LA-UR- 08-8078

Approved for public release; distribution is unlimited.

Title:	Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach
Author(s):	D. Paraskevis, Katholieke Universiteit Leuven O. Pybus, University of Oxford G. Magiorkinis, Katholieke Universiteit Leuven A. Hatzakis, Katholieke Universiteit Leuven Thomas Leitner, Los Alamos National Laboratory et al.
Intended for:	PLOS ~~Met~~ Medicine

Los Alamos
NATIONAL LABORATORY
EST. 1943

Form 836 (7/06)

Classification: Biological Sciences, Microbiology

Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach

D. Paraskevis[1,2*], O. Pybus[3], G. Magiorkinis[1], A. Hatzakis[1], A.M.J. Wensing[4], D.A. van de Vijver4, J. Albert[5], G. Angarano[6], B. Åsjö[7], C. Balotta[8], E. Boeri[9], R. Camacho[10], M.-L. Chaix[11], D. Costagliola[12], A. De Luca[13], C. de Mendoza[14], I. Derdelinckx[1], Z. Grossman[15], O. Hamouda[16], R. Hemmer[17], I.M. Hoepelman[4], A. Horban[18], K. Korn[19], C. Kücherer[16], T. Leitner[5], C. Loveday[20], E. MacRae[20], I. Maljkovic[5], L. Meyer[21], C. Nielsen[22], E.L.M. Op de Coul[23], V. Ormaasen[24], L. Perrin[25], E. Puchhammer-Stöckl[26], F. Roman[17], L. Ruiz[27], M. Salminen[28], J.-C. Schmit[17], R. Schuurman[4], V. Soriano[14], J. Stanczak[18], M. Stanojevic[29], K. Van Laethem[1], M. Violin[8], S. Yerly[25], M. Zazzi[30], C.A. Boucher[4] and A.-M. Vandamme[1], for the SPREAD Programme

[1]Katholieke Universiteit Leuven, Rega Institute for Medical research, Minderbroederstraat 10, B-3000 Leuven, Belgium

[2]National Retrovirus Reference Center, Department of Hygiene Epidemiology and Medical Statistics, Medical School, University of Athens, M. Asias 75, GR-11527, Athens, Greece

[3]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.

[4]University Medical Center Utrecht, Department of Virology, G04.614, Heidelberglaan 100, 3584 CX, Utrecht

[5]Dept of Virology, Swedish Institute for Infectious Disease Control (SMI), SE-171 82 Solna, Sweden

[6]University of Foggia, Clinic of Infectious Diseases, Ospedali Riuniti - Via L. Pinto 71100 Foggia, Italy

[7]Center for Research in Virology, University of Bergen, Bergen High Technology Center, N-5020 Bergen, Norway

[8]University of Milan, Institute of Infectious and Tropical Diseases, Via Festa del Perdono 7, 20122 Milano, Italy

[9]Diagnostica and Ricerca San Raffaele, Centro San Luigi, I.R.C.C.S. Istituto Scientifico San Raffaele, Milan, Italy,

[10]Universidade Nova de Lisboa, Laboratorio de Virologia, Rua da Junqueira 96 1349-008 Lisboa

[11]Hôpital Necker-Enfants Malades, 149 rue de Sèvres, 75743 Paris cedex 15, France

[12]INSERM U263 et SC4, Faculté de médecine Saint-Antoine, Université Pierre et Marie Curie, 27 rue de Chaligny, F-75571 Paris, France

[13] Department of Infectious Diseases, Catholic University, L.go A. Gemelli, 8 00168 Rome, Italy

[14]Hospital Carlos III, Hospital Carlos III, Madrid, Spain

[15]Natl. HIV Reference Lab, Central Virology, Public Health Laboratories, MOH Central Virology, Sheba Medical Center, 2 Ben-Tabai Street, Israel

[16]Robert Koch Institut (RKI), Nordufer 20, 13353 Berlin, Germany

[17] Centre Hospitalier de Luxembourg, Retrovirology Laboratory, National service of Infectious Diseases, 4 Rue Barblé, L-1210 Luxembourg

[18]Hospital for Infectious Diseases, Center for Diagnosis & Therapy Warsaw
37, Wolska Str. 01-201 Warszawa, Poland
[19]University of Erlangen, Schlossplatz 4, D-91054 Erlangen, Germany
[20]ICVC Charity Laboratories, 3d floor, Apollo Centre Desborough Road
High Wycombe, Buckinghamshire, HP11 2QW, UK
[21]Inserm, U822, Le Kremlin-Bicêtre, F-94276, France
[22]Statens Serum Institut Copenhagen, Retrovirus Laboratory, department of virology,
building 87, Division of Diagnostic Microbiology 5, Artillerivej
2300 Copenhagen Denmark
[23]Laboratory for Infectious Diseases and Perinatal Screening, National Institute for Public
Health and the Environment (RIVM), 3720 BA Bilthoven
Bilthoven, The Netherlands
[24]Ullevaal University Hospital, Department of Infectious Diseases
Kirkeveien 166, N-0407 Oslo, Norway
[25]Les Hopitaux Universitaires de Geneve, Department of Internal Medicine
Division of Infectious Diseases, Geneva, Switzerland
[26]Institute of Virology, Medical University Vienna, Kinderspitalgasse 15, Vienna, Austria
[27]IrsiCaixa Foundation, Hospital Germans Trias i Pujol, Ctra. de Canyet s/n, 08916
Badalona (Barcelona), Spain
[28]National Public Health Institute, HIV laboratory and department of infectious disease
epidemiology, Mannerheimintie 166, FIN-00300 Helsinki, Finland
[29]University of Belgrade School of Medicine, Institute of Microbiology and Immunology
Virology Department, Dr Subotica 1, 11000 Belgrade, Serbia
[30]Section of Microbiology, Department of Molecular Biology, University of Siena, Italy

*To whom correspondence should be addressed:

Dimitrios Paraskevis, PhD
National Retrovirus Reference Center,
Department of Hygiene Epidemiology and Medical Statistics,
Medical School, University of Athens,
M. Asias 75, Athens,
GR-11527, GREECE
Tel: +30 210 7486382/7462090
Fax: +30 210 7462190
Email address: **dparask@cc.uoa.gr, dparask@med.uoa.gr**

Manuscript information: ???

Abstract

The prevalence and the origin of HIV-1 subtype B, the most prevalent circulating clade among the long-term residents in Europe, has extensively been studied. However the spatial diffusion of the epidemic from the perspective of the virus has not previously been traced. In the current study we inferred the migration history of HIV-1 subtype B by way of a phylogeography of viral sequences sampled from 17 European countries and the USA. Migration events were inferred from viral phylogenies by character reconstruction using parsimony. According to the significant migratory pathways, we show that there are considerable differences across Europe. Specifically, the UK, Spain, Greece, Italy, and Poland provide sources shedding HIV-1; Austria, Belgium, Denmark, France, Germany, Israel, Luxembourg, Switzerland and the Netherlands on the other hand, are migratory targets while for Portugal and Sweden we inferred significant bidirectional migration. Notably, the USA provides also source for HIV-1 migration towards Europe. Subtype B phylogeographies provided a new insight about the significant pathways of virus dispersal across Europe, suggesting that intervention strategies should also address tourists, travellers and migrants.

Introduction

Pandemic HIV-1 group M infection originated in Africa from the simian immunodeficiency virus (SIVcpz) infecting chimpanzees (1-6). The introduction of HIV-1 into Europe occurred mainly through homosexual contacts or needle sharing in or from the USA (7-11), or through heterosexual contacts with individuals from Central Africa (12, 13). At the beginning of the HIV-1 epidemic (the early 1980's) the prevalence of HIV-1 infection was higher among men having sex with other men (MSM) than among heterosexuals. For this reason subtype B - which was identified at a high prevalence among MSM in the USA - was the predominant clade in Europe. The prevalence of non-B subtypes in Europe has been increasing over the last years (14-29). However, the AIDS epidemic among the long-term residents is still dominated by viruses assigned to subtype B (30, 31)

RNA viruses provide measurably evolving populations characterized by very high nucleotide substitution rate (32, 33). Phylogenies can be used for molecular epidemiology studies and notably they contain information about temporal and spatial dynamics of the virus (34). The latter is the geographic pattern of viral lineages sampled from different localities, also termed as phylogeography, tracking the migration of the virus. For several viral infections, the dispersal of the parasite and its host cannot be easily tracked, therefore suggesting that phylogenies may be a better way to monitor migratory pathways of the virus. This methodology has been recently applied to phylogeographic studies of influenza A (H5N1) (35) and HCV (36) epidemics showing the pathways of viral dispersal.

Thus, phylogenies are the 'state of the art' in characterizing viral genealogy and evolution and also serve as tools to track migration for organisms for which there is no other way to monitor their dispersal. Although several phylogenetic studies have analyzed HIV-1 clades by geographic region in Europe, none has inferred the history of virus's migration through its phylogeny. In the present study, we inferred the migration history of HIV-1 virus among 17 countries in Europe, as well as between Europe and the USA, by way of a phylogeography of subtype B sequences.

Results

Migration events were inferred through virus phylogenies by using the Slatkin and Maddison's method (37) (illustrated in Figure 1). Trees were built by Bayesian method and countries from which sequences were sampled were assigned to each of tip of $4x10^4$ credible trees estimated by Metropolis Coupled Markov Chains Monte Carlo (MC^3). Inclusion of a large number of trees takes into account phylogenetic uncertainty, because migration events are estimated over a set of credible trees than a single one. Phylogenies of subtype B sequences from 17 countries in Europe and Israel (Supplementary information Table 1) showed no considerable grouping of sequences by country, therefore suggesting high levels of mixing within Europe (Figure 2). However, the total number of location changes between countries (migration events) for all trees was significantly lower than expected by chance under the null hypothesis of panmixis. This finding indicates that although there is a high level of HIV dispersal between countries, there is still geographic subdivision among the subtype B lineages analyzed (Figure 2).

Statistical Phylogeography To test the significance of specific pathways of mobility across different countries, we estimated the expected number of changes, under the null hypothesis of complete geographic mixing, for each pair of countries (Supplementary information Tables 2 and 3), as described previously (36). The results of this test showed major differences across Europe (Supplementary information Figure 1a, b). In particular, for 6 countries (Austria, Denmark, France, Germany, Israel and Switzerland) no significant exporting migration was observed; whereas the UK, Spain, Greece, Italy, Poland, Portugal and Sweden, on the other hand, appeared as source of subtype B mobility (high levels of exporting migration; "From")* to other countries (Supplementary information Figure 1a, b). Notably, UK's and Greece's migratory targets were dispersed to 6 and 8 countries, respectively; while high levels of HIV migration (>2.5) was detected from Italy to Austria and from Portugal to Luxembourg. On the other hand, Belgium, Luxembourg, Norway and the Netherlands showed only limited export of HIV-1 subtype B (Supplementary information Figure 1a, b).

Major migratory targets of HIV-1 subtype B (importing migration; "To") were Austria, Belgium, Denmark, France, Germany, Luxembourg, Portugal, Sweden, Switzerland and the Netherlands• (Supplementary information Figure 1c, d), while limited migration was observed into the UK, Greece, Israel, Norway and Poland (Supplementary information Figure 1c, d). No significant migration was detected, on the other hand, to Spain and Italy.

Based on these findings, evidence for directional HIV dispersion was detected where UK, Spain, Greece, Italy, and Poland acted as sources of migration events ("exporters") (Figure 3); Austria, Belgium, Denmark, France, Germany, Israel♠, Luxembourg, Switzerland and the Netherlands provided migratory targets ("importers") (Figure 3), while significant bidirectional HIV migration was found for Sweden and Portugal (Figure 3). For Norway no significantly importing or exporting migration was detected.

---

* In case that significant migration was detected from a country to more than 2 others, the former was designated as "exporter"; Portugal was classified within the "exporters" due to the high migration (>2.5) inferred from Portugal towards Luxembourg.

• A similar criterion as for the "From" migration was used to assign countries.

♠ Israel was classified among the "importers" due to the absence of significant exporting migration from this country.

In addition to the statistical phylogeography, we estimated total mobility by way of summing up migration events from and to each country (Supplementary information Figure 2a, b and c; shown also in detail in Supplementary information Table 4). Israel, Norway Switzerland and the Netherlands, show the highest ratio of importing versus exporting HIV-1 migration (upper 5th percentile), while Greece, Italy, Spain and Poland show the lowest ratio (Supplementary information Figure 2c). The former provide the major migratory "importers", while the latter provide the areas seeding migration. Notably, this classification is in accordance with the statistical phylogeography results.

Inferring migration between Europe and the USA. Moreover, we estimated the history of HIV-1 subtype B dispersal between Europe and the USA by including HIV-1 sequences sampled from 10 US cities. Statistical phylogeography revealed 6 significant migratory pathways from the USA to Spain, Germany, Sweden, the UK, Greece and Luxembourg and from the latter 3 countries to the USA. These findings suggest that the USA is still a source of HIV-1 migration towards Europe (Figure 4), but on the other hand HIV-1 significant migratory pathways were also detected from Europe into the USA.

Discussion

Our study provided important clues about HIV-1 subtype B spatial diffusion, defining the major sources and targets for migratory events, as well as localities with bidirectional viral dispersion. Moreover, we show that there was no extensive clustering of lineages by country, suggesting that transmission of subtype B, represented by the phylogenies of the sampled sequences, mostly occurred across individual countries.

In particular, Greece, Italy and Spain attract many tourists, especially from Central Europe, thus suggesting that HIV dispersal from Southern to Central Europe may, at least in part, occur by tourists infected during their holidays. The non-importing nature of the subtype B infection in these countries is probably related to the fact that there is less mobility into Southern Europe from other countries, where this subtype is prevalent. Poland showed a similar pattern of HIV migration related to the high prevalence of injecting drug users (50%) among the total HIV-infected population, and, therefore, explaining the spread of the infection within the country. Additionally, exporting migration is probably related to the emigration of Polish labourers to other European countries. Interestingly, HIV infections from Poland were mainly exported to countries in proximity (e.g. Belgium, Luxembourg, Norway and the Netherlands), except for Israel, possibly due to the historical links between the Jewish community of Poland and Israel. For the UK, on the other hand, a more disperse spatial migratory pattern was inferred. The reason why the UK was also classified among the "exporters" is probably due to the high population mobility towards the UK (labour and student mobility, tourism, immigration, etc).

For countries classified among the HIV migratory targets (Austria, Belgium, Denmark, Luxembourg, France, Germany, Israel, Switzerland, and the Netherlands) the epidemic was mainly imported due to the high HIV mobility to these countries: The Netherlands is among the countries in Europe with the most diverse geographical origin among newly diagnosed patients, confirmed by the high percentage of non-Dutch individuals among the newly HIV-infected patients in 2003 (38). Moreover, because of its policies, the Netherlands attracts foreign drug users and male homosexuals, two populations known to be at higher risk for HIV infection. Similarly, there is a high proportion of non-long term residents among the subtype B infected population in Israel. For Norway this percentage is comparable to other European countries, thus suggesting that the importing subtype B epidemic is the result of international travelling of the Norwegians rather than immigration. Another significant pathway was tracked from Italy to Austria, in accordance with the high inflow from Italy during recent years (http://www.migrationinformation.org/datahub/countrydata/). Similarly, countries identified as the major source for HIV diffusion (Poland, UK, Spain and Greece) rank among those with the highest rate of population mobility towards Germany (http://www.migrationinformation.org/datahub/countrydata/). Moreover, the fact that 13% of the population of Luxembourg is of Portuguese origin provides a plausible explanation for the migratory pathway from Portugal.

We should note that subtype B migratory pathways inferred through viral phylogenies cannot be directly validated by other sources of information (epidemiological figures, mobility and immigration information, tourism, etc) for the reason that this information is not subtype-specific. Moreover, due to the high mobility of population within Europe and the complexity of the epidemic spread, information about the locus of infection for an individual doesn't necessarily match with the geographic origin of the source. On the other hand, phylogenetic analysis of viral sequences provides a realistic approach for the reconstruction of HIV transmission chains or networks (39-44), therefore

suggesting that statistical phylogeography is appropriate for inferring the spatial dispersal of a viral epidemic. Additionally in our study, we used the largest available dataset of subtype B sequences (CATCH) available at the time of analysis, to characterize relationships between viruses infecting individuals from 17 European countries. However, one of the potential drawbacks is that migratory pathways can change over time according to the patterns of the epidemic spread.

HIV-1 subtype B phylogeographies provide a new insight into the pathways of virus migration across Europe. Specifically, we estimated significant migration and described in detail the country-wise movement among the sampled countries. The results suggest that mobility of the virus matches mobility of the host, such that in order to reduce further spread of the epidemic, prevention measures should not only be directed towards national populations, but also towards migrants, travellers and tourists who are the major sources and targets of HIV dispersal.

Methods

Samples

European sequences HIV-1 partial reverse transcriptase (RT) sequences were sampled from HIV-1 seropositive individuals who had never received antiretroviral drugs (ARV) as described previously (45, 46). Specifically, partial RT sequences were sampled from 17 countries in Europe including Israel. Due to unavailability of protease PR sequences from one country, all PR sequences were excluded from the analysis. All sequences were collected from geographically distinct centres across the participating countries, except for Belgium and the Netherlands, where HIV-1 sequences were sampled from a single geographic area. The samples were collected during 1996-2003 and only those classified as subtype B were included in the analysis. The subtyping process was performed by phylogenetic analysis as described previously (45). The sequences were collected for a retrospective study of prevalence of HIV-1 resistance in naïve individuals in Europe (CATCH) (45). This dataset provides the largest available dataset of HIV-1 sequences sampled in Europe so far. The prevalence of the transmission risk groups among the study population is shown in Supplementary information Table 1.

US sequences HIV-1 partial RT sequences used in the migration analysis were downloaded from the Stanford HIV drug resistance database (http://hivdb.stanford.edu/). These sequences were used before for studying the epidemiology of antiretroviral drug resistance among drug-naïve HIV-1-infected persons in the USA (46). This dataset was selected for the following reasons: 1) the samples were collected at a similar time period with the "European" dataset and 2) HIV-1 sequences were collected from drug-naïve persons of different ethnicities enrolled from 10 different US cities spanning from East to West US (New York, Newark, Miami, New Orleans, Houston, Detroit/Grand Rapids, Denver, San Diega, San Francisco) (46); as a result the US and the European datasets are comparable to each other in terms of sampling period, the use of sequences from drug-naïve patients and the large geographic coverage. For the migration study, 100 HIV-1 subtype B sequences were randomly selected among the total number of sequences (1082) sampled from the USA.

Phylogenetic analyses

Sampling strategy. For the estimation of country-wise clustering (migration), first we need to infer the phylogenies of the sequences under study. One of the issues to be addressed was how many sequences needed to be included for each country. The dataset size needs to be large enough as: 1) to include most of the available information from each country and 2) to estimate rare migration events. On the other hand, we had to restrict the number of sequences to keep the computation time needed for phylogenetic inference reasonable, while maintaining an informative number of sequences required for the calculation of migration events. For this reason, we performed a preliminary analysis of migration for 4 countries including 10, 20 or 25 sequences per country. For each dataset, we tested whether the distribution of the total number of migration events across the set of all credible trees differed significantly from a distribution of randomly generated trees (phylogenetic inference was performed by Bayesian method). The results of this preliminary analysis showed that with 25 sequences per country, the largest number of countries reached significantly different migration levels than compared to the

distribution for a random set of trees ($P<0.01$). Consequently, we included 25 sequences per country in the analyses datasets. As a result of choosing an equal number of strains per country, irrespective of the prevalence or the total number of infected individuals across Europe, we calculated the relative mobility per infected individual. Therefore, the numbers in the migration matrices are directly comparable reflecting actual differences in mobility between countries.

To increase the number of analyzed sequences and, thus to minimize the potential sampling bias, we created 4 distinct datasets according to the following strategy: 1) 25 sequences were randomly selected for each country and, more specifically, by implementing random sampling without replacement. The same process was repeated for 4 rounds (yielding 4 datasets) in total, each time within the remaining sequences for each country. 2) If the number of sequences was not adequate (<100) for 4 rounds, sequences were re-sampled randomly from the existing datasets without replacement. Therefore, each sequence was presented only once over all rounds, except for countries with less than 100 subtype B sequences.

In particular, for Greece, Israel, Norway and the Netherlands for which less than 26 subtype B sequences were available in the CATCH study[1]; all sequences were included in each round. For countries with >100 subtype B sequences, 4 different sets of 25 sequences were included in each dataset. Finally, for countries with an intermediate number of sequences (25< N< 100), the 4 sets were partially overlapping, but sequences were presented only once in each round. Phylogenetic analyses for the estimation of the migration process were performed in 4 different datasets each consisting of 399 sequences sampled as described above. Specifically, the analyzed sequences (1001) were from Austria (64), Belgium (57), Denmark (72), France (100), Germany (53), Greece (25), Israel (21), Italy (84), Luxembourg (100), The Netherlands (16), Norway (12), Poland (34), Portugal (29), Spain (84), Sweden (75), Switzerland (100) and the United Kingdom (75).

To estimate the number of migration events between Europe and the USA, phylogenetic analyses were performed in 4 datasets each one including the same set sequences from Europe (dataset 1 as described in sampling strategy paragraph plus 25 sequences from the USA (46) with a total number of 424 sequences per dataset (399 from Europe + 25 from the USA).

Alignment and phylogenetic tree reconstruction. The alignment of the subtype B partial RT sequences sampled from 1427 individuals was performed using CLUSTAL W version 1.74 (47) and manually edited according to the encoded reading frame. The edited length of the alignment was 531 nucleotides. The reported prevalence of resistance among individuals infected with subtype B in our study is 12.9% (45) and previous studies have shown that especially in partial *pol*, resistance mutations may result in incorrect clustering of the HIV-1 sequences under study (43). Therefore, phylogenetic analyses was performed using only the 3rd codon positions of the alignment (177 nucleotides) to avoid bias by selective pressure and thus minimize potential error due to parallel evolution through accumulation of mutations at resistance sites.

Phylogenetic trees for each dataset were inferred by Bayesian method under the HKY model (base frequency parameters and transitions/transversions ratio are estimated) including a ! distributed rates heterogeneity among sites as implemented in MrBayes (v 3.0B4) (48). For the European datasets, four Metropolis Coupled Markov Chains Monte

---

[1] We should note at the time of the analysis no additional RT sequences for these countries were available at the NCBI sequence database

Carlo (MC^3) chains ran for 2-3 x 10^7 generations while for the European-US datasets (MC^3) chains ran for 3-3.5 x 10^7 generations with a burnin of 1-2.5 x 10^7, depending on the dataset. For all runs, the estimated sampling size (ESS) was calculated for a random sample of windows by using Tracer program (v 1.2) (http://evolve.zoo.ox.ac.uk/software.html?id=tracer). Appropriate mixing was achieved in the posterior target distribution (ESS>100). Alignment of DNA sequences and phylogenetic analyses were performed in a linux cluster of 8 nodes.

Inference of migration events.

All credible trees (10^4), for each dataset, were converted to midpoint rooted trees by using PAUP*4.0b10 (49) and used for the estimation of the HIV-1 migration events by using the cladistic approach first described by Slatkin and Maddison (37), as implemented in MacClade (50). Specifically, all the nodes of the inferred trees were assigned with a character according to the geographic origin (e.g. 0, 1, 2, 3 for Austria, Belgium, Denmark, France, etc). The algorithm reconstructs "ancestral" states that in our case correspond to countries, at each internal node by the criterion of parsimony (37). Parsimony selects the reconstruction that minimizes the total number of steps on the tree. Ambiguous changes were not counted.

When two branches from 2 different locations (e.g. 0 and 1) join with each other, and thus more than one character can be reconstructed at the node, then the ancestor state at the internal node is assigned to be the union of the two characters [0, 1] that is assigned a migration event. If this number between two groups of sequences remains low, the possibility for migration events between these particular groups also remains low.

Specifically, the migration events between HIV-1 sequences sampled in different locations were estimated for each dataset according to the following method: 1) for nodes with more than one equally parsimonious reconstructions (e.g. 0, 1 or 0), implicit examination of all most parsimonious reconstructions (MPRs) was used in case of a big number of MPRs (51, 52), while explicit examination was used in case of a small number of MPR, as implemented in MacClade. As a result, for a particular type of character change, e.g. (1-31, 35-37, 39-48, 52-54) MacClade reports a minimum, a maximum and a average number of (34) changes estimated over all possible MPRs. We estimated the average number of migration events for each tree used in the analyses. 2) Polytomies that correspond to nodes with more than two descendant nodes were interpreted as regions of uncertain evolution (soft polytomies) as implemented in MacClade.

Construction of migration matrices For each dataset a 17x17 migration matrix was estimated between HIV-1 sequences sampled in different European countries. Each migration event was calculated as the median of the distribution estimated from all trees (10^4) used in the analysis. In the matrix, all 'from' events and 'to' events are pooled per country. The final migration matrix was based on the total number of trees ($4x10^4$) after combining the 4 datasets. More specifically each migration event was the median of the distribution of the total number of trees. Similarly, a 18x18 migration matrix was estimated using sequences sampled from Europe and the USA. The final matrix was based on the total number of trees from the 4 datasets.

Hypothesis testing

To assess whether the migration pattern within the study population exhibited panmixis, we tested if the distribution of the total number of migration events estimated

from the set of all credible trees is significantly different from the distribution of migration events estimated from a set of random trees (10^4) of equal number of sequences as the dataset under study. This analysis was performed using MacClade (50).

Statistical phylogeography To further estimate which migration events were significantly different from the expected number of changes under the null hypothesis of full geographic mixing of HIV-1 sequences, 10^3 credible trees were selected among the total number (10^4) from a single dataset (1 tree was selected at every step of 10). The sampling locations of all countries were randomized 20 times and the migration events were re-calculated for the whole set of trees in each dataset (10^3). The average number of the migration events estimated over the 20 replicates gave the expected number of changes under panmixis. Benjamini-Hochberg (BH) correction was applied for the multiple corrections test run for the HIV-1 subtype B migration matrices.

The differences between the observed and the expected values indicate the levels of HIV-1 country-dependent structure in the dataset, and thus also of the relative mobility of the virus between countries. This strategy allowed estimating significant differences also when an unequal number of strains were included per country.

Acknowledgements

We thank the patients and doctors throughout Europe, for their consent and support for the study. The study was supported in part the European Commission (QLK2-CT-2001-01344) by the Hellenic Scientific Society for the Study of AIDS and STDs, by the Belgian AIDS Reference Laboratory fund and the Belgian Fonds voor Wetenschappelijk Onderzoek (F.W.O. nr G.0266.04).

References

1. Peeters, M., Honore, C., Huet, T., Bedjabaga, L., Ossari, S., Bussi, P., Cooper, R. W. & Delaporte, E. (1989) *Aids* **3,** 625-30.
2. Peeters, M., Fransen, K., Delaporte, E., Van den Haesevelde, M., Gershy-Damet, G. M., Kestens, L., van der Groen, G. & Piot, P. (1992) *Aids* **6,** 447-51.
3. Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M. & Hahn, B. H. (1999) *Nature* **397,** 436-41.
4. Hahn, B. H., Shaw, G. M., De Cock, K. M. & Sharp, P. M. (2000) *Science* **287,** 607-14.
5. Sharp, P. M., Bailes, E., Gao, F., Beer, B. E., Hirsch, V. M. & Hahn, B. H. (2000) *Biochem Soc Trans* **28,** 275-82.
6. Heeney, J. L., Rutjens, E., Verschoor, E. J., Niphuis, H., ten Haaft, P., Rouse, S., McClure, H., Balla-Jhagjhoorsingh, S., Bogers, W., Salas, M., Cobb, K., Kestens, L., Davis, D., van der Groen, G., Courgnaud, V., Peeters, M. & Murthy, K. K. (2006) *J Virol* **80,** 7208-18.
7. Brunet, J. B., Bouvet, E. & Massari, V. (1984) *Ann N Y Acad Sci* **437,** 334-9.
8. Glauser, M. P. & Francioli, P. (1984) *Eur J Clin Microbiol* **3,** 55-8.
9. Melbye, M., Biggar, R. J., Ebbesen, P., Sarngadharan, M. G., Weiss, S. H., Gallo, R. C. & Blattner, W. A. (1984) *Br Med J (Clin Res Ed)* **289,** 573-5.
10. Casado, C., Urtasun, I., Saragosti, S., Chaix, M. L., de Rossi, A., Cattelan, A. M., Dietrich, U. & Lopez-Galindez, C. (2000) *AIDS Res Hum Retroviruses* **16,** 299-304.
11. Lukashov, V. V., Kuiken, C. L., Vlahov, D., Coutinho, R. A. & Goudsmit, J. (1996) *AIDS Res Hum Retroviruses* **12,** 1179-83.
12. Clumeck, N., Sonnet, J., Taelman, H., Mascart-Lemone, F., De Bruyere, M., Vandeperre, P., Dasnoy, J., Marcelis, L., Lamy, M., Jonas, C. & et al. (1984) *N Engl J Med* **310,** 492-7.
13. Vittecoq, D., May, T., Roue, R. T., Stern, M., Mayaud, C., Chavanet, P., Borsa, F., Jeantils, P., Armengaud, M., Modai, J. & et al. (1987) *Lancet* **1,** 612-5.
14. Chaix, M. L., Descamps, D., Harzic, M., Schneider, V., Deveau, C., Tamalet, C., Pellegrin, I., Izopet, J., Ruffault, A., Masquelier, B., Meyer, L., Rouzioux, C., Brun-Vezinet, F. & Costagliola, D. (2003) *Aids* **17,** 2635-43.
15. Descamps, D., Chaix, M. L., Andre, P., Brodard, V., Cottalorda, J., Deveau, C., Harzic, M., Ingrand, D., Izopet, J., Kohli, E., Masquelier, B., Mouajjah, S., Palmer, P., Pellegrin, I., Plantier, J. C., Poggi, C., Rogez, S., Ruffault, A., Schneider, V., Signori-Schmuck, A., Tamalet, C., Wirden, M., Rouzioux, C., Brun-Vezinet, F., Meyer, L. & Costagliola, D. (2005) *J Acquir Immune Defic Syndr* **38,** 545-52.
16. Deroo, S., Robert, I., Fontaine, E., Lambert, C., Plesseria, J. M., Arendt, V., Staub, T., Hemmer, R., Schneider, F. & Schmit, J. C. (2002) *Aids* **16,** 2461-7.
17. Machuca, R., Bogh, M., Salminen, M., Gerstoft, J., Kvinesdal, B., Pedersen, C., Obel, N., Nielsen, H. & Nielsen, C. (2001) *Scand J Infect Dis* **33,** 697-701.
18. Sonnerborg, A., Durdevic, S., Giesecke, J. & Sallberg, M. (1997) *AIDS Res Hum Retroviruses* **13,** 343-5.

19. Snoeck, J., Van Laethem, K., Hermans, P., Van Wijngaerden, E., Derdelinckx, I., Schrooten, Y., van de Vijver, D. A., De Wit, S., Clumeck, N. & Vandamme, A. M. (2004) *J Acquir Immune Defic Syndr* **35,** 279-85.
20. Lospitao, E., Alvarez, A., Soriano, V. & Holguin, A. (2005) *HIV Med* **6,** 313-20.
21. Boni, J., Pyra, H., Gebhardt, M., Perrin, L., Burgisser, P., Matter, L., Fierz, W., Erb, P., Piffaretti, J. C., Minder, E., Grob, P., Burckhardt, J. J., Zwahlen, M. & Schupbach, J. (1999) *J Acquir Immune Defic Syndr* **22,** 174-9.
22. Op de Coul, E. L., Coutinho, R. A., van der Schoot, A., van Doornum, G. J., Lukashov, V. V., Goudsmit, J. & Cornelissen, M. (2001) *Aids* **15,** 2277-86.
23. Paraskevis, D., Magiorkinis, E., Magiorkinis, G., Sypsa, V., Paparizos, V., Lazanas, M., Gargalianos, P., Antoniadou, A., Panos, G., Chrysos, G., Sambatakou, H., Karafoulidou, A., Skoutelis, A., Kordossis, T., Koratzanis, G., Theodoridou, M., Daikos, G. L., Nikolopoulos, G., Pybus, O. G. & Hatzakis, A. (2007) *J Infect Dis* **196,** 1167-76.
24. Bobkov, A., Kazennova, E., Selimova, L., Bobkova, M., Khanina, T., Ladnaya, N., Kravchenko, A., Pokrovsky, V., Cheingsong-Popov, R. & Weber, J. (1998) *AIDS Res Hum Retroviruses* **14,** 669-76.
25. Holmes, E. C., Zhang, L. Q., Robertson, P., Cleland, A., Harvey, E., Simmonds, P. & Leigh Brown, A. J. (1995) *J Infect Dis* **171,** 45-53.
26. Brown, A. J., Lobidel, D., Wade, C. M., Rebus, S., Phillips, A. N., Brettle, R. P., France, A. J., Leen, C. S., McMenamin, J., McMillan, A., Maw, R. D., Mulcahy, F., Robertson, J. R., Sankar, K. N., Scott, G., Wyld, R. & Peutherer, J. F. (1997) *Virology* **235,** 166-77.
27. Thomson, M. M. & Najera, R. (2001) *Clin Infect Dis* **32,** 1732-7.
28. Liitsola, K., Holmstrom, P., Laukkanen, T., Brummer-Korvenkontio, H., Leinikki, P. & Salminen, M. O. (2000) *Scand J Infect Dis* **32,** 475-80.
29. Palma, A. C., Araujo, F., Duque, V., Borges, F., Paixao, M. T. & Camacho, R. (2007) *Infect Genet Evol* **7,** 391-8.
30. Paraskevis, D., Wensing, AMJ,Vercauteren, J,van de Vijver, DA, Albert, J, Asjo, B, et al, on behalf of the SPREAD-programme (2006) in *1st International Workshop on HIV Transmission*, Toronto, Canada), pp. Abstract No 34, p 31.
31. Abecasis, A., Wensing, A., Paraskevis, D., Vercauteuren, D., Van de Vijver, D., Albert, J. & et al, o. b. o. t. S.-p. (2008) in *Sixth European HIV Drug Resistance Workshop*, Budapest, Hungary).
32. Drummond, A. J., Pybus, O. G., Rambaut, A., Forsberg, R. & Rodrigo, A. (2003) *TRENDS in Ecology and Evolution* **18,** 481-488.
33. Coffin, J. M. (1995) *Science* **267,** 483-9.
34. Leitner, T., Escanilla, D., Franzen, C., Uhlen, M. & Albert, J. (1996) *Proc Natl Acad Sci U S A* **93,** 10864-9.
35. Wallace, R. G., Hodac, H., Lathrop, R. H. & Fitch, W. M. (2007) *Proc Natl Acad Sci U S A* **104,** 4473-8.
36. Nakano, T., Lu, L., Liu, P. & Pybus, O. G. (2004) *J Infect Dis* **190,** 1098-108.
37. Slatkin, M. & Maddison, W. P. (1989) *Genetics* **123,** 603-13.
38. Wensing, A. M. J., Vercauteren, J., van de Vijver, D. A., Albert, J., Asjo, B., Balotta, C., Camacho, R., Coughlan, S., Grossman, Z., Horban, A., Kucherer, C., Nielsen, C., Paraskevis, D., Loke, W. C., Poggensee, G., Puchhammer-Stockl, E., Riva C, Ruiz L, Schmit JC, Schuurman R, Salminen M, Sonnerborg A, Stanojevic M, Struck D, Vandamme AM, Boucher CAB (2008) *Aids* **22,** 625-35.

39. Brenner, B. G., Roger, M., Routy, J. P., Moisi, D., Ntemgwa, M., Matte, C., Baril, J. G., Thomas, R., Rouleau, D., Bruneau, J., Leblanc, R., Legault, M., Tremblay, C., Charest, H. & Wainberg, M. A. (2007) *J Infect Dis* **195,** 951-9.
40. Lewis, F., Hughes, G. J., Rambaut, A., Pozniak, A. & Leigh Brown, A. J. (2008) *PLoS Med* **5,** e50.
41. Hue, S., Pillay, D., Clewley, J. P. & Pybus, O. G. (2005) *Proc Natl Acad Sci U S A* **102,** 4425-9.
42. Hue, S., Clewley, J. P., Cane, P. A. & Pillay, D. (2005) *Aids* **19,** 449-50.
43. Lemey, P., Derdelinckx, I., Rambaut, A., Van Laethem, K., Dumont, S., Vermeulen, S., Van Wijngaerden, E. & Vandamme, A. M. (2005) *J Virol* **79,** 11981-9.
44. Paraskevis, D., Magiorkinis, E., Magiorkinis, G., Kiosses, V. G., Lemey, P., Vandamme, A. M., Rambaut, A. & Hatzakis, A. (2004) *J Mol Evol* **59,** 709-17.
45. Wensing, A. M., van de Vijver, D. A., Angarano, G., Asjo, B., Balotta, C., Boeri, E., Camacho, R., Chaix, M. L., Costagliola, D., De Luca, A., Derdelinckx, I., Grossman, Z., Hamouda, O., Hatzakis, A., Hemmer, R., Hoepelman, A., Horban, A., Korn, K., Kucherer, C., Leitner, T., Loveday, C., MacRae, E., Maljkovic, I., de Mendoza, C., Meyer, L., Nielsen, C., Op de Coul, E. L., Ormaasen, V., Paraskevis, D., Perrin, L., Puchhammer-Stockl, E., Ruiz, L., Salminen, M., Schmit, J. C., Schneider, F., Schuurman, R., Soriano, V., Stanczak, G., Stanojevic, M., Vandamme, A. M., Van Laethem, K., Violin, M., Wilbe, K., Yerly, S., Zazzi, M. & Boucher, C. A. (2005) *J Infect Dis* **192,** 958-66.
46. Weinstock, H. S., Zaidi, I., Heneine, W., Bennett, D., Garcia-Lerma, J. G., Douglas, J. M., Jr., LaLota, M., Dickinson, G., Schwarcz, S., Torian, L., Wendell, D., Paul, S., Goza, G. A., Ruiz, J., Boyett, B. & Kaplan, J. E. (2004) *J Infect Dis* **189,** 2174-80.
47. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res* **22,** 4673-80.
48. Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. (2001) *Science* **294,** 2310-4.
49. Swofford, D. L. (1998) (Sinaeur Associates, Sunderland, Mass).
50. Maddison, W. P., Maddison, D. R. (2003) (Sinaeur Associates, Sunderland, Mass).
51. Sankoff, D., Cedergren, R.J. (1983) in *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison.*, ed. Sankoff, D., Kruskal, J.B. (Addison-Wesley, Reading, Mass), pp. 253-263.
52. Rinsma, I., Hendy, M. & Penny, D. (1990) *Math Biosci* **98,** 201-10.
53. Corbet, S., Muller-Trutwin, M. C., Versmisse, P., Delarue, S., Ayouba, A., Lewis, J., Brunak, S., Martin, P., Brun-Vezinet, F., Simon, F., Barre-Sinoussi, F. & Mauclere, P. (2000) *J Virol* **74,** 529-34.
54. Van Heuverswyn, F., Li, Y., Neel, C., Bailes, E., Keele, B. F., Liu, W., Loul, S., Butel, C., Liegeois, F., Bienvenue, Y., Ngolle, E. M., Sharp, P. M., Shaw, G. M., Delaporte, E., Hahn, B. H. & Peeters, M. (2006) *Nature* **444,** 164.

Figure Legends

Figure 1. This tree contains 8 sequences sampled from 2 countries (A and B). Tips (HIV-1 sequences) were labelled according to its sampling country. **A.** If there are no epidemiological links between the two populations A and B, viral sequences will consist of two monophyletic groups, therefore representing distinct epidemics. **B.** In case that an individual sampled within population B acquired the infection in geographic area A, one branch sampled from population B would cluster within the monophyletic clade of the population A. The migration pattern for each country was estimated by counting "state" (county label) changes at each internal node of the tree by the criterion of parsimony. For each country we counted "exporting" (From) and "importing" (To) migration events. Specifically, as shown in Fig. 1b, a state change (A-B) is counted as an exporting migration event for country A and as importing for B. In our study migration events correspond to mobility of HIV-1 strains or infections and, therefore, inferred exporting or importing migration events are proportional to country-wise mobility of HIV-1 subtype B strains.

Figure 2. Single phylogenetic tree inferred for subtype B sequences sampled from Europe. Branches are shown in different colours and styles by country of origin. Changes in colour represent migration events.

Figure 3. Significant HIV migratory pathways across Europe. Arrowheads indicate the targets of migration shown in different colours and styles by country of origin. Due to space limitations, migration from Luxembourg to Belgium was not shown on the map.

Figure 4. Significant migration events From and To the USA as estimated from the statistical phylogeography analysis.

Supplementary Information

Supplementary information Figure 1. Significant HIV exporting (**A** and **B**) and importing (**C** and **D**) migration events between different countries as estimated by statistical phylogeography study For all countries, 25 sequences were included per analysis, except for the Netherlands (21), Israel (16), and Norway (12). This lower number of sequences explains why the significantly high migration count for these countries is lower than for the other countries. Country code as in table 1.

Supplementary information Figure 2. Total Exporting (**A**), Importing (**B**) and ratio of Importing/Exporting (**C**) migration events across Europe. Country code as in table 1. Migration for Israel, Norway and the Netherlands is not reported in figures A and B for the reason that they are not comparable to the other countries.

Supplementary information Table 1 *Proportion of transmission risk groups among the study population*

Country	Risk groups					Sum
	MSM	**IDUs**	**Heterosexuals**	**Others**	**Unknown**	
United Kingdom (GBR)	47 (63%)	0 (0%)	3 (4%)	0 (0%)	25 (33%)	**75**
Austria (AUT)	1 (2%)	3 (5%)	0 (0%)	0 (0%)	60 (94%)	**64**
Belgium (BEL)	32 (56%)	2 (4%)	7 (12%)	4 (7%)	12 (21%)	**57**
Denmark (DNK)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	72 (0%)	**72**
Spain (ESP)	40 (48%)	18 (21%)	13 (15%)	0 (0%)	13 (15%)	**84**
France (FRA)	74 (74%)	0 (0%)	17 (17%)	0 (0%)	9 (9%)	**100**
Germany (DEU)	48 (91%)	2 (4%)	1 (2%)	0 (0%)	2 (4%)	**53**
Greece (GRC)	9 (36%)	1 (4%)	2 (8%)	0 (0%)	13 (52%)	**25**
Israel (ISR)	6 (29%)	4 (19%)	7 (33%)	1 (5%)	3 (14%)	**21**
Italy (ITA)	24 (29%)	16 (19%)	30 (36%)	0 (0%)	14 (17%)	**84**
Luxembourg (LUX)	55 (55%)	16 (16%)	21 (21%)	0 (0%)	8 (8%)	**100**
Netherlands (NLD)	11 (69%)	1 (6%)	1 (6%)	0 (0%)	3 (19%)	**16**
Norway (NOR)	7 (58%)	1 (8%)	3 (25%)	0 (0%)	1 (8%)	**12**
Poland (POL)	1 (3%)	17 (50%)	2 (6%)	0 (0%)	14 (41%)	**34**
Portugal (PRT)	6 (21%)	4 (14%)	9 (31%)	0 (0%)	10 (34%)	**29**
Sweden (SWE)	30 (40%)	1 (1%)	7 (9%)	0 (0%)	37 (49%)	**75**
Switzerland (CHE)	54 (54%)	10 (10%)	32 (32%)	0 (0%)	4 (4%)	**100**
Sum	445 (44%)	96 (10%)	155 (15%)	5 (0.5%)	300 (30%)	**1001**

Supplementary information Table 2 *Number of calculated migration events between countries.*

To From	GBR	AUT	BEL	DNK	ESP	FRA	DEU	GRC	ISR	ITA	LUX	NLD	NOR	POL	PRT	SWE	CHE	Sun (Fro
GBR		1.04	1.13	1.56	1.16	1.71	1.47	1.37	0.86	1.21	1.16	0.54	0.34	0.67	1.36	1.43	1.43	**18.4**
AUT	0.75		0.48	0.60	0.56	0.43	0.75	0.51	1.01	1.22	0.99	0.66	0.70	0.10	0.46	0.17	0.75	**10.1**
BEL	1.14	0.75		1.06	1.03	1.38	1.55	1.09	1.00	0.82	1.29	0.69	0.60	0.69	1.13	0.79	1.07	**16.0**
DNK	1.00	0.76	0.93		1.15	1.11	1.37	1.01	0.67	0.95	0.94	0.79	0.49	0.70	0.88	1.09	1.02	**14.8**
ESP	1.15	1.03	1.46	1.22		1.83	1.16	0.96	1.00	1.16	1.11	1.20	0.53	0.83	1.08	1.18	1.68	**18.5**
FRA	1.22	0.52	1.36	1.14	1.24		0.96	0.85	0.81	0.65	0.86	0.84	0.59	0.80	1.12	1.16	1.34	**15.4**
DEU	1.06	0.86	1.43	1.30	0.69	1.10		1.10	0.51	1.05	0.87	0.66	0.42	0.56	0.77	0.82	1.13	**14.3**
GRC	1.99	1.00	1.40	1.68	1.41	1.61	1.61		0.79	1.34	1.18	0.99	0.36	0.30	1.86	1.72	1.31	**20.5**
ISR	0.43	0.33	0.75	0.40	0.42	0.38	0.41	0.24		0.64	0.56	0.17	0.15	0.56	0.42	0.47	0.54	**6.8**
ITA	1.04	3.32	1.23	1.54	1.26	1.17	1.35	1.05	1.52		1.05	1.50	0.36	0.13	0.83	0.64	1.98	**19.9**
LUX	0.76	0.44	1.48	0.86	1.08	0.72	1.06	0.90	1.00	0.81		0.66	0.43	0.73	1.39	1.47	0.83	**14.6**
NLD	0.27	0.30	0.34	0.42	0.41	0.49	0.45	0.29	0.15	0.41	0.38		0.44	0.33	0.65	0.46	0.54	**6.3**
NOR	0.18	0.38	0.25	0.27	0.17	0.33	0.32	0.23	0.06	0.10	0.23	0.38		0.33	0.38	0.38	0.22	**4.2**
POL	1.12	0.58	1.61	1.11	1.10	1.42	1.44	0.68	1.71	0.65	1.62	1.11	1.00		1.10	1.12	1.26	**18.6**
PRT	1.07	0.57	1.19	0.68	1.00	0.91	0.74	1.10	1.01	0.68	2.60	0.89	0.43	0.45		0.66	0.85	**14.8**
SWE	1.53	0.42	0.91	1.20	1.21	1.39	1.30	1.53	0.71	1.00	1.63	0.91	0.63	1.00	1.01		1.50	**17.8**
CHE	0.95	0.58	1.01	0.89	1.01	1.06	0.92	0.69	0.75	1.00	0.74	0.65	0.33	0.74	0.82	1.00		**13.1**
Sum (To)	**15.65**	**12.88**	**16.93**	**15.92**	**14.92**	**17.02**	**16.87**	**13.59**	**13.55**	**13.69**	**17.20**	**12.61**	**7.80**	**8.92**	**15.24**	**14.54**	**17.45**	

For all countries, per dataset, 25 sequences were sampled except Israel (ISR), the Netherlands (NLD) and Norway (NOR) for which 21, 16 and 12 sequenc were available, respectively. Migration events are thus directly comparable except for ISR, NLD and NOR. To obtain total migration events per country, t results need to be scaled according to the number of infections per country.

Supplementary information Table 3 *Differences between the observed and the expected migration events. Highlighted cells in yello and cyan denote significantly higher and lower migration numbers compared to panmixis, respectively*

From \ To	GBR	AUT	BEL	DNK	ESP	FRA	DEU	GRC	ISR	ITA	LUX	NLD	NOR	POL	PRT	SWE	CHE
GBR		-0.12	-0.10	0.30	-0.17	0.48	0.15	0.24	-0.15	-0.03	-0.07	-0.25	-0.24	-0.47	0.34	0.21	0.24
AUT	-0.47		-0.67	-0.62	-0.68	-0.85	-0.44	-0.68	-0.06	0.09	-0.12	-0.34	0.02	-0.94	-0.80	-0.93	-0.34
BEL	-0.02	-0.39		-0.25	-0.28	0.12	0.37	-0.05	0.04	-0.37	0.01	-0.14	0.04	-0.55	-0.13	-0.47	-0.14
DNK	-0.33	-0.53	-0.43		-0.07	-0.13	0.13	-0.20	-0.38	-0.26	-0.30	-0.09	-0.04	-0.52	-0.32	-0.13	-0.11
ESP	-0.15	-0.21	0.23	-0.03		0.66	-0.16	-0.24	-0.05	-0.15	0.06	0.41	-0.14	-0.45	-0.28	-0.02	0.31
FRA	-0.17	-0.72	0.05	-0.06	0.09		-0.20	-0.41	-0.17	-0.54	-0.34	-0.04	-0.04	-0.44	0.02	-0.19	0.10
DEU	-0.25	-0.47	0.05	0.02	-0.57	-0.16		-0.14	-0.57	-0.19	-0.47	-0.10	-0.18	-0.81	-0.30	-0.50	-0.19
GRC	0.77	-0.23	0.25	0.35	0.09	0.26	0.36		-0.31	0.00	-0.21	0.25	-0.22	-1.02	0.45	0.31	-0.05
ISR	-0.37	-0.60	-0.04	-0.40	-0.29	-0.48	-0.42	-0.54		-0.29	-0.34	-0.39	-0.26	-0.25	-0.34	-0.40	-0.34
ITA	-0.10	2.16	0.02	0.31	0.08	0.01	0.13	-0.07	0.52		-0.12	0.66	-0.14	-1.07	-0.52	-0.47	0.80
LUX	-0.46	-0.88	0.20	-0.34	-0.04	-0.49	-0.18	-0.34	-0.25	-0.44		-0.17	-0.15	-0.49	0.04	0.27	-0.30
NLD	-0.21	-0.31	-0.17	-0.09	-0.06	-0.01	0.05	-0.05	-0.24	-0.19	-0.12		0.09	-0.24	0.20	0.03	0.04
NOR	-0.08	0.02	-0.08	0.03	-0.10	0.04	0.07	-0.06	-0.15	-0.10	-0.09	0.10		0.12	0.04	0.08	-0.06
POL	-0.25	-0.45	0.33	-0.03	-0.18	0.17	0.19	-0.47	0.66	-0.63	0.45	0.20	0.52		-0.10	-0.22	0.07
PRT	0.00	-0.55	-0.05	-0.46	-0.18	-0.11	-0.27	-0.10	-0.02	-0.64	1.37	0.19	-0.18	-0.85		-0.47	-0.39
SWE	0.30	-0.72	-0.34	-0.15	-0.04	0.08	0.07	0.27	-0.46	-0.21	0.41	0.15	0.05	-0.32	-0.23		0.23
CHE	-0.29	-0.60	-0.21	-0.20	-0.18	-0.21	-0.41	-0.58	-0.31	-0.23	-0.48	-0.15	-0.29	-0.49	-0.48	-0.25	

Negative values indicate less mobility than expected by chance, whereas positive values indicate more mobility than expected. For Norway, The Netherlands and Isra less than 25 strains were included resulting in a lower number of migration events to be expected. In such cases, the panmixis hypothesis test allows to estimate whether not a significantly higher (or lower) number of migration events were observed.

Supplementary information Table 4 *Ranking of Exporting, Importing and ratio of Importing/Exporting migration events*

Country	Exporting migration events	Country	Importing Migration events	Country	*Importing/Exporting*
GRC	20.55	**CHE**	17.45	**NLD**	1.99
ITA	19.96	**LUX**	17.20	**ISR**	1.98
POL	18.62	**FRA**	17.02	**NOR**	1.86
ESP	18.58	**BEL**	16.93	**CHE**	1.33
GBR	18.42	**DEU**	16.87	**AUT**	1.27
SWE	17.87	**DNK**	15.92	**DEU**	1.18
BEL	16.06	**GBR**	15.65	**LUX**	1.18
FRA	15.47	**PRT**	15.24	**FRA**	1.10
DNK	14.84	**ESP**	14.92	**DNK**	1.07
PRT	14.81	**SWE**	14.54	**BEL**	1.05
LUX	14.61	**ITA**	13.69	**PRT**	1.03
DEU	14.31	**GRC**	13.59	**GBR**	0.85
CHE	13.15	**ISR***	13.55	**SWE**	0.81
AUT	10.15	**AUT**	12.88	**ESP**	0.80
ISR*	6.86	**NLD***	12.61	**ITA**	0.69
NLD*	6.33	**POL**	8.92	**GRC**	0.66
NOR*	4.21	**NOR***	7.80	**POL**	0.48

* *Migration figures (Exporting and Importing migration events) estimated for Israel, Norway and the Netherlands cannot be compared to the other countries given that for these countries a smaller number of sequences was included in the analysis*