HIV infections reported in the EU [1]. This study provides an analysis of HIV-1 transmission clusters circulating in Portugal, reconstructed from HIV-1 genomic sequences of 772 Portuguese-speaking migrants diagnosed between 2001 and 2017 enrolled in the Portuguese HIV-1 Resistance cohort. To obtain control sequences for transmission cluster analysis, we collected all the available HIV-1 genomes of non-B and CRF subtypes deposited in the HIV Los Alamos database. For subtype B, we used all sequences from Brazil available in LosAlamos as this subtype was only detected among Brazilian migrants. We also included 2973 sequences from Portugal that originated from the HIV-1 Drug Resistance database. Our study is the first to investigate the origin of HIV-1 infection acquisition in migrants in Portugal using this approach and among the first to investigate the structure of these regional HIV-1 phylogenies in greater detail, using a large-scale sequence dataset, dense reference sequence sampling and associating multiple clinical and demographic factors.

Around 40% of the overall individuals included in the analysis were involved in clusters. Although this observation is consistent with other reports of HIV-1 epidemic dispersal in Portugal [9,20], it indicates an important role of HIV-1-positive migrants in the dispersal of HIV-1 infection.

Transmission cluster analyses were carried out involving the most prevalent subtypes identified among immigrants living in Portugal subtypes non-B were more prevalent among immigrants from PALOPS, whereas patients from Brazil were mainly infected with subtype B [2,21]. Of note, the subtypes identified in the immigrants were largely consistent with the ones driving the epidemic in their countries of origin. However, in contrast to subtype B, CRF02_AG, was less likely to be in a cluster, these findings are in line with other studies in Europe, where subtype non-B seems to be associated with the heterosexual population, and therefore less frequently found in transmission clusters [22,23].

The number of same country migrant sequences that we observed was significantly greater than expected by chance, including among pairs with evidence of postmigration acquisition of HIV. Our results corroborate studies conducted in Europe where HIV transmission among

migrants from sub-Saharan Africa is very common within their community [24,25]. However, it is worth of comment that migrants from Brazil had a more important epidemiological link with the Portuguese community. This, perhaps, can be explained by potentially higher levels of inclusion within the Portuguese native community.

It is important to note that, herein, we do not assume the origin of infection based on the time and place of the most recent common ancestor (tMRCA) but based on the total number of sequences belonging to a specific region. This approach has some limitations, given the fact that sampling bias can artificially cause clustering of sequences from the same origin because of undersampling from other geographic origins. In fact, it is legitimate to pose the hypothesis of an artificial grouping of immigrants infected with subtype B with samples of Portuguese patients. To overcome such possible criticism, we included a higher proportion of control samples from Brazil than from Portugal: 5623 control samples from Brazil vs. 1657 samples from Portugal were included. On the other hand, our approach is the only possible approach to analyze such a large dataset, as Bayesian phylogeographic approaches are limited to smaller datasets, given its computational complexity.

Prior research based largely on self-report or $CD4^+$ cell count testing suggested that most migrants living with HIV were infected prior to migration [26]. Recent investigations from Europe utilizing more robust methods indicate that high proportions of migrants acquire HIV infection after migration [27,28]. Our data suggest that the highest number of HIV infections among the migrants' population in Portugal occurred after the migratory process. These data corroborate previous studies conducted in the migrant population from African, where the transmission clusters are mainly composed of immigrants. For non-B subtypes, which are more prevalent among migrants from PALOPs, infection occurred mostly within their communities. However, when we checked the immigrants from Brazil, the majority of infections occurred in clusters of Portuguese origin, suggesting that they were infected by the local population. Brazilian immigrants infected with subtypes G and CRF02_AG showed a very similar profile, where 58 and 60% of infections, respectively, suggest transmission between Portuguese patients and immigrants [29,30].

Our study has some limitations. The exact source of infection for an individual cannot always be identified as there may be intermediate links that have not been sampled. In the present study, as we investigated the potential geographic origin of HIV transmissions and not the putative source, our estimations are expected to be credible, except in the case of migrants' sequences that did not fall within phylogenetic clusters from specific geographic areas.

Using a large-scale dataset constituting protease and partial RT sequences from unique patients from migrants, which were linked to demographic and clinical data, we identified that a high proportion (>40%) of sequences belong to a transmission cluster and that most HIV-1 infections within transmission clusters were acquired postmigration. We also found evidence that differential patterns of HIV-1 infection acquisition when comparing immigrants from different geographic origins, with immigrants from PALOPs acquiring infection mostly within their communities, as opposed to Brazilian immigrants that were infected mostly in the context of transmission clusters with Portuguese. This probably reflects the lower level of integration of PALOP immigrants when compared with the ones originating from Brazil. These findings are of crucial importance for public health as this knowledge can be used to inform the design of prevention strategies for communicable diseases.

Acknowledgements

We would like to thank the patients and all the members of the Portuguese HIV-1 Resistance Study Group: Fátima Gonçalves (Laboratório de Biologia Molecular (LMCBM, SPC, CHLO-HEM), Lisbon, Portugal), Isabel Diogo (Laboratório de Biologia Molecular (LMCBM, SPC, CHLO-HEM), Lisbon, Portugal), Joaquim Cabanas (Laboratório de Biologia Molecular (LMCBM, SPC, CHLO-HEM), Lisbon, Portugal), Ana Patrícia Carvalho (Laboratório de Biologia Molecular (LMCBM, SPC, CHLO-HEM), Lisbon, Portugal), Sandra Fernandes (Laboratório de Biologia Molecular (LMCBM, SPC, CHLO-HEM), Lisbon, Portugal), Inês Costa (Laboratório de Biologia Molecular (LMCBM, SPC, CHLO-HEM), Lisbon, Portugal), Kamal Mansinho (Serviço de Doenças Infeciosas, Centro Hospitalar de Lisboa Ocidental, Hospital de Egas Moniz, Lisbon, Portugal), Ana Cláudia Miranda (Serviço de Doenças Infeciosas, Centro Hospitalar de Lisboa Ocidental, Hospital de Egas Moniz, Lisbon, Portugal), Isabel Aldir (Serviço de Doenças Infeciosas, Centro Hospitalar de Lisboa Ocidental, Hospital de Egas Moniz, Lisbon, Portugal), Fernando Ventura (Serviço de Doenças Infeciosas, Centro Hospitalar de Lisboa Ocidental, Hospital de Egas Moniz, Lisbon, Portugal), Jaime Nina (Serviço de Doenças Infeciosas, Centro Hospitalar de Lisboa Ocidental, Hospital de Egas Moniz, Lisbon, Portugal), Fernando Borges8, Emília Valadas (Serviço de Infeciologia, Hospital de Santa Maria, Centro Hospitalar de Lisboa Norte, Lisbon, Portugal), Manuela Doroana (Serviço de Infeciologia, Hospital de Santa Maria, Centro Hospitalar de Lisboa Norte, Lisbon, Portugal), Francisco Antunes (Serviço de Infeciologia, Hospital de Santa Maria, Centro Hospitalar de Lisboa Norte, Lisbon, Portugal), Nuno Marques (Serviço de Infeciologia, Hospital Garcia da Orta, Almada, Portugal), Maria João

Aleixo (Serviço de Infeciologia, Hospital Garcia da Orta, Almada, Portugal), Maria João Águas (Serviço de Infeciologia, Hospital Garcia da Orta, Almada, Portugal), Júlio Botas (Serviço de Infeciologia, Hospital Garcia da Orta, Almada, Portugal), Teresa Branco (Serviço de Infeciologia, Hospital Dr Fernando da Fonseca, Amadora, Portugal), Patrícia Pacheco (Serviço de Infeciologia, Hospital Dr Fernando da Fonseca, Amadora, Portugal), Luís Duque (Serviço de Infeciologia, Hospital Dr Fernando da Fonseca, Amadora, Portugal), José Vera (Serviço de Medicina Interna, Centro Hospitalar do Barreiro-Montijo, Portugal), Luís Tavares (Serviço de Medicina Interna, Centro Hospitalar do Barreiro-Montijo, Portugal), Inês Vaz Pinto (Unidade funcional VIH/SIDA, Hospital de Cascais, Cascais, Portugal), José Poças (Serviço de Infeciologia, Centro Hospitalar de Setúbal, Setúbal, Portugal), Joana Sá (Serviço de Infeciologia, Centro Hospitalar de Setúbal, Setúbal, Portugal), António Diniz (Serviço de Pneumologia, Hospital Pulido Valente, Centro Hospitalar de Lisboa Norte, Lisbon, Portugal), Margarida Serrado (Serviço de Pneumologia, Hospital Pulido Valente, Centro Hospitalar de Lisboa Norte, Lisbon, Portugal), Ana Mineiro (Serviço de Pneumologia, Hospital Pulido Valente, Centro Hospitalar de Lisboa Norte, Lisbon, Portugal), Flora Gomes (Serviço de Obstetrícia—Maternidade Alfredo da Costa, Centro Hospitalar de Lisboa Central, Lisbon, Portugal), Cristina Guerreiro (Serviço de Obstetrícia—Maternidade Alfredo da Costa, Centro Hospitalar de Lisboa Central, Lisbon, Portugal), Carlos Santos (Serviço de Medicina Interna-Centro Hospitalar do Algarve, Hospital de Portimão, Portimão, Portugal), Domitília Faria (Serviço de Medicina Interna-Centro Hospitalar do Algarve, Hospital de Portimão, Portimão, Portugal), Paula Fonseca (Serviço de Medicina Interna-Centro Hospitalar do Algarve, Hospital de Portimão, Portimão, Portugal), Paula Proença (Serviço de Infeciologia-Centro Hospitalar do Algarve, Hospital de Faro, Faro, Portugal), Telo Faria (Unidade Local de Saúde do Baixo Alentejo, Hospital José Joaquim Fernandes, Beja, Portugal), Eugénio Teófilo (Serviço de Infeciologia, Hospital de Santo António dos Capuchos, Centro Hospitalar de Lisboa Central, Lisbon, Portugal), Sofia Pinheiro (Serviço de Infeciologia, Hospital de Santo António dos Capuchos, Centro Hospitalar de Lisboa Central, Lisbon, Portugal), Isabel Germano (Serviço de Medicina 1.4, Centro Hospitalar de Lisboa Central, Hospital de São José, Lisbon, Portugal), Umbelina Caixas (Serviço de Medicina 1.4, Centro Hospitalar de Lisboa Central, Hospital de São José, Lisbon, Portugal), Margarida Bentes Jesus (Serviço de Medicina 1.4, Centro Hospitalar de Lisboa Central, Hospital de São José, Lisbon, Portugal), Nancy Faria (Serviço de Doenças Infectocontagiosas, Hospital Central do Funchal, Funchal, Madeira, Portugal), Ana Paula Reis (Serviço de Doenças Infectocontagiosas, Hospital Central do Funchal, Funchal, Madeira, Portugal), Graça Amaro (Serviço de Doenças Infeciosas, Hospital Distrital de Santarém, Santarém, Portugal), Fausto Roxo (Serviço de Doenças Infeciosas, Hospital Distrital de Santarém, Santarém, Portugal), Ricardo Abreu (Serviço de Infeciologia, Unidade de Local de Saúde de Matosinhos, Hospital Pedro Hispano, Matosinhos, Portugal) and Isabel Neves (Serviço de Infeciologia, Unidade de Local de Saúde de Matosinhos, Hospital Pedro Hispano, Matosinhos, Portugal).

Lead author for group: Kamal Mansinho (tp.eduas-nim.olhc@ohnisnamk; moc.liamg@mcbcana)

Funding: this study was supported by European Funds through grant 'Bio-Molecular and Epidemiological Surveillance of HIV Transmitted Drug Resistance, Hepatitis Co-Infections and Ongoing Transmission Patterns in Europe – BEST HOPE – (project funded through HIVERA: Harmonizing Integrating Vitalizing European Research on HIV/Aids, grant 249697); by L''Oréal Portugal Medals of Honor for Women in Science 2012 (financed through L''Oréal Portugal, Comissão Nacional da Unesco and Fundação para a Ciência e Tecnologia (FCT – http://www.fct.pt)); by FCT for funds to GHTM-UID/Multi/04413/2013; to the MigrantHIV project (financed by FCT: PTDC/DTP-EPI/7066/2014) to the INTEGRIV project (PTDC/SAU-INF/31990/2017) and to the MARVEL project (PTDC/SAU-PUB/4018/2021); by Gilead Génese HIVLatePresenters and by Associação Portuguesa para o Estudo CLínico da SIDA (Bolsa Ricardo Camacho-Bolsa de Investigação na Infecção por VIH- APECS/Gilead).

Authors' contributions: conceptualization: V.P. and A.A.; data curation: I.D., S.F., M.P., A.P., D.A. and V.P.; formal analysis: V.P., A.P., M.M., P.L. and A.A.; funding acquisition: A.A.; investigation: V.P., M.P., P.G., A.V., R.C. and P.R.S.G.; methodology: V.P., A.P., M.P. and A. A.; project administration: A.A.; resources: P.G., R.C., A. V. and P.R.S.G.; supervision: A.A.; validation: V.P., D.A. and M.P.; writing – original draft: V.P. and A.A.; writing – review and editing: V.P. and A.A.

Conflicts of interest

There are no conflict of interests.

References

1. HIV/AIDS surveillance in Europe. Published online first: 2018. doi:10.2900/523723.
2. Pimentel V, Pingarilho M, Alves D, Diogo I, Fernandes S, Miranda M, *et al*. **Molecular epidemiology of hiv-1 infected migrants followed up in portugal: trends between 2001–2017.** *Viruses* 2020; **12**:268.
3. Cortes Martins H, Ricardo Jorge Isabel Aldir D, Nacional para P, Novembro Doutor Ricardo Jorge L. *VIH e SIDA: a situação em _Portugal a de dezembro de _Documento n^{o} 149 Nacional de Saúde _ Instituto*. Available at: www.insa.pt. [Accessed 3 September 2019]
4. Tumiotto C, Bellecave P, Recordon-Pinson P, Groppi A, Nikolski M, Fleury H. **Diversity of HIV-1 in Aquitaine, Southwestern**

France, 2012–2016. *AIDS Res Hum Retroviruses* 2018; **34**:471–473.
5. Beloukas A, Psarris A, Giannelou P, Kostaki E, Hatzakis A, Paraskevis D. **Molecular epidemiology of HIV-1 infection in Europe: an overview.** *Infect Genet Evol* 2016; **46**:180–189.
6. Carvalho A, Costa P, Triunfante V, Branca F, Rodrigues F, Santos CL, *et al*. **Analysis of a local HIV-1 epidemic in Portugal highlights established transmission of non-B and non-G subtypes.** *J Clin Microbiol* 2015; **53**:1506–1514.
7. Pingarilho M, Pimentel V, Diogo I, Fernandes S, Miranda M, Pineda-Pena A, *et al*. **Increasing prevalence of HIV-1 transmitted drug resistance in Portugal: implications for first line treatment recommendations.** *Viruses* 2020; **12**:1238.
8. Wertheim JO, Leigh Brown AJ, Hepler NL, Mehta SR, Richman DD, Smith DM, Kosakovsky Pond SL. **The Global Transmission Network of HIV-1.** *J Infect Dis* 2014; **209**:304–313.
9. Pineda-Peña A-C, Pingarilho M, Li G, Vrancken B, Libin P, Gomes P, *et al*., Portuguese HIV-1 Resistance Study Group. **Drivers of HIV-1 transmission: the Portuguese case.** *PLoS One* 2019; **14**:e0218226.
10. Theys K, Libin P, Pineda-Peña A-C, Nowé A, Vandamme A-M, Abecasis AB. **The impact of HIV-1 within-host evolution on transmission dynamics.** *Curr Opin Virol* 2018; **28**:92–101.
11. Billock RM, Powers KA, Pasquale DK, Samoff E, Mobley VL, Miller WC, *et al*. **Prediction of HIV transmission cluster growth with statewide surveillance data.** *JAIDS J Acquir Immune Defic Syndr* 2019; **80**:152–159.
12. Stoner MCD, Dennis AM, Hughes JP, Eshleman SH, Sivay MV, Hudelson SE, *et al*. **Characteristics associated with human immunodeficiency virus transmission networks involving adolescent girls and young women in Human Immunodeficiency Virus Prevention Trials Network 068 Study.** *Sex Transm Dis* 2019; **46**:e46–e49.
13. De Oliveira T, Deforche K, Cassol S, Rambaut A VA. BioAfrica - HIV-1 & 2 Subtyping Tool Web Interface. Available at: http://www.bioafrica.net/subtypetool/html/subtypinghiv.html. [Accessed 8 February 2019]
14. Libin PJK, Deforche K, Abecasis AB, Theys K. **VIRULIGN: fast codon-correct alignment and annotation of viral genomes.** *Bioinformatics* 2019; **35**:1763–1765.
15. Ragonnet-Cronin M, Hodcroft E, Hué S, Fearnhill E, Delpech V, Brown AJL, *et al*. **Automated analysis of phylogenetic clusters.** *BMC Bioinformatics* 2013; **14**:317.
16. Bennett DE, Camacho RJ, Otelea D, Kuritzkes DR, Fleury H, Kiuchi M, *et al*. **Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update.** *PLoS One* 2009; **4**:e4724.
17. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjörnsson J. **Defining HIV-1 transmission clusters based on sequence data.** *AIDS* 2017; **31**:1211–1222.
18. Romero-Severson EO, Bulla I, Leitner T. **Phylogenetically resolving epidemiologic linkage.** *Proc Natl Acad Sci USA* 2016; **113**:2690–2695.
19. Li L, Han N, Lu J, Li T, Zhong X, Wu H, *et al*. **Genetic characterization and transmitted drug resistance of the HIV type 1 epidemic in men who have sex with men in Beijing, China.** *AIDS Res Hum Retroviruses* 2013; **29**:633–637.
20. Araújo PMM, Carvalho A, Pingarilho M, Abecasis AB, Osório NS. **Characterization of a large cluster of HIV-1 A1 infections detected in Portugal and connected to several Western European countries.** *Sci Rep* 2019; **9**:7223.
21. Hemelaar J, Elangovan R, Yun J, Dickson-Tetteh L, Fleminger I, Kirtley S, *et al*., WHO–UNAIDS Network for HIV Isolation Characterisation. **Global and regional molecular epidemiology of HIV-1, 1990–2015: a systematic review, global survey, and trend analysis.** *Lancet Infect Dis* 2019; **19**:143–155.
22. Paraskevis D, Beloukas A, Stasinos K, Pantazis N, De Mendoza C, Bannert N, *et al*., CASCADE collaboration of EuroCoord. **HIV-1 molecular transmission clusters in nine European countries and Canada: Association with demographic and clinical factors.** *BMC Med* 2019; **17**:4.
23. Lorenzin G, Gargiulo F, Caruso A, Caccuri F, Focà E, Celotti A, *et al*. **Prevalence of non-B HIV-1 subtypes in north italy and analysis of transmission clusters based on sequence data analysis.** *Microorganisms* 2019; **8**:36.
24. Burns FM, Arthur G, Johnson AM, Nazroo J, Fenton KA. **United Kingdom acquisition of HIV infection in African residents in London: more than previously thought.** *AIDS* 2009; **23**:262–266.
25. Xiridou M, van Veen M, Coutinho R, Prins M. **Can migrants from high-endemic countries cause new HIV outbreaks among heterosexuals in low-endemic countries?** *AIDS* 2010; **24**:2081–2088.
26. Hamers FF, Downs AM. **The changing face of the HIV epidemic in western Europe: what are the implications for public health policies?** *Lancet* 2004; **364**:83–94.
27. Ross J, Cunningham CO, Hanna DB. **HIV outcomes among migrants from low-income and middle-income countries living in high-income countries.** *Curr Opin Infect Dis* 2018; **31**:25–32.
28. Desgrées-du-Loû A, Pannetier J, Ravalihasy A, Gosselin A, Supervie V, Panjo H, *et al*., Parcours Study Group. **Sub-Saharan African migrants living with HIV acquired after migration, France, ANRS PARCOURS study, 2012 to 2013.** *Eurosurveillance* 2015; **20**:doi: 10.2807/1560-7917.ES.2015.20.46.30065.
29. Velasco-de-Castro CA, Grinsztejn B, Veloso VG, Bastos FI, Pilotto JH, Fernandes N, Morgado MG. **HIV-1 diversity and drug resistance mutations among people seeking HIV diagnosis in voluntary counseling and testing sites in Rio de Janeiro, Brazil.** *PLoS One* 2014; **9**:e87622.
30. Machado LFA, Ishak MOG, Vallinoto ACR, Lemos JAR, Azevedo VN, Moreira MRC, *et al*. **Molecular epidemiology of HIV type 1 in Northern Brazil: identification of subtypes C and D and the introduction of CRF02_AG in the Amazon Region of Brazil.** *AIDS Res Hum Retroviruses* 2009; **25**:961–966.