

# Accidental Father-to-Son HIV-1 Transmission During the Seroconversion Period

Ifeanyi Ezeonwumelu,<sup>1,2</sup> Inês Bártolo,<sup>1</sup> Francisco Martin,<sup>1</sup> Ana Abecasis,<sup>3</sup> Teresa Campos,<sup>4</sup> Ethan O. Romero-Severson,<sup>5</sup> Thomas Leitner,<sup>5</sup> and Nuno Taveira<sup>1,2</sup>

## Abstract

A 4-year-old child born to an HIV-1 seronegative mother was diagnosed with HIV-1, the main risk factor being transmission from the child's father who was seroconverting at the time of the child's birth. In the context of a forensic investigation, we aimed to identify the source of infection of the child and date of the transmission event. Samples were collected from the father and child at two time points about 4 years after the child's birth. Partial segments of three HIV-1 genes (*gag*, *pol*, and *env*) were sequenced and maximum likelihood (ML) and Bayesian methods were used to determine direction and estimate date of transmission. Neutralizing antibodies were determined using a single cycle assay. Bayesian trees displayed a paraphyletic–monophyletic topology in all three genomic regions, with the father's host label at the root, which is consistent with father-to-son transmission. ML trees found similar topologies in *gag* and *pol* and a monophyletic–monophyletic topology in *env*. Analysis of the time of the most recent common ancestor of each HIV-1 gene population indicated that the child was infected shortly after the father. Consistent with the infection history, both father and son developed broad and potent HIV-specific neutralizing antibody responses. In conclusion, the direction of transmission implicated the father as the source of transmission. Transmission occurred during the seroconversion period when the father was unaware of the infection and was likely accidental. This case shows how genetic, phylogenetic, and serological data can contribute for the forensic investigation of HIV transmission.

**Keywords:** HIV father-to-son transmission, forensic investigation of HIV transmission, phylogenetic analysis, neutralizing antibodies

## Introduction

WHILE MOTHER-TO-CHILD (MTC) HIV transmission still occurs in many parts of the world and has been phylogenetically investigated many times, few father-to-child transmissions have been reported. In one case, a 12-year-old daughter was found to be likely infected by her father, although no obvious transmission route was found.<sup>1</sup> In another unusual case, a father apparently infected one of his three children, a 5-year-old boy, while the mother had not been infected.<sup>2</sup> Several modes of non-MTC HIV transmission have been reported, including unsafe needle practices (nosocomial),<sup>3</sup> surrogate breastfeeding,<sup>4</sup> premastication of food,<sup>5</sup> and sexual abuse.<sup>6</sup>

Phylogenetic analysis can aid in source attribution in such cases. In this study, we report a case of an alleged transmission of HIV-1 from a father to his son in Portugal that led to a criminal investigation of sexual abuse since the mother was HIV negative and no other risk factors were identified.

## Methods

### Subjects

Subject CC1 (child) was born on 8th April 2009 to an HIV seronegative mother and was diagnosed with HIV-1 infection when he was 4 years old, in January 2013 (Supplementary

<sup>1</sup>HIV Evolution, Epidemiology and Prevention, Research Institute for Medicines (iMed.Ulisboa), Faculdade de Farmácia, Universidade de Lisboa, Lisboa, Portugal.

<sup>2</sup>Centro de Investigação Interdisciplinar Egas Moniz (CiiEM), Instituto Universitário Egas Moniz (IUEM), Monte de Caparica, Portugal.

<sup>3</sup>Global Health and Tropical Medicine, GHTM, Instituto de Higiene e Medicina Tropical, IHMT, Universidade Nova de Lisboa, UNL, Lisboa, Portugal.

<sup>4</sup>Departamento de Pediatria, Hospital Prof. Dr. Fernando Fonseca EPE, Amadora, Portugal.

<sup>5</sup>Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico.

Fig. S1; Supplementary Data are available online at [www.liebertpub.com/aid](http://www.liebertpub.com/aid)). Antiretroviral therapy (ART) was started in March 2013. Subject CC2 (the boy's father) was diagnosed with HIV-1 infection on the 28th of April 2009. At that time his viral load was 16,000 RNA copies/mL and the number of CD4<sup>+</sup> T lymphocytes was 700 per mm<sup>3</sup>. He had been recently diagnosed with varicella zoster virus infection and syphilis and was undergoing treatment with acyclovir and penicillin. During this time, he developed large vesicles all over his body that leaked fluids profusely. He initiated ART in January 2012 with tenofovir, emtricitabine, and efavirenz, when the viral load was 200,000 RNA copies/mL and CD4<sup>+</sup> T cell number 330 per mm<sup>3</sup>, which led to undetectable viral load. Besides the possibility of sexual abuse, the only risk factor identified for HIV transmission was that the father took care of the umbilical cord during the first days of life, such that transmission could have occurred through contact of the open wound in the umbilical cord with the HIV-1 infected fluids that were oozing profusely from the father's skin vesicles.

Blood samples were collected from both father and son at two different time points, 20th March 2009 and 12th December 2009.

#### *Amplification, cloning, and sequencing*

Genomic DNA was extracted from peripheral blood mononuclear cells using the Wizard Genomic DNA Purification Kit (Promega) as per the manufacturer's instructions. PCR amplification of the p17 (*gag*), integrase (*pol*), and C2 V3 (*env*) regions of HIV-1, cloning, and sequencing were performed as previously described.<sup>7</sup> At least 10 clones per each genomic region were sequenced for each subject. Sequences have been assigned GenBank accession numbers MG273182–MG273261.

#### *Phylogenetic analyses*

Multiple sequence alignment of the derived HIV-1 sequences was performed using MAFFT v7 under the L-INS-i algorithm<sup>8</sup> with manual editing of aligned sequences performed in MEGA 7.0.21.<sup>9</sup> Sequence subtyping was performed using REGA HIV subtyping tool version 3.<sup>10</sup> HIV subtyping indicated the query sequences to be HIV-1 subtype G. Thus, for the database controls (DBC), we retrieved HIV-1 subtype G sequences from the Los Alamos National Laboratory HIV database (LANL HIV DB) using the HIV BLAST tool and the geography (Portugal) search interface ([www.hiv.lanl.gov/content/sequence/HIV/mainpage.html](http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html)). Reference subtypes G and D sequences were also retrieved, where subtype D reference sequences were used as outgroup for rooting of the overall phylogeny.

The sequence pair-wise diversity in each gene fragment from CC1 and CC2 were independently determined for each sampling time (20 March 2013 and 12 December 2013) using a Kimura-2 parameter model as implemented in MEGA 7.0.21.<sup>9</sup>

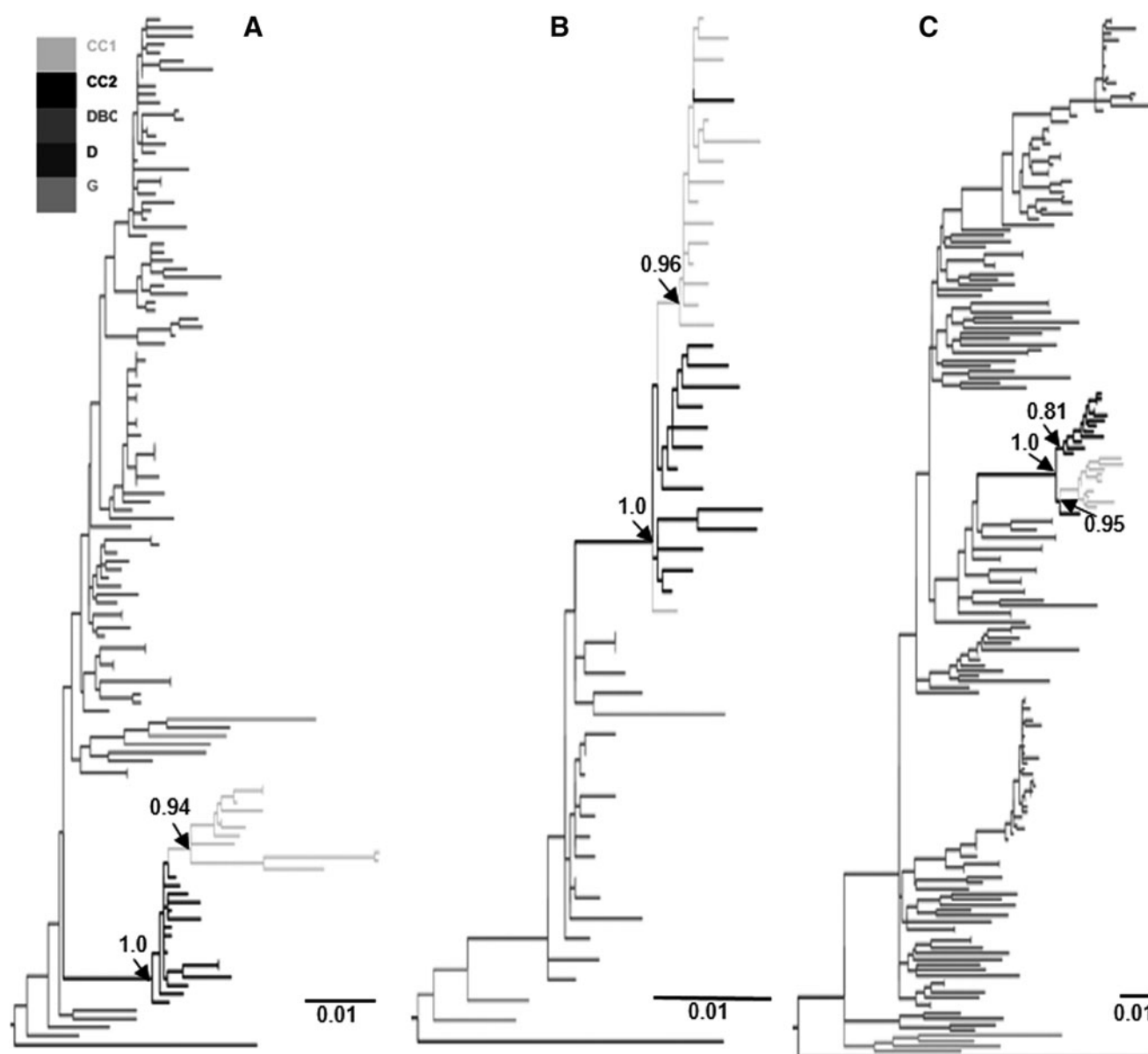
Maximum likelihood (ML) phylogenetic reconstruction was performed using PhyML 3.1<sup>11</sup> as implemented in Sea-view 4.5.4<sup>12</sup> with a bio-neighbor-joining (BioNJ) starting tree, and tree optimization parameters: nearest neighbor interchange and subtree pruning and regrafting heuristic search. Branch supports for the ML trees were inferred based

on the approximate likelihood ratio test (aLRT).<sup>13</sup> Initial Bayesian inference using Markov-chain Monte Carlo (MCMC) sampling as implemented in MrBayes 3.2.6<sup>14</sup> was used to compare trees with the ML results. Two independent runs of four coupled chains per run were performed for  $5 \times 10^6$  generations with trees sampled every 1,000 generations to produce 5,000 posterior tree samples. The burn-in was set at 10% of the initial posterior tree samples, and convergence of chains assumed for ESS values >200 for all the posterior parameters as viewed in Tracer 1.6 (<http://tree.bio.ed.ac.uk/software/tracer>). Figtree was used for the visualization and annotation of phylogenetic trees. Additionally,  $2 \times 30,000$  MrBayes posterior tree samples were obtained for topology hypothesis testing and inference of transmission frequency in the case subjects (CC1 and CC2) for the *env* and *gag* datasets.<sup>7</sup> Appropriate substitution models for the datasets were determined in jModeltest 2.1<sup>15</sup> using 11 substitution schemes and 88 models of substitution. The corrected Akaike information criterion and Bayesian Information criterion scores from the model test were used to select the best-suited substitution models for the inference of the ML and Bayesian trees, respectively.

For the determination of time to the most recent common ancestor (tMRCA) and evolutionary rates of the viral sequences, a Bayesian MCMC approach was performed using BEAST 1.8.4.<sup>16</sup> Prior specifications were set in BEAUti 1.8.4.<sup>16</sup> Analysis of the tMRCA was performed using strict and uncorrelated lognormal relaxed molecular clock models, and logistic growth and skygrid dynamic population size as tree priors. Because the skygrid model failed to describe the dataset, only the logistic tree prior was carried on throughout the analysis with the logistic growth rate fixed at 0.01. Normal distributed priors were specified for the root height with a mean of 4 years for the son and 5 years for the father based on the epidemiological data (Supplementary Fig. S1), with a standard deviation of 2 years to allow for uncertainty and variance in these estimates. Because the mother was HIV negative and could not have infected the son, the prior was truncated at 4.68 years corresponding to the time of his birth after which he could have been infected. Three independent MCMC chains with random seed numbers were run sufficiently long for each dataset to ensure convergence with ESS >200 for all parameters as viewed in Tracer 1.6 (<http://tree.bio.ed.ac.uk/software/tracer>). The log files were combined in LogCombiner 1.8.4 [39] and 10% of the initial posterior MCMC samples were discarded as burn-in. The best molecular clock model was determined based on estimation of a Bayes' Factor from the posterior MCMC samples using a stepping-stone approach<sup>17</sup> with chains run for  $1 \times 10^6$  generations and 50 path steps.

#### *Interpretation of phylogenetic topology*

In a recent work, we showed that the shape (topology) of a phylogeny computed from independent samples from two infected persons is related to how those two persons are related to one another in a transmission network.<sup>18</sup> We determine the tree topology and root label by propagating host tip labels (CC1 and CC2) toward the root until the parent of the current node has a different label. Thus, we find the smallest set of internal nodes in the phylogeny such that all the tips have the same host label. We refer to this part of the tree as a



**FIG. 1.** ML tree reconstructions for *gag* (A), *pol* (B), and *env* (C) viral gene fragments. Clades of viral sequences from query sequences, CC1 (*ash*) and CC2 (*black*), database subtype G controls (DBC) and subtype G reference sequences are colored differently. Subtype D reference sequences, shown in *gray*, were used as outgroup for rooting of the overall phylogeny. Relevant node supports (aLRT) associated with the case sequences are indicated. CC2 and CC1 formed a strongly supported cluster in which a subset of sequences from CC2 were paraphyletic with respect to CC1 forming a paraphyletic–monophyletic (PM) topology in both *gag* (A) and *pol* datasets (B) and a strongly supported monophyletic cluster characterized as monophyletic–monophyletic (MM) topology in the *env* dataset (C). Similar topological relationships were also inferred using a Bayesian statistical approach (Supplementary Fig. S2). aLRT, approximate likelihood ratio test; DBC, database controls; ML, maximum likelihood.

clade. The tree topology in our method is determined by the number of clades for each label and the label at the root. If each host only has one clade, we call this topology monophyletic/monophyletic (MM); if one host has more than one clade and the other only has one—that is, all of the tips of one host are clustered with one another and embedded in the broader tree of the other host—we call this topology paraphyletic/monophyletic (PM); if both hosts have multiple clades, we call this topology polyphyletic/paraphyletic (PP). We have shown that

the PM and PP topologies often arise in direct transmission cases where a direct transmission event forms one (PM) or multiple (PP) clades that are then sampled in the recipient.

#### Selective pressure analysis

The HYPHY package hosted on the Datamonkey open server<sup>19</sup> was used for analysis of selective pressure. The single-likelihood ancestor counting (SLAC), mixed-effects

model of evolution (MEME), fixed-effects likelihood (FEL), and relaxed-effects likelihood (REL) methods were used with statistical significance at  $p < .1$  (SLAC, MEME, and FEL) and a Bayes factor (BF) cutoff value of 50 (REL).

### Antibody neutralization

Plasma antibody neutralization was assessed in TZM-bl cells with seven *env*-pseudoviruses from a panel of global HIV-1 reference isolates as described.<sup>20</sup>

## Results

### Transmission linkage and direction

ML and Bayesian reconstruction inferred similar phylogenetic trees for all datasets with the sequences from CC1 and CC2 forming a strongly supported transmission cluster within subtype G (Fig. 1 and Supplementary Fig. S2). For the *env* dataset, CC1 and CC2 taxa displayed a monophyletic–monophyletic (MM) topological relationship in the ML tree (Fig. 1C). However, the Bayesian posterior probability (pp) for tree topology indicated a paraphyletic–monophyletic (PM) tree topology at  $pp=0.84$  over a MM topology, with CC2 inferred at the root ( $pp=0.82$ ). In addition, based on the *env* dataset, there was a strong support for the transmission of a single lineage of virus between CC2 and CC1 ( $pp=0.97$ ). For the *gag* fragment there was an even stronger evidence for a PM topology ( $pp=0.98$ ) with CC2 again inferred at the root ( $pp=0.98$ ), which was consistent with the inferred ML phylogenetic tree (Fig. 1A). A transmission of a single lineage of HIV-1 virus between CC2 and CC1 was strongly supported for *gag* ( $pp=1.00$ ). For the *pol* dataset, again, there was strong

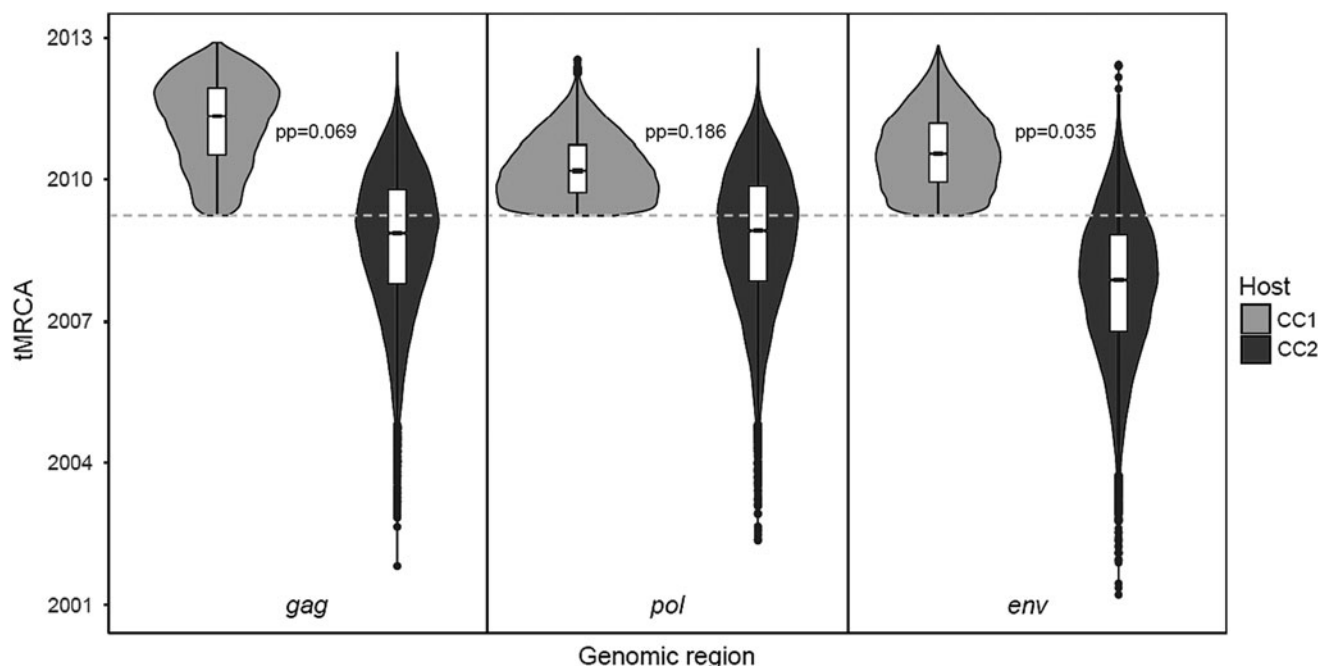
evidence in the ML tree for a PM topology with CC2 paraphyletic with respect to CC1 ( $aLRT=0.97$ ) (Fig. 1B). Altogether, the Bayesian and ML tree analyses support the hypothesis that the father infected his son on a single occasion.

### Evaluation of time of transmission

Assuming that CC2 like CC1 was also infected by a single virus lineage, we can estimate a most recent bound for when each individual was infected by the tMRCA of the individual HIV-1 populations. The estimated tMRCA consistently showed that CC1 most probably became HIV-1 infected at a later date than CC2 for all analyzed genomic fragments (mean tMRCA *gag* CC1=March 2011, CC2=November 2008; *pol* CC1=April 2010, CC2=November 2008, *env* CC1=August 2010, CC2=November 2007) (Fig. 2). Interestingly, CC2's tMRCA agrees well with symptoms of acute HIV-1 infection in early 2009.

### Evolutionary trends and immune pressure assessments

Using an uncorrelated lognormal relaxed clock, the within-host median evolutionary rates were higher in CC1 than CC2 (*env*: 21 and 5.0; *gag*: 24 and 4.1; *pol*: 2.9 and  $2.4 \times 10^{-3}$  median substitutions/site per year, respectively). Evaluating the full posterior distributions of these estimates showed that they differed significantly in *env* and *gag*, but not in *pol* ( $pp=0.013$ ;  $pp=0.012$ ;  $pp=0.65$ , respectively). CC1 had many more codons affected by potential selective forces in *gag* and *pol* relative to CC2; most codons were negatively selected in CC1 and positively selected in CC2 (Supplementary Table S1). In contrast, similar levels of positive



**FIG. 2.** Violin plots of the posterior distribution of estimated dates of the most recent common ancestor (tMRCA) for the *gag*, *pol*, and *env* sequences of CC1 and CC2. Standard boxplots are inserted in each violin plot. The dashed line indicates the birth of CC1, where the normal distributed prior of CC1 was truncated. The posterior probability (pp) in each genomic comparison shows the probability that CC2 was infected after CC1. The tMRCA is given as calendar year. tMRCA, time to the most recent common ancestor.

selection were detected in *env*. Consistent with this antibody selection footprint, plasma samples from CC1 and CC2 had similar high titers of neutralizing antibodies against HIV-1 isolates (Supplementary Fig. S3).

## Discussion

We applied genetic, phylogenetic, and serological analyses to the investigation of an unusual father-to-son HIV-1 transmission case. The initial suspicion of sexual abuse prompted a criminal investigation, which eventually led to no criminal charges. Because of the lack of anal and genital lesions in the child, HIV-1 transmission by rape was excluded. It has previously been shown that the fluid from skin blisters and similar vesicular body fluids can have high load of infectious virions.<sup>21</sup> Moreover, exposure to bleeding skin lesions has been proposed as the most probable transmission route in a similar case of father-to-child HIV-1 transmission.<sup>22</sup> Therefore, we hypothesized that infection of the child might have occurred during the first days of life by accidental contact with the infectious fluid exuding from the father's skin blisters.

The Bayesian phylogenetic analysis showed a PM topology with the father inferred at the root in all three genomic regions. This topology has been strongly associated with direction of transmission, where the root host label typically indicates the donor.<sup>7</sup> Hence, our phylogenetic results are consistent with a father-to-son transmission implicating the father as the child's source of infection. The estimated dates of the tMRCA of the individual HIV-1 populations agreed with the acute HIV-1 infection of the father in early 2009 and infection of the child shortly after birth. However, we were unable to amplify HIV-1 DNA from biopsy samples collected from one skin blister of the father. This negative result may be due to severe nucleic degradation in the paraffin-embedded tissue samples or to the absence of HIV-infected cells in the particular blister we analyzed. Hence, at this time, it is impossible to provide a definitive explanation for the transmission route in this case.

Interestingly, we found that the *env* ML tree showed a MM topology, while *gag* and *pol* ML trees as well as all Bayesian analyses showed PM trees. This iterates the importance of evaluating more than one genomic region, as it is unlikely to have identical sampling artefacts across several regions. In addition, it shows that evaluating a single tree is not enough, even if it is the ML estimate, because many nearly as good trees that also plausibly explain the observed sequence data may differ in their topology. It is thus better to analyze the overall phylogenetic patterns in a full posterior tree sample. However, it should be noted that limited sampling can render PM trees MM and that the paraphyletic signal deteriorates over time especially in rapidly evolving regions such as C2 V3 in *env*.<sup>18</sup> This may explain the MM topology in *env* in our case since the time lag between the putative transmission event and sampling was about 4 years.

Median evolutionary rates in *env* C2 V3, a major antibody neutralizing domain, and *gag*, a target for cytotoxic T lymphocytes, were significantly higher in the child than in the father. This is consistent with the virus adaptation to the developing immune control in the child and with viral suppression due to effective ART in the father.<sup>23</sup> Early initiation of ART in the father may also explain the differences in selective pressure seen in *pol* in both patients. Remarkably, father and

child exhibited a similarly potent and broad neutralizing antibody response providing additional support for a similar infection time, which should be around the birth of the child.<sup>24</sup>

In conclusion, PM phylogeny, root host label, timing analysis, selection analysis, and neutralizing antibody profiling, all supported the father-to-son HIV-1 transmission shortly after the birth of the child when the father was seroconverting and was unaware of his HIV status. Consequently, the case was not taken to court.

## Acknowledgments

Research reported in this publication was supported by the NIAID/NIH under award number R01AI087520, by European Funds through grant 'Bio-Molecular and Epidemiological Surveillance of HIV Transmitted Drug Resistance, Hepatitis Coinfections and Ongoing Transmission Patterns in Europe (BEST HOPE) (project funded through HIVERA: Harmonizing Integrating Vitalizing European Research on HIV/Aids, grant 249697) and by grants PTDC/SAU-EPI/122400/2010, VIH/SAU/0029/2011, and PTDC/DTP-EPI/7066/2014 from Fundação para a Ciência e Tecnologia (FCT), Portugal. Global Health and Tropical Medicine Center was funded through FCT (UID/Multi/04413/2013). Inês Bártole was supported by a postdoc fellowship (SFRH/BPD/76225/2011) from FCT, Portugal. A.A. was supported by Fundação para a Ciência e Tecnologia (FCT) through Investigador FCT Program.

## Author Disclosure Statement

There are no conflicts of interest to report.

## References

- Shao J, Wang J, Abubakar YF, *et al.*: Genetic relatedness of human immunodeficiency virus-1 (HIV-1) strains in a 12-year-old daughter and her father in a household setting. *Arch Virol* 2014;159:1385–1391.
- Murugan S, Anburajan R: Father to child transmission of human immunodeficiency virus disease while sero-discordant status of the mother is maintained. *Indian J Sex Transm Dis* 2013;34:60–61.
- de Oliveira T, Pybus OG, Rambaut A, *et al.*: Molecular Epidemiology: HIV-1 and HCV sequences from Libyan outbreak. *Nature* 2006;444:836–837.
- Goedhals D, Rossouw I, Hallbauer U, Mamabolo M, de Oliveira T: The tainted milk of human kindness. *Lancet* 2012;380:702.
- Ivy WI, Dominguez KL, Rakhmanina NY, *et al.*: Premastication as a route of pediatric HIV transmission: Case-control and cross-sectional investigations. *J Acquir Immune Defic Syndr* 2012;59:207–212.
- 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* 2000;31:1098–1100.
- Romero-Severson EO, Bulla I, Hengartner N, *et al.*: Donor-recipient identification in Para- and Poly-phyletic Trees Under Alternative HIV-1 transmission hypotheses using approximate bayesian computation. *Genetics* 2017;207:1089–1101.
- Katoh K, Standley DM: MAFFT Multiple Sequence Alignment Software version 7: Improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.

9. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;33:1870–1874.
10. Pineda-Peña AC, Faria NR, Imbrechts S, *et al.*: Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. *Infect Genet Evol* 2013;19:337–348.
11. Guindon S, Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003;52:696–704.
12. Gouy M, Guindon S, Gascuel O: SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010;27:221–224.
13. Anisimova M, Gascuel O: Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 2006;55:539–552.
14. Ronquist F, Teslenko M, van der Mark P, *et al.*: MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 2012;61:539–542.
15. Darriba D, Taboada GL, Doallo R, Posada D: jModelTest 2: More models, new heuristics and parallel computing. *Nat Methods* 2012;9:772–772.
16. Drummond AJ, Suchard MA, Xie D, Rambaut A: Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012;29:1969–1973.
17. Xie W, Lewis PO, Fan Y, Kuo L, Chen MH: Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst Biol* 2011;60:150–160.
18. Romero-Severson EO, Bulla I, Leitner T: Phylogenetically resolving epidemiologic linkage. *Proc Natl Acad Sci U S A* 2016;113:2690–2695.
19. Delpont W, Poon AFY, Frost SDW, Kosakovsky Pond SL: Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 2010;26:2455–2457.
20. deCamp A, Hraber P, Bailer RT, *et al.*: Global panel of HIV-1 Env reference strains for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol* 2014;88:2489–2507.
21. Correia O, Delgado L, Santos C, Miranda AM: HIV-1 in blister fluid of a patient with toxic epidermal necrolysis and AIDS. *Lancet* 1994;344:1432–1433.
22. Salvatori F, De Martino M, Galli L, Vierucci A, Chieco-Bianchi L, De Rossi A: Horizontal transmission of human immunodeficiency virus type 1 from father to child. *AIDS Res Hum Retroviruses* 1998;14:1679–1685.
23. Simonetti FR, Kearney MF. Review: Influence of ART on HIV genetics. *Curr Opin HIV AIDS*. 2015;10:49–54.
24. Borrow P, Moody MA: Immunologic characteristics of HIV-infected individuals who make broadly neutralizing antibodies. *Immunol Rev* 2017;275:62–78.

Address correspondence to:

*Nuno Taveira*

*Faculdade de Farmácia*

*HIV Evolution*

*Epidemiology and Prevention*

*Research Institute for Medicines/Instituto*

*de Investigação do Medicamento (iMed.Ulisboa)*

*Universidade de Lisboa*

*Lisboa 1649-003*

*Portugal*

*E-mail: ntaveira@ff.ulisboa.pt*