Phylogenetic analysis as a forensic tool in HIV transmission investigations

Ana B. Abecasis[a,b], Marta Pingarilho[a] and Anne-Mieke Vandamme[a,b]

Because HIV is a fast-evolving virus, HIV genomic sequences of several individuals can be used to investigate whether they belong to a transmission network. Since the infamous 'Florida dentist case' in the beginning of the 1990s, phylogenetic analyses has been recurrently used in court settings as a forensic tool in HIV transmission investigations, for example cases where one or more complainants allege that a defendant has unlawfully infected them with HIV. Such cases can arise both in the context of HIV-specific criminal laws – in countries where transmission of HIV infection is specifically criminalized – or in the context of general laws, for example, by applying physical or sexual assault laws to HIV-related cases. Although phylogenetic analysis as a forensic technique for HIV transmission investigations has become common in several countries, the methodologies have not yet been standardized, sometimes giving rise to unwarranted conclusions. In this literature review, we revisit HIV court case investigations published in the scientific literature, as well as the methodological aspects important for the application and standardization of phylogenetic analyses methods as a forensic tool. Phylogenetic methodologies are improving quickly, such that more recently, phylogenetic relatedness, directionality of transmission and timing of nodes in the tree are used to assess whether the phylogenetic transmission analysis is consistent with or contradicting the charges. We find that there has been a lack of consistency between methods used in court case investigations and that it is essential to define guidelines to be used by phylogenetic forensic experts in HIV transmission cases in court.

AIDS 2018, **32**:543–554

Keywords: court, forensics, HIV, phylogenetics, transmission

Introduction

(1) HIV's fast evolutionary rate allows to study transmission networks of HIV-1 infection.
(2) HIV transmission investigation as a forensic tool happens frequently. It can occur in the context of criminal or civil law, whether or not under an HIV-specific law.
(3) The currently debated hot topics in the context of forensic HIV transmission investigations include whether direction of transmission can be ascertained, whether intermediary links can be excluded and whether it can provide evidence for the timing of HIV transmission.

The genetic divergence of HIV is so large, that the same individual houses a swarm of genetically slightly different viruses: 'quasispecies' [1]. During transmission, one or a few of these virus variants are transmitted and

[a]Global Health and Tropical Medicine (GHTM), Instituto de Higiene e Medicina Tropical/Universidade Nova de Lisboa (IHMT/UNL), Lisbon, Portugal, and [b]Clinical and Epidemiological Virology, Department of Microbiology and Immunology, Rega Institute for Medical Research, Clinical and Epidemiological Virology, KU Leuven – University of Leuven, Leuven, Belgium.

Correspondence to Ana B. Abecasis, PhD, MD, PharmD, Global Health and Tropical Medicine (GHTM), Instituto de Higiene e Medicina Tropical/Universidade Nova de Lisboa (IHMT/UNL), Lisbon, Portugal.

E-mail: ana.abecasis@ihmt.unl.pt
Received: 26 June 2017; revised: 27 November 2017; accepted: 11 December 2017.

DOI:10.1097/QAD.0000000000001728

ISSN 0269-9370

subsequently diverge in the newly infected individual, such that epidemiologically linked individuals never have the same virus [2]. This enormous genetic variation has been successfully used for research at epidemic scale, to study evolutionary history and phylogeography [3,4], and at a local scale, to study transmission networks of HIV-1 patients [5,6].

Phylogenetic inference has become a standard way to characterize HIV-1 transmission networks. This can be used to understand transmission, to aid in the development of better public health prevention initiatives [5,6] or to study transmission of drug resistance [7]. In addition, it has been used in court as part of the evidence in attempts to prove, or to disprove, transmission between defendant and complainant.

Cases of HIV transmission investigations occur under criminal or civil law. Criminal law deals with crime and legal punishment of criminal offenses. Civil law deals with disputes between individuals, between organizations and/or between the two. Under civil law, the outcome primarily involves compensation for injuries or damages, and/or an injunction; under criminal law, a defendant is found innocent or guilty, and if guilty subject to custodial or noncustodial punishment often accompanied by public condemnation via the media and, sometimes, lifetime sex offender registration (in jurisdictions where HIV-related prosecutions are considered sexual assaults). Criminal laws related to this issue vary from country to country. Prosecutions can occur under HIV-specific laws (around 70 countries have HIV-specific criminal laws) or under general criminal or public health laws, such as endangerment, poisoning, physical or sexual assault and even murder. At least 32 countries are known to have applied general or public health laws, and three (Australia, Denmark and United States) have applied both HIV-specific and general criminal laws [8,9]. In criminal cases addressing allegations of HIV-1 transmission, HIV forensics are sometimes relied upon to prove both timing and direction of transmission to show that the defendant infected the complainant and that the defendant was aware of their HIV-positive diagnosis at the time of alleged transmission [10].

DNA profiling technology has been successfully used to link suspects to crime scenes; identify victims of accidents, disasters and wars or exonerate wrongly convicted prisoners, as the human genome remains relatively unchanged during our lifetime. However, for HIV-1 infected individuals, the problem becomes much more challenging, as each patient houses a quasispecies of dynamically evolving viral strains. How different the viruses in two epidemiologically linked individuals are depends on many factors and can still not be reliably predicted. Yet, phylogenetic inference attempts to recapture the molecular epidemiology of the viruses. When a-priori hypotheses can be drawn based on contact tracing, phylogenetic analyses can be used to draw limited conclusions about the epidemiological links between the involved individuals, in the context of the relationship of the viruses with appropriate local controls. Consequently, phylogenetic analysis has been recurrently used as a forensic technique to investigate whether the relationship between viruses infecting a set of individuals is compatible or in contradiction with the virus having passed directly between them [10].

Yet, there are concerns regarding phylogenetic analysis, particularly whether it can indicate direction of transmission, timing of HIV transmission and whether intermediary links can be excluded [10,11]. The establishment of timing of infection, in particular, is highly relevant, as one of the requirements for required state of mind (*mens rea*), for example recklessness, that should be proven, is that infection took place after the defendant knew his/her HIV-positive status. More recent cases have used phylogenetic and population genetics analyses to estimate an approximate time window of infection.

Beyond research purposes, phylogenetic analyses in court involve taking additional, rigorous care regarding what the results can prove in a particular case when a defendant is being prosecuted. In this context, this review puts together courts cases around the world, of crucial importance for HIV forensics. As phylogenetic analyses to investigate transmissions is expanding globally, the most rigorous standards must apply to the appropriate use of those techniques in court. A narrative of all published cases is available in supplementary material (Suppl. 1, http://links.lww.com/QAD/B207). Although HIV transmission investigations in forensics have become very common, the scientific publication of such investigations is decreasing, given the lack of novelty. As such, many of the reported cases are not so recent anymore.

In court, the results of phylogenetic analyses need to be put in the context of other types of evidence

The phylogenetic investigation is only one of the many steps needed to frame what can be concluded from the phylogenetic tree: that is to conclude whether the tree is consistent with or contradicting the charges. Two other very important aspects are determining the window period of infection based on serological or molecular results, and performing thorough contact tracing to trace potential other sources of infection to be included in the analysis. Given that treatment can prevent transmission [12,13], it is also important to determine whether the defendant was infectious at the time of the event. We

could not find data on this type of evidence in the articles discussed, so we cannot comment on how these have been used for the interpretation of phylogenetic results.

One specific single-tree topology is compatible with several alternative transmission scenarios

It is impossible to know for certain that all persons involved in the transmission network have been sampled. Whether links are missing needs to be assessed through contact tracing, and these depend on testimonies of defendant and complainant, and potential other witnesses. Therefore, in reconstructing a transmission history from a phylogenetic tree in the context of forensic investigations, one should never assume that all links are known. Complainants have to recall or be willing to fully disclose all risk contacts [14].

Phylogenetic analysis results should be interpreted with caution, whether using population sequencing, multiple clones or next-generation sequencing (NGS). When using population sequencing, if phylogenetic analysis shows that the two strains under investigation are more related to each other than to control strains, all of the alternative scenarios are also plausible (Fig. 1):

(1) defendant was infected by complainant, not the other way around;
(2) there is a third party with a similar viral strain, linking defendant and complainant;
(3) both complainant and defendant were infected by one or more third parties with similar viral strains;
(4) the complainant was already HIV-positive and was reinfected with another strain, either by the defendant or by a third party [10].

In many cases reviewed herein (Table 1 and Suppl. 1, http://links.lww.com/QAD/B207), monophyletic clustering was significant and court investigations suggested that intermediate contacts could be excluded, whereas no other risk contacts could be identified. How contact tracing was performed was often not described, in part because those that perform contact tracing are not the same as those that perform phylogenetic analysis. Yet, it is often clear that other pieces of evidence were used to support or discard hypothesis of transmission (e.g. timing of seroconversion, physical examination, HIV status of other family members, testimonies of other individuals) [15–18].

Phylogenetics can be used to exonerate individuals

There are examples where the information present in the topology of the tree was used to exonerate individuals. In the nosocomial infection case reported in [19], where two nurses were a potential source of infection of a patient, the virus from nurse 1 clustered significantly in a separate cluster. Phylogenetic analysis was crucial to exclude nurse 1 as source. Also, the topology of the tree reconstructed in [20] was used as evidence against transmission from an HIV-1-infected surgeon to his patient. It seems indeed logical to exonerate defendants that cluster significantly separate from the complainant. However, the longer the time since the events, the higher the chance that onward transmission is separating viruses infecting linked

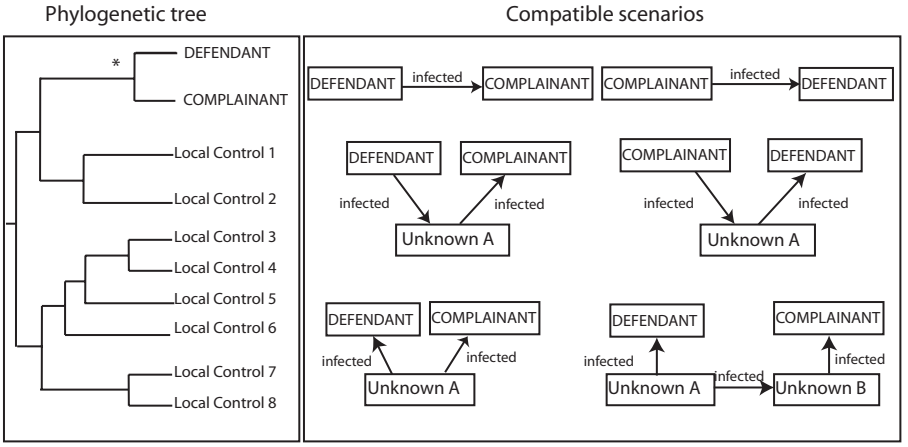

Fig. 1. Hypothetical phylogenetic tree for an investigation of HIV transmission. When the defendant and the complainant form a monophyletic cluster that is significantly supported, alternative scenarios cannot be excluded. Because complete sampling of all patients with HIV (or at least from the epidemiologically relevant population who are local to the parties under investigation) is not feasible, it is impossible to know if (and how many) other individuals belong to the same transmission chain. *Statistically significant support of that cluster.

Table 1. List of transmission investigations where phylogenetic analysis was used as a forensic technique.

Forensic context	Type of case	Contact tracing	Phylogenetic method	Genomic regions	Criteria for selection of controls	Analysis of genomic signatures	Outcome	Sentence	Publication & date
First Florida dentist case	Epidemiological investigation	1100 patients tested (7 found to be infected)	MP	env	Two HIV clinics located within 90 miles from dentist	Yes	Evidence did not allow to discard the hypothesis of transmission	–	Ou *et al.* (1992) [24]
Swedish rape case	Criminal case	NA	ML, MP and NJ	gag, pol	Same geographic region, same risk group	Yes	Evidence did not allow to discard the hypothesis of transmission	Convicted	Albert (1993, 1994) [15,18]
Blood transfusion case	Epidemiological investigation	416 patients tested	Pairwise distance, ML, MP and NJ	gag	10 HIV-1 strains, mainly isolated in the United States	No	Evidence against transmission	–	Holmes (1993) [20]
Second Florida dentist case	Epidemiological investigation	1279 patients tested (28 found to be infected)	NJ	env	Information not available	No	Evidence against transmission	–	Jaffe (1994) Suppl. 1, http://links.lww.com/QAD/B207 – [6]
First Florida dentist case	Epidemiological investigation	1100 patients tested (7 found to be infected)	MP	env	Two HIV clinics located within 90 miles from dentist	Yes	Inconclusive	–	DeBry (1995) [27]
First Florida dentist case	Epidemiological investigation	1100 patients tested (7 found to be infected)	Cladogram estimation + nested analyses	env	Two HIV clinics located within 90 miles from dentist	No	Evidence did not allow to discard the hypothesis of transmission	–	Crandall (1995) Suppl. 1, http://links.lww.com/QAD/B207 – [4]
UK obstetrician/ gynecologist	Epidemiological investigation	Patients upon whom the infected doctor had performed surgical procedures in the preceding 10 years; 475 additional people who asked to be tested after knowing the case	Pairwise distance, ML, NJ	env	Female HIV-1 positive member of the same 'sex circle'; an established male-to-female transmission pair; an established transmission trio; public reference sequences	No	Evidence against transmission	–	Arnold (1995) Suppl. 1, http://links.lww.com/QAD/B207 – [8] Hochuli (1995) Suppl. 1, http://links.lww.com/QAD/B207 – [9]
Glenochil prison outbreak	Epidemiological investigation	Of 378 inmates who were incarcerated during the survey period, 227 (60%) agreed to counseling and 162 (43%) to an HIV test	ML	gag, env	Two previous inmates of that prison who were putatives sources of the outbreak, wife of one of the inmates, unrelated Scottish patients, unrelated subtype B	No	Evidence supporting a prison outbreak derived from a single source	–	Taylor (1995), BMJ [53] Yirrell (1997) [54]
Patient to surgeon infection	Epidemiological investigation	3004 patients who had performed surgical procedures with the surgeon after the moment he was infected	Pairwise distance, NJ, ML	env, gag	Reference sequences of different HIV-1 subtypes	No	Evidence did not allow to discard the hypothesis of transmission	–	Blanchard (1998) Suppl. 1, http://links.lww.com/QAD/B207 – [11]
German sexual abuse case	Criminal case	NA	NJ	gag, env	Information not available	No	Evidence did not allow to discard the hypothesis of transmission	Defendant confessed crime after being confronted with data; no info about court sentence	Banaschack (2000) [16]
Paris surgery case	Epidemiological investigation	2 HIV-positive nurses of the surgical staff	ML and MP	pol, env	HIV-1 infected individuals from the same geographic area	Yes	Evidence against transmission from nurse 1; no evidence against transmission from nurse 2	–	Goujon (2000) [19]
Reckless transmission of HIV in Australia	Criminal case	Sexual contacts of the suspect and contact tracing	Pairwise distance, ML, MP and NJ	gag, env	Patients from the same geographic region without epidemiological link to the suspect	Yes	Evidence did not allow to discard the hypothesis of transmission	Defendant convicted for knowingly and recklessly transmitting HIV	Birch (2000) [21]
Glenochil reckless HIV transmission	Criminal case	This was the same analysis as the Glenochil prison outbreak, used in a second court case	ML	gag, env	This was the same analysis as the Glenochil prison outbreak, used in a second court case	no	Evidence did not allow to discard the hypothesis of transmission	5 years imprisonment	Bird (2001) [56]
Denmark coach case	Criminal case	10 boys trained by the suspect	ML, MP and NJ	gag, pol, env	14 unrelated HIV-1 infected individuals from the same geographic area representing different transmission risk groups	Yes	Evidence did not allow to discard the hypothesis of transmission	Defendant sentenced to 6 years imprisonment	Machuca (2001) [17]

Table 1 (*continued*)

Forensic context	Type of case	Contact tracing	Phylogenetic method	Genomic regions	Criteria for selection of controls	Analysis of genomic signatures	Outcome	Sentence	Publication & date
Louisiana gastroenterologist	Criminal case	All sexual contacts of the defendant	MP, ME, Bayesian	pol, env	32 controls selected from the same local metropolitan area	No	Evidence did not allow to discard the hypothesis of transmission	Defendant convicted to 50 years for attempted second-degree murder	Metzker (2002) Suppl. 1, http://links.lww.com/QAD/B207 – [27]
Sexual assault during professional activities	Criminal case	NA	ML, Bayesian, Statistical Parsimony Networks	env, pol	Controls collected from two local hospitals based on epidemiological criteria geographic area, time of infection, age-matching and risk group + 10 database BLAST controls per complainant	No	Evidence did not allow to discard the hypothesis of transmission	–	Lemey (2005) [28]
Libya foreign medical staff case	Criminal case	NA	ML and Bayesian MCMC, Molecular Clock analysis	gag	Public reference sequences closest to the sequences of the HIV infected children, selected using BLAST	No	Support for long standing nosocomial transmission in the hospital	All defendants freed	de Oliveira (2006) [46]
Washington case and Texas case	Two criminal cases	No other contacts	NJ, ML, Bayesian	pol, env	Local controls and Genbank sequences closest to the sequences of the cases, selected using BLAST	No	Evidence did not allow to discard the hypothesis of transmission	Texas: Defendant charged with intentionally, knowingly and recklessly causing 'serious bodily injury' to 6 female partners; Washington case: Defendant charged with intending to inflict 'great bodily harm' to 17 female partners	Scaduto (2010) [44]
Sex parties in The Netherlands	Criminal case	Individuals attending the private meetings and sex parties organized by the three suspects	NJ, Bayesian	env, pol	Reference sequences representative of the Dutch HIV epidemic, generated from blood plasma of subtype B-infected outpatients visiting the Academic Medical Center in Amsterdam, The Netherlands, between 2005 and 2009	No	Evidence did not allow to discard the hypothesis of transmission	Defendants convicted (not clear from the literature what were the charges)	van der Kuyl (2011) [45]
Shared breastfeeding	Epidemiological investigation	Family of the infant	ML, Bayesian	pol	100 subtype C strains from the same geographical area (Free State Province, South Africa)	No	Evidence did not allow to discard the hypothesis of transmission	–	Goedhals (2012) [29]
Valencian anesthetist	Epidemiological investigation	All patients who had undergone surgery in the 2 hospitals where the suspect worked	NJ, ML	NS5B, E1-E2	44 samples from persons infected with HCV-1a in the city of Valencia who were not related to the outbreak, based on the epidemiological evidence, whose sera had been stored at −80 °C in local hospitals. These samples were used as local controls	No	Evidence did not allow to discard the hypothesis of transmission	Defendant convicted of professional malpractice	González-Candelas F (2013) [47]
Sharing of manicure instruments	Epidemiological investigation	NA	Bayesian, Molecular Clock Analysis	env, pol	Patients from the same geographic region; reference sequences from public databases	Yes (analysis of DRM)	Evidence didn't allow to discard the hypothesis of transmission from the suspected source	–	Matsuda (2014) [48]
Yunnan sexual transmission	Criminal case	NA	Pairwise distance, NJ, Bayesian, recombination profile (bootscanning)	gag, pol, env	Local controls selected from individuals based on epidemiological criteria geographical area, risk group and time of diagnosis	No	Evidence did not allow to discard the hypothesis of transmission	NA	Chen (2015) Suppl. 1, http://links.lww.com/QAD/B207 – [29]

Full narrative of the cases can be found in Supplementary material, http://links.lww.com/QAD/B207, where the case is listed according to type of charge and identified with the forensic context from column 1 mentioned in bold. In the table, the cases are ordered per publication date, oldest cases first. BLAST, basic local assignment search tool; DRM, drug resistance mutation; MCMC, markov chain monte carlo; ML, maximum likelihood; MP, maximum parsimony; NA, not available; NJ, neighbor joining.

individuals, as can be appreciated in [21]. In that case, a time frame of more than 2 years potentially resulted in separation of a direct link by onward transmission.

Potential confounding factors in phylogenetic analyses, such as convergent evolution, should be considered and corrected for in forensic investigations

A confounder that influences topology is convergent or parallel evolution, as illustrated in [22]. When reconstructing a transmission chain of treated patients, Lemey *et al.* found that phylogenetic analysis based on the *pol* gene, was incompatible with the known transmission history. However, after eliminating codon positions associated with drug resistance, the phylogeny became compatible with the transmission history. Contrarily to *pol*, results generated from the gp41 alignment were fully compatible with the transmission history. Indeed, some of the samples had been taken after therapy failure, and the difference between the *pol* and *env* tree could thus be attributed to drug-selective pressure resulting in a pattern of parallel evolution. Since that article, drug-resistance mutations are usually stripped from *pol* alignments before performing phylogenetic analyses of transmission investigations. Other factors potentially leading to convergent evolution include the human leukocyte antigens type of the patients and HIV adaptation to neutralizing antibodies acting on *env* [23,24].

Analyses of HIV-1 transmission clusters need to be put in the context of the local epidemic

An HIV-1 transmission cluster has been defined as a set of HIV-1 sequences that are aggregated nonrandomly, linked to their epidemiology [25]. Even when two individuals are phylogenetically and epidemiologically linked, this evidence on its own cannot prove that person A infected person B. The amount of genetic variation between the HIV strains of the person who transmitted the infection and the person who received it depends on many factors including the time passed since infection, therapy taken, immune pressure and potential subsequent events of superinfection. Therefore, the topology and evolutionary distances of a tree joining defendant and complainant(s) must be placed in context of the epidemic. The inclusion of controls in the analyses, together with anyone identified in contact tracing, and eventually other known linked infections, is critically important to translate the phylogenetic link present in a reconstructed transmission cluster into a meaningful epidemiological link.

The first Florida dentist case already used local controls and contact tracing. However, it was criticized for how controls were selected: they were sampled in clinics that were not close enough to the potential transmission setting [26–28]. Following this case, it became consensual across most of the studies (Table 1) that the set of controls should contain as many local sequences as possible with no known direct link with the case. The criteria for local controls strains were most commonly the same geographic origin, same risk group or same subtype as the complainant and as the defendant, often different sets if these criteria were different between both parties.

More recently, it has become common practice to also include the most similar publicly available sequences selected as database controls, using BLAST search [29,30]. The goal is to find sequences clustering with the query sequence, and since a potential clustering sequence is not necessarily the most similar one, it is customary to add at least 10 most similar sequences per query sequence [31]. This is because the power to exonerate the defendant, if indeed he/she was not the source of the infection, depends on how close the control strains are to the case strains. The closer the control strains, the more likely that the strains from defendant and complainant will get separated in the tree if not linked. However, for most of the cases mentioned (Table 1), this was either not done or not reported (not available).

It is often difficult to obtain relevant controls if no local sequences have been deposited in the public databases. Then forensic scientists have to resort to sampling the local infected population and this poses ethical questions. The establishment of countrywide databases containing sequences and limited anonymized information from all sampled HIV-1 patients is a public health effort in many countries; however, not all consortia are prepared to share their data in part to avoid their use in court cases.

The controls used so far are epidemiologically and temporally relevant to the parties under investigation. In some studies, contact tracing was used to define the set of controls; in others, it was defined based on risk group or geographic or subtype criteria, but most frequently there is no mention of using BLAST searches. Thus, controls included in the analyses were selected using inconsistent criteria.

Analyzing the *pol* genomic region has advantages over using other genomic regions for court investigations

In Table 1, we list genomic regions used for the analyses performed in different court case investigations. Most cases use *env*; some combine the analysis of *env* with that of another genomic region, either *gag* or *pol*. Only one case analyzed three genomic regions (*gag, env* and *pol*) [17]. Leitner *et al.* [32] concluded that the accuracy of the reconstructed tree topology was more dependent on the

amount of genetic information taken into account than on the phylogenetic reconstruction methods used. When investigating the Swedish transmission chain, involving nine patients, he found that the V3 genomic fragment generated more accurate phylogenetic history reconstructions than the p17 fragment. However, the combination of the two dataset V3 + p17 generated even better results [33]. Also Holmes *et al.* [20] established p17 as an epidemiological informative genomic region to be used in transmission cases.

Lemey and Vandamme [34] performed a systematic investigation of the most suitable genetic region to reconstruct three known transmission clusters by performing phylogenetic analysis of genetic fragments extracted from a full-genome alignment of the transmission clusters along with control sequences. The sliding window approach showed that, whereas some clusters were highly supported in some regions, others were not supported at all in the same region. The only consistent result was that the larger the region, the more reliable the reconstruction. *pol* seemed to be particularly consistent for all investigated transmission clusters, provided the fragment used was large enough. However, as described above, when using *pol*, resistance-related positions need to be removed. The analysis of *pol* also provides an advantage for selecting control sequences, as many sequences available from drug resistance testing are from *pol*. However, many of the cases reviewed herein were performed before the era of routine drug resistance testing.

These observations argue in favor of the analysis of two regions to obtain as much information as possible. Also, potential recombination can be assessed.

Different phylogenetic reconstruction methods should be used, as a measure to assess how dependent the results are on the method used

Subtyping, multiple-sequence alignment, phylogenetic reconstruction and population genetic analysis are important steps in forensics investigations. Several alignment algorithms exist [35], yet for HIV transmission investigations, and particularly for *pol,* the level of complexity of the alignments is very low, and all algorithms have been shown to perform well [36]. For phylogenetic reconstruction, distance methods, particularly neighbor-joining, have been shown to perform better than parsimony methods, whereas ML methods perform better than distance methods. In any case, using a correct model to estimate evolutionary distances has been shown important to generate the correct tree [37–39]. Despite being much slower, Bayesian estimation seems to perform better than any other method. Posterior probabilities, however, seem to overestimate the reliability of the clades, whereas bootstrap support is generally considered to be over conservative. However, contradicting previous studies that reported Bayesian estimation as better performing than ML, Wertheim *et al.* [40] found that the ML time-free tree topology was strikingly closer to the true tree than the posterior distribution of Bayesian trees, but comparisons between bootstrapped ML trees and the Bayesian posterior distribution of trees showed they were quantitatively similar [37,40–44]. Pattengale *et al.* [45] analyzed the number of ML bootstrap replicates necessary to reach convergence and found that this number is highly dependent on the dataset. In their analyses of 'stopping criteria', the number of replicates deemed sufficient typically varied between 100 and 500. Similarly, the length of the Markov Chain Monte Carlo runs needed for the Bayesian analysis is highly variable depending on the dataset.

Many of the studies described herein did not use Bayesian approach to phylogenetic estimation, most likely because such investigations were performed when it was still very time consuming. Yet, for most court cases, at least two different methods were used, which allows to circumvent any over or under conservativeness of different phylogenetic approaches. Ideally, all methods are concordant, but if the two most reliable methods, ML and Bayesian, are discordant, then care should be taken when interpreting the results.

Phylogenetic investigation of direction of transmission can only be performed in the context of next-generation sequencing or of multiple clones (or samples) representing the patients' viral quasispecies

In cases where transmission is supported by both phylogenetic and other evidence, additional questions become important, namely direction of transmission, direct vs. indirect and timing of transmission. For these more complex investigations, it is important to rely on sequencing methods more complex than Sanger sequencing, such as NGS or sequencing of multiple clones and/or samples. Direction of transmission, in particular, can only be investigated when multiple viral strains are available from the complainant and the defendant. These give a representation of the quasispecies present in each patient and not only of the predominant viral strain. In addition, direction of transmission is more reliably assessed when multiple samples, and especially early samples, are available. At the same time, any discrepancies between early and late samples could uncover superinfection, if contamination and sample mix-up can be excluded.

In theory, providing molecular evidence for direction of transmission (source to recipient) is possible if a

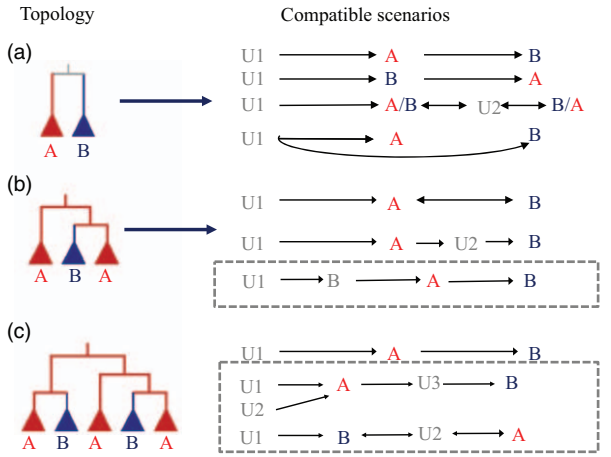

Fig. 2. Classes of topological signal when estimating direction of transmission according to Romero-Severson *et al.* and corresponding compatible scenarios. Compatible scenarios indicated inside dotted boxes have not been described by Romero-Severson *et al.* and have been identified here, based on previously described court cases. When colored gray, these indicate unsampled individuals (U1, U2, U3 and B). When host A (red) is epidemiologically linked to host B (blue), the resulting virus populations upon sampling may relate to each other such that: (a) both populations are monophyletic (MM): If MM relationship is observed, no claims can be made about direction of transmission: A can have infected B or vice-versa, A and B can be linked through an intermediary host U2 (unsampled 2), or both A and B can have been infected by a common source U1 (unsampled 1). (b) one is paraphyletic and the other monophyletic (PM): According to Romero-Severson, if a PM relationship is observed, it is indicative that either host A (paraphyletic) infected host B (monophyletic) or that host A infected an intermediary host U2, which then infected host B. However, we identify yet another possible scenario: B and A have been infecting each other back and forth in a long term relationship, but an earlier sampled virus population of B is lacking and therefore A is paraphyletic to B. (c) one is paraphyletic and the other polyphyletic relative to the first (PP). According to Romero-Severson, if a PP relationship is observed, it is indicative both of direction of transmission and of direct transmission, for example that A infected B, without intermediary hosts involved. However, we identify alternative scenarios, based on a Dutch court case [48]: in a network of individuals, where several individuals have been infecting each other back and forth, for example in sex parties and/or where at least one individual transmits a dual infection to several others. In this case, A could have infected U3 with a dual infection and U3 would then infect B (so, indirect transmission by unsampled U2) or, again B could have been unsampled earlier in time and therefore, B could have infected A and then vice-versa, or there is a complex relationship between U1, A and B and potentially even another U2. Adapted with permission [11].

paraphyletic relationship is observed (Fig. 2). Paraphyly results from a significant genetic bottleneck when establishing productive infection to a recipient, as the majority (>75%) of productive infections derive from a single virus [46]. This argument was used in [47], along with evidence from contact tracing, not only about the directionality of the transmission, but also arguing there was a direct transmission between defendant and several complainants. Similarly, Romero-Severson *et al.* [11] recently provided evidence from simulations that an observed paraphyletic–monophyletic phylogenetic relationship corresponds to a high level of confidence that the directionality of infection is from the paraphyletic to the monophyletic partner, still not excluding a potential third contact in between. It can be argued that this reasoning is not necessarily valid if multiple back and forth transmission events can have occurred, for example when the relationship between defendant and complainant lasted for a considerable period. This is especially true when early samples from complainant but not from defendant are available (Fig. 2b).

It is however possible to exclude one direction of transmission from the establishment of the time window of infection. If the time windows do not overlap, then the one infected later cannot be the source of infection.

One recent modeling study claims that direct transmission can be proven if a paraphyletic–polyphyletic relationship is observed between source and recipient

In addition, a paraphyletic–polyphyletic relationship could indicate not only transmission directionality from the paraphyletic to the nested polyphyletic partner, but also that the transmission occurred directly from source (paraphyletic) to recipient (polyphyletic) [11] (Fig. 2). These findings should however be interpreted with caution, as: usually only one variant establishes the new infection resulting in a paraphyletic–monophyletic relationship, in which case direct transmission cannot be proven; simulations – as the ones used here – require assumptions that are not necessarily true in the context of a court case and the claims were only tested in three different known transmission scenarios involving only two individuals. Cases that end up in court are often not simple, but can involve individuals in a relationship, such that transmission back and forth cannot be excluded. Then directionality in paraphyletic-monophyletic relationship can be reversed, depending on which variants were sampled or disappeared. Similarly, several court cases are known where multiple individuals are involved in a sexual network where virus passes on repeatedly in the network, and various of the individuals could even be superinfected. Furthermore, such findings are only applicable in cases where different clones and/or samples of each patients' viral population are available. This ideal scenario is not available in most court cases described in the literature, and often transmission clusters involve several individuals, even if only two of them are investigated (e.g. [48]).

It is still not possible to confidently estimate the timing of transmission using molecular clock models

Establishing the timing of particular nodes in the tree is highly relevant to infer whether the tree is compatible with infection taking place around the time of the events described in the charges. Time since infection has previously been identified as an important variable in analyses of this type [27]. The determination of the time window of HIV-1 infection should be done from laboratory testing, but can be narrowed down using phylogenetic analysis. This time window is important in a court setting for several reasons:

(1) To know if the defendant was in fact infected by the time of the events described in the charges;
(2) To confirm that complainant was infected around the time of the events described in the charges;

There are three published cases [49–51] where timing of infection was estimated, using similar methodologies: strict or relaxed molecular clock estimation using the BEAST framework [52]. In 1993, it was found that, although assuming a strict molecular clock allows inferring accurate trees for clock-like datasets, when the datasets violate the clock-like assumption, the performance of the tree reconstruction is extremely poor [53]. More recently, Wertheim *et al.* [40] evaluated the accuracy of different clock assumptions on phylogenetic tree estimations and showed that relaxed clock models, particularly the exponential and lognormal relaxed clock models, were more accurate than the strict clock and Bayesian or ML time-free models to infer the correct phylogeny. Importantly, strict clock models consistently estimated less accurate phylogenies than time-free models and assuming a strict molecular clock on nonclock-like data can severely decrease the accuracy of the tree estimation [40]. Similarly, as with direction of transmission, accuracy of estimating a time window of transmission from phylogenetic information improves when multiple clones/samples are available.

With regard to timing of transmission, one needs to be aware that a phylogenetic tree constructed from viral genetic sequences is a gene tree, not a transmission tree [54]. The variant(s) transmitted usually predate the transmission date, and conversely, some transmitted variants might have disappeared from the viral quasispecies after transmission. Therefore, the time to the most recent common ancestor of a clade joining the variants of the recipient with the variants of the donor needs not coincide with the transmission event. This is called the pretransmission interval (Fig. 3).

Discordance between timing of nodes in a tree and the transmission time is elegantly illustrated in a article by

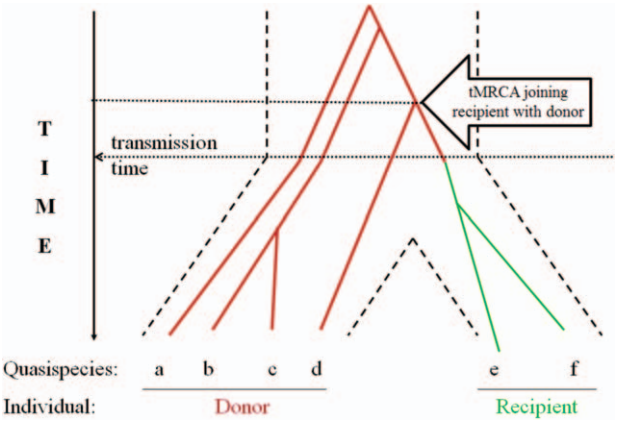

Fig. 3. Pretransmission interval. When genetic sequences are used to build phylogenetic trees, the resulting tree is a gene tree. Such a tree can be used to reconstruct historical events such as a transmission event; however, the timing of the node where donor and recipient join do not necessarily coincide with the transmission event. Even when only one variant was transmitted, the timing of that node may predate the transmission event (adapted with permission from [55]).

Vrancken *et al.* [56], who showed that in the known transmission history under study, ancestral viruses were preferentially transmitted and this led to inconsistencies between the known transmission time interval and the one that was estimated in the genealogy. Paradoxically, in the context of a molecular clock model, a preferential persistence of ancestral viruses strains upon transmission may result in more similar source-recipient lineages than expected based on their transmission time and bias the divergence time estimates toward more recent times [56]. In some of the cases described [49,51], the estimated timing of nodes in the reconstructed phylogenetic trees has been used as an estimate of the window period of infection. Also, Gonzalez-Candelas *et al.* [50] found that the largest inconsistencies found between the estimated timing of infection from the tree and independently derived estimates by the prosecution during the trial corresponded to patients infected in the beginning and in the end of the outbreak, indicating that molecular clock estimates might not be robust enough especially at the extremes of the time distribution of the nodes of the phylogenetic tree.

Estimating the window period of infection based on other tests, such as avidity or ambiguity counting or even on multiassay algorithms is also possible and has been shown to have a reliable performance at population level, it can however not be used at individual level [57]. However, none of the criminal cases described used that type of methodology. The use of NGS in the context of criminal cases can be important for estimating timing of infection: a recent study indicated that genetic diversity calculated from NGS data enables a more accurate estimation of coalescence times, which can give information on time of infection, even many years after

infection. The most precise estimates of coalescence times were obtained using average pairwise distance or site entropy based on third codon positions in the *pol* gene, where viral diversity increases approximately linearly during at least 8 years after infection [58].

The use of public health data in these types of court investigations raises major ethical issues that should be discussed

The Glenochil prison outbreak discussed in Suppl. 1, http://links.lww.com/QAD/B207 [59,60] led to the first prosecution of HIV transmission in Scotland, data and samples from the prison outbreak were seized to build a case against a former inmate, who was found guilty of recklessly transmitting HIV to a female sexual partner [61]. This case raised major ethical issues, as it illustrated how public health data can be used in court and stressed the need for privacy rules. Since that case, some progress has been made in the communication between public health and law official such as to design the best laws that compromise between damage to the individual and damage to public health.

Criminalization of HIV-1 transmission does compromise the general public health when individuals at risk for infection avoid diagnosis, especially as many laws and prosecutions do not need to prove transmission, but rather rely on allegations of nondisclosure of known HIV-positive status or prohibit potential or perceived exposure to HIV during sex [62]. This overly broad use of the criminal law is not recommended by leading global bodies including Joint United Nations Programme on HIV/Acquired Immune Deficiency Syndrome [63] and the Global Commission on HIV and the Law [64].

Now that newer techniques improve the power of phylogenetics to narrow down the source of an infection, contact tracing and phylogenetics are more and more used in public health to track down super-transmitters and treat them to prevent ongoing transmission [65,66]. Public health bodies are very concerned how this will be used in court cases. A recent gathering of the Phylogenetics And Networks for Generalized HIV Epidemics in Africa consortium discussed the ethical issues on how to make cohort sequences public, including avoiding exposure of cohort patients to prosecution [67].

Conclusion

Phylogenetic reconstruction is used in court in the context of other, various, types of evidence. Phylogenetic analysis is therefore only one of the steps needed for the investigation, on its own it can never exclude the presence of hypothetical third parties involved. We reviewed the published cases where phylogenetic analysis was used in the context of forensic HIV transmission investigations. These reflect the past practice. It is difficult to establish from these sources how phylogenetic analysis is currently done in forensic context. Early cases received much more attention than more recent cases, and most of the published cases date from many years ago.

Epidemiological investigation and contact tracing were the initial step of most of the cases reviewed herein (Table 1). Epidemiological investigation gives information on control sequences to be included, and contact tracing through interviews allows to trace other potential sources of infection. When phylogenetic analysis indicates that the two HIV strains under investigation are not related, phylogenetic evidence on its own has been regarded as solid enough to exclude the possibility of direct transmission. Importantly, such investigations have been used to exonerate suspects. On the other hand, when the phylogenetic analysis indicates that the two HIV strains are monophyletic, it has been considered as a piece of evidence that, together with epidemiological data and other types of evidence, contributed to convictions in court.

Significantly, none of the studies listed here reported that nothing could be concluded from the data collected. We do however think that, in some context, where an adequate epidemiological investigation has not been performed, the phylogenetic expert can testify and say that there is not enough information to draw conclusions. An extensive number of cases could not reach a sound conclusion, even when the case strains cluster together, and this is usually because of lack of sufficient evidence and lack of an appropriate epidemiological investigation [47,51].

An additional important aspect is the determination of the window period of infection based on actual serological or molecular results. This is usually not the task of the phylogenetic expert, but recent evidence indicates that timing the most recent common ancestor of nodes in trees contributes to understand the timing of infection. A few cases reported have used this information [49]. Also, with the generalization of NGS methods, assessing direction of transmission and intermediate links may become possible through phylogenetics and used as evidence in court. However, these methods are still debated. Reliability of such estimations, especially in the context of court cases, should therefore be more extensively investigated and validated, before starting to be used routinely in court.

In summary, phylogenetic investigations have been shown useful for the analysis of HIV transmission in forensics and many promising advances in research might empower its use in future cases. At the same time, its use in public health settings is of growing concern to public health bodies, as their best tools to track down sources of

infection to treat them and prevent further infections can be used against these sources in court. This conflict between individual and public health is being discussed widely and will hopefully result in an acceptable balance in future laws.

Acknowledgements

A.B.A. performed the literature review and designed, wrote and reviewed the article. M.P. performed the literature review and wrote the article. A.-M.V. performed the literature review and designed, wrote and reviewed the article.

A.B.A. was supported by Fundação para a Ciência e Tecnologia (FCT) through Investigador FCT Program. M.P. was funded 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) and by FCT (Projecto MigrantHIV; grant PTDC/DTP-EPI/7066/2014). Global Health and Tropical Medicine Center is funded through FCT (UID/Multi/04413/2013). The current work was supported in part by Fonds voor Wetenschappelijk Onderzoek Vlaanderen (grant G.0692.14), by KU Leuven (Program Financing no PF/10/018) and by the VIROGENESIS project which receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 634650.

Conflicts of interest

There are no conflicts of interest.

References

1. Goodenow M, Huet T, Saurin W, Kwok S, Sninsky J, Wain-Hobson S. **HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions.** *J Acquir Immune Defic Syndr* 1989; **2**:344–352.
2. Rambaut A, Posada D, Crandall KA, Holmes EC. **The causes and consequences of HIV evolution.** *Nat Rev Genet* 2004; **5**:52–61.
3. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, *et al*. **HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations.** *Science* 2014; **346**:56–61.
4. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, *et al*. **Unifying the spatial epidemiology and molecular evolution of emerging epidemics.** *Proc Natl Acad Sci U S A* 2012; **109**:15066–15071.
5. Bezemer D, Cori A, Ratmann O, van Sighem A, Hermanides HS, Dutilh BE, *et al*. **Dispersion of the HIV-1 epidemic in men who have sex with men in the Netherlands: a combined mathematical model and phylogenetic analysis.** *PLoS Med* 2015; **12**:e1001898discussion e1001898.
6. Ragonnet-Cronin M, Lycett SJ, Hodcroft EB, Hué S, Fearnhill E, Brown AE, *et al*. **Transmission of non-B HIV subtypes in the United Kingdom is increasingly driven by large non-heterosexual transmission clusters.** *J Infect Dis* 2016; **213**:1410–1418.
7. Pineda-Peña A-C, Schrooten Y, Vinken L, Ferreira F, Li G, Trovão NS, *et al*. **Trends and predictors of transmitted drug resistance (TDR) and clusters with TDR in a local Belgian HIV-1 epidemic.** *PLoS One* 2014; **9**:e101738.
8. Bernard EJ, Cameron S. *Global trends in HIV criminalization [Abstract THPED428]*. Durban: 21st International AIDS Conference; 2016, http://programme.aids2016.org/Abstract/Abstract/7264.
9. Global Network of People Living with HIV (GNP+). *GNP+. 2010 Global Criminalisation Scan Report.*Global Network of People Living with HIV (GNP+); 2010.
10. Bernard EJ, Azad Y, Vandamme AM, Weait M, Geretti AM. **HIV forensics: pitfalls and acceptable standards in the use of phylogenetic analysis as evidence in criminal investigations of HIV transmission.** *HIV Med* 2007; **8**:382–387.
11. Romero-Severson EO, Bulla I, Leitner T. **Phylogenetically resolving epidemiologic linkage.** *Proc Natl Acad Sci U S A* 2016; **113**:2690–2695.
12. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, *et al*. **Prevention of HIV-1 infection with early antiretroviral therapy.** *N Engl J Med* 2011; **365**:493–505.
13. Eshleman SH, Hudelson SE, Redd AD, Wang L, Debes R, Chen YQ, *et al*. **Analysis of genetic linkage of HIV from couples enrolled in the HIV Prevention Trials Network 052 trial.** *J Infect Dis* 2011; **204**:1918–1926.
14. Abecasis AB, Geretti AM, Albert J, Power L, Weait M, Vandamme A-M. **Science in court: the myth of HIV fingerprinting.** *Lancet Infect Dis* 2011; **11**:78–79.
15. Albert J, Wahlberg J, Leitner T, Escanilla D, Uhlén M. **Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 pol and gag genes.** *J Virol* 1994; **68**:5918–5924.
16. Banaschak S, Werwein M, Brinkmann B, Hauber I. **Human immunodeficiency virus type 1 infection after sexual abuse: value of nucleic acid sequence analysis in identifying the offender.** *Clin Infect Dis Off Publ Infect Dis Soc Am* 2000; **31**:1098–1100.
17. Machuca R, Jørgensen LB, Theilade P, Nielsen C. **Molecular investigation of transmission of human immunodeficiency virus type 1 in a criminal case.** *Clin Diagn Lab Immunol* 2001; **8**:884–890.
18. Albert J, Wahlberg J, Uhlén M. **Forensic evidence by DNA sequencing.** *Nature* 1993; **361**:595–596.
19. Goujon CP, Schneider VM, Grofti J, Montigny J, Jeantils V, Astagneau P, *et al*. **Phylogenetic analyses indicate an atypical nurse-to-patient transmission of human immunodeficiency virus type 1.** *J Virol* 2000; **74**:2525–2532.
20. Holmes EC, Zhang LQ, Simmonds P, Rogers AS, Brown AJ. **Molecular investigation of human immunodeficiency virus (HIV) infection in a patient of an HIV-infected surgeon.** *J Infect Dis* 1993; **167**:1411–1414.
21. Birch CJ, McCaw RF, Bulach DM, Revill PA, Carter JT, Tomnay J, *et al*. **Molecular analysis of human immunodeficiency virus strains associated with a case of criminal transmission of the virus.** *J Infect Dis* 2000; **182**:941–944.
22. Lemey P, Derdelinckx I, Rambaut A, Van Laethem K, Dumont S, Vermeulen S, *et al*. **Molecular footprint of drug-selective pressure in a human immunodeficiency virus transmission chain.** *J Virol* 2005; **79**:11981–11989.
23. Holmes EC, Zhang LQ, Simmonds P, Ludlam CA, Brown AJ. **Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient.** *Proc Natl Acad Sci U S A* 1992; **89**:4835–4839.
24. Matthews PC, Leslie AJ, Katzourakis A, Crawford H, Payne R, Prendergast A, *et al*. **HLA footprints on human immunodeficiency virus type 1 are associated with interclade polymorphisms and intraclade phylogenetic clustering.** *J Virol* 2009; **83**:4605–4615.
25. Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjörnsson J. **Defining HIV-1 transmission clusters based on sequence data.** *AIDS Lond Engl* 2017; **31**:1211–1222.
26. Ou CY, Ciesielski CA, Myers G, Bandea CI, Luo CC, Korber BT, *et al*. **Molecular epidemiology of HIV transmission in a dental practice.** *Science* 1992; **256**:1165–1171.
27. Smith TF, Waterman MS. **The continuing case of the Florida dentist.** *Science* 1992; **256**:1155–1156.

28. DeBry RW, Abele LG, Weiss SH, Hill MD, Bouzas M, Lorenzo E, *et al*. **Dental HIV transmission?** *Nature* 1993; **361**:691.
29. Lemey P, Van Dooren S, Van Laethem K, Schrooten Y, Derdelinckx I, Goubau P, *et al*. **Molecular testing of multiple HIV-1 transmissions in a criminal case.** *AIDS Lond Engl* 2005; **19**:1649–1658.
30. Goedhals D, Rossouw I, Hallbauer U, Mamabolo M, de Oliveira T. **The tainted milk of human kindness.** *Lancet Lond Engl* 2012; **380**:702.
31. Paraschiv S, Banica L, Nicolae I, Niculescu I, Abagiu A, Jipa R, *et al*. **Epidemic dispersion of HIV and HCV in a population of co-infected Romanian injecting drug users.** *PLoS One* 2017; **12**:e0185866.
32. Leitner T, Albert J. **Reconstruction of HIV-1 transmission chains for forensic purposes.** *AIDS Rev* 2000; **24**:241–251.
33. Leitner T, Escanilla D, Franzén C, Uhlén M, Albert J. **Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis.** *Proc Natl Acad Sci U S A* 1996; **93**:10864–10869.
34. Lemey P, Vandamme A-M. **Exploring full-genome sequences for phylogenetic support of HIV-1 transmission events.** *AIDS Lond Engl* 2005; **19**:1551–1552.
35. Abecasis A, Vandamme AM, Lemey P. Sequence alignment in HIV computational analysis. In: *HIV Sequence Compendium 2006/2007*. Leitner T, Foley B, Hahn BH, Marx P, McCutchan F, Mellors J, *et al*. (editors). Los Alamos, New Mexico: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory.
36. Bahr A, Thompson JD, Thierry J-C, Poch O. **BAliBASE (Benchmark Alignment dataBASE): enhancements for repeats, transmembrane sequences and circular permutations.** *Nucleic Acids Res* 2001; **29**:323–326.
37. Hall BG. **Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences.** *Mol Biol Evol* 2005; **22**:792–802.
38. Kuhner MK, Felsenstein J. **A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates.** *Mol Biol Evol* 1994; **11**:459–468.
39. Nei M. **Phylogenetic analysis in molecular evolutionary genetics.** *Annu Rev Genet* 1996; **30**:371–403.
40. Wertheim JO, Sanderson MJ, Worobey M, Bjork A. **Relaxed molecular clocks, the bias-variance trade-off, and the quality of phylogenetic inference.** *Syst Biol* 2010; **59**:1–8.
41. Alfaro ME, Zoller S, Lutzoni F. **Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence.** *Mol Biol Evol* 2003; **20**:255–266.
42. Douady CJ, Delsuc F, Boucher Y, Doolittle WF, Douzery EJP. **Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability.** *Mol Biol Evol* 2003; **20**:248–254.
43. Cummings MP, Handley SA, Myers DS, Reed DL, Rokas A, Winka K. **Comparing bootstrap and posterior probability values in the four-taxon case.** *Syst Biol* 2003; **52**:477–487.
44. Erixon P, Svennblad B, Britton T, Oxelman B. **Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics.** *Syst Biol* 2003; **52**:665–673.
45. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. **How many bootstrap replicates are necessary?** *J Comput Biol J Comput Mol Cell Biol* 2010; **17**:337–354.
46. Kariuki SM, Selhorst P, Ariën KK, Dorfman JR. **The HIV-1 transmission bottleneck.** *Retrovirology* 2017; **14**:22.
47. Scaduto DI, Brown JM, Haaland WC, Zwickl DJ, Hillis DM, Metzker ML. **Source identification in two criminal cases using phylogenetic analysis of HIV-1 DNA sequences.** *Proc Natl Acad Sci U S A* 2010; **107**:21242–21247.
48. van der Kuyl AC, Jurriaans S, Back NKT, Sprenger HG, van der Werf TS, Zorgdrager F, *et al*. **Unusual cluster of HIV type 1 dual infections in Groningen, The Netherlands.** *AIDS Res Hum Retroviruses* 2011; **27**:429–433.
49. de Oliveira T, Pybus OG, Rambaut A, Salemi M, Cassol S, Ciccozzi M, *et al*. **Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak.** *Nature* 2006; **444**:836–837.
50. González-Candelas F, Bracho MA, Wróbel B, Moya A. **Molecular evolution in court: analysis of a large hepatitis C virus outbreak from an evolving source.** *BMC Biol* 2013; **11**:76.
51. Matsuda EM, Coelho LPO, Pimentel VF, Onias HB, Brigido LF de M. **An HIV-1 transmission case possibly associated with manicure care.** *AIDS Res Hum Retroviruses* 2014; **30**:1150–1153.
52. Drummond AJ, Suchard MA, Xie D, Rambaut A. **Bayesian phylogenetics with BEAUti and the BEAST 1.7.** *Mol Biol Evol* 2012; **29**:1969–1973.
53. Huelsenbeck J, Hillis DM. **Success of phylogenetic methods in the four-taxon case.** *Syst Biol* 1993; **42**:247–264.
54. Romero-Severson E, Skar H, Bulla I, Albert J, Leitner T. **Timing and order of transmission events is not directly reflected in a pathogen phylogeny.** *Mol Biol Evol* 2014; **31**:2472–2482.
55. Vandamme AM. **Basic concepts of molecular evolution.** In: Lemey P, Salemi M, Vandamme AM,editors. *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. New York: Cambridge University Press; 2009.
56. Vrancken B, Rambaut A, Suchard MA, Drummond A, Baele G, Derdelinckx I, *et al*. **The genealogical population dynamics of HIV-1 in a large transmission chain: bridging within and among host evolutionary rates.** *PLoS Comput Biol* 2014; **10**:e1003505.
57. Cousins MM, Konikoff J, Sabin D, Khaki L, Longosz AF, Laeyendecker O, *et al*. **A comparison of two measures of HIV diversity in multiassay algorithms for HIV incidence estimation.** *PLoS One* 2014; **9**:e101043.
58. Puller V, Neher R, Albert J. **Estimating time of HIV-1 infection from next-generation sequence diversity.** *PLoS Comput Biol* 2017; **13**:e1005775.
59. Taylor A, Goldberg D, Emslie J, Wrench J, Gruer L, Cameron S, *et al*. **Outbreak of HIV infection in a Scottish prison.** *BMJ* 1995; **310**:289–292.
60. Yirrell DL, Robertson P, Goldberg DJ, McMenamin J, Cameron S, Leigh Brown AJ. **Molecular investigation into outbreak of HIV in a Scottish prison.** *BMJ* 1997; **314**:1446–1450.
61. Bird SM, Brown AJ. **Criminalisation of HIV transmission: implications for public health in Scotland.** *BMJ* 2001; **323**:1174–1177.
62. UNAIDS. *Report of the expert meeting on the scientific, medical, legal and human rights aspects of criminalisation of HIV non-disclosure, exposure and transmission, 31 August – 2 September 2011*. UNAIDS; 2012.
63. UNAIDS; UNDP. *Policy brief: criminalization of HIV transmission. Geneva, July 2008; UNAIDS guidance note: ending overly-broad criminalisation of HIV nondisclosure, exposure and transmission: critical scientific, medical and legal considerations*. Geneva: UNAIDS; 2013.
64. Global Commission on HIV and the Law. HIV and the Law: Risks, Rights and Health. Global Commission on HIV and the Law, 2012. https://hivlawcommission.org/.
65. Wertheim F O, Kosakovsky Pond S L, Forgione L A, Mehta S R, Murrell B, Shah S, *et al*. **Social and genetic networks of HIV-1 transmission in New York City.** *PLoS Pathog* 2017; **13**:e1006000.
66. Lin H, He N, Zhou S, Ding Y, Qiu D, Zhang T, *et al*. **Behavioral and molecular tracing of risky sexual contacts in a sample of Chinese HIV-infected men who have sex with men.** *Am J Epidemiol* 2013; **177**:343–350.
67. Cohen J. **Pinpointing HIV spread in Africa poses risks.** *Science* 2017; **356**:568–569.