

RESEARCH ARTICLE

# Deciphering Multiplicity of HIV-1C Infection: Transmission of Closely Related Multiple Viral Lineages

Vlad Novitsky<sup>1,2</sup>, Sikhulile Moyo<sup>2,3</sup>, Rui Wang<sup>1,4</sup>, Simani Gaseitsiwe<sup>2</sup>, M. Essex<sup>1,2\*</sup>

**1** Harvard T. H. Chan School of Public Health, Boston, Massachusetts, United States of America, **2** Botswana Harvard AIDS Institute Partnership, Gaborone, Botswana, **3** Division of Medical Virology, Stellenbosch University, Tygerberg, South Africa, **4** Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, Massachusetts, United States of America

\* [messex@hsph.harvard.edu](mailto:messex@hsph.harvard.edu)



OPEN ACCESS

**Citation:** Novitsky V, Moyo S, Wang R, Gaseitsiwe S, Essex M (2016) Deciphering Multiplicity of HIV-1C Infection: Transmission of Closely Related Multiple Viral Lineages. PLoS ONE 11(11): e0166746. doi:10.1371/journal.pone.0166746

**Editor:** Dimitrios Paraskevis, National and Kapodistrian University of Athens, GREECE

**Received:** August 19, 2016

**Accepted:** November 2, 2016

**Published:** November 28, 2016

**Copyright:** © 2016 Novitsky et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The GenBank accession numbers of the viral sequences used in this study are KC628761–KC630726 and KX644184–KX644757. <http://www.ncbi.nlm.nih.gov/nuccore/>.

**Funding:** The primary HIV-1C infection study in Botswana (Tshedimoso Study) was supported and funded by the NIH/NIAID (R01 AI057027). SM was supported by the Oak Foundation Fellowship (Grant # OUSA-12-025), by the NIH/FIC (D43 TW009610), and by Stellenbosch University, Division of Medical Virology. RW was supported by the NIH/NIAID (R37 AI051164). The funders had no role in study

## Abstract

### Background

A single viral variant is transmitted in the majority of HIV infections. However, about 20% of heterosexually transmitted HIV infections are caused by multiple viral variants. Detection of transmitted HIV variants is not trivial, as it involves analysis of multiple viral sequences representing intra-host HIV-1 quasispecies.

### Methodology

We distinguish two types of multiple virus transmission in HIV infection: (1) HIV transmission from the same source, and (2) transmission from different sources. Viral sequences representing intra-host quasispecies in a longitudinally sampled cohort of 42 individuals with primary HIV-1C infection in Botswana were generated by single-genome amplification and sequencing and spanned the V1C5 region of HIV-1C *env* gp120. The Maximum Likelihood phylogeny and distribution of pairwise raw distances were assessed at each sampling time point ( $n = 217$ ; 42 patients; median 5 (IQR: 4–6) time points per patient, range 2–12 time points per patient).

### Results

Transmission of multiple viral variants from the same source (likely from the partner with established HIV infection) was found in 9 out of 42 individuals (21%; 95 CI 10–37%). HIV super-infection was identified in 2 patients (5%; 95% CI 1–17%) with an estimated rate of 3.9 per 100 person-years. Transmission of multiple viruses combined with HIV super-infection at a later time point was observed in one individual.

### Conclusions

Multiple HIV lineages transmitted from the same source produce a monophyletic clade in the inferred phylogenetic tree. Such a clade has transiently distinct sub-clusters in the early stage of HIV infection, and follows a predictable evolutionary pathway. Over time, the gap

design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

between initially distinct viral lineages fills in and initially distinct sub-clusters converge. Identification of cases with transmission of multiple viral lineages from the same source needs to be taken into account in cross-sectional estimation of HIV recency in epidemiological and population studies.

## Introduction

The majority of HIV infections are associated with transmission of a single founder virus, with transmission of multiple HIV-1 lineages occurring in about 20% of heterosexual cases [1–9]. Multivariant HIV transmission is higher in men who have sex with men (MSM) and injection drug users, reaching about 30–40% [10–13], although no difference in multiplicity of HIV transmission between modes of HIV transmission was also reported [14]. Multiple HIV lineages could be transmitted at a single encounter, or on multiple occasions over the course of HIV infection. The latter scenario is known as an HIV super-infection [7, 15–25]. In the HPTN 052 study, analysis of transmitted HIV's helped to distinguish linked and unlinked viral transmissions and solidify the study findings [26–29]. The multiplicity of breakthrough HIV infections can be an important outcome in vaccine trials [11].

Identification of the multiplicity of virus transmission in HIV infection is challenging because it requires multiple viral sequences representing intra-host HIV quasispecies. Molecular techniques, such as single-genome amplification and sequencing, or next-generation sequencing, can be applied to address multiplicity of HIV transmission, which remains a subject of special studies.

The term *multiplicity* of virus transmission in the context of HIV infection has not been well defined with the exception of two extreme scenarios, transmission of a single founder virus and HIV super-infection. High homogeneity of viral quasispecies soon after infection is associated with the effective transmission of a single HIV variant (which does not exclude transmission of multiple but undetected, or extinguished, variants). Similarly, distinct viral lineages separated by other patients' sequences in the phylogenetic tree provide compelling evidence for transmission of multiple HIV variants, often as a super-infection. However, the interpretation of intermediate scenarios remains uncertain, as well as thresholds and criteria for multiplicity of HIV transmission. Technically, even a single nucleotide difference between identified intra-host HIV quasispecies could be interpreted as transmission of *multiple* viral variants. However, the clinical or epidemiological relevance of transmitted HIV quasispecies with minor differences is still unclear.

Multivariant HIV infection has been associated with elevated HIV-1 RNA set point [22, 25, 30–33] and faster disease progression [34–37], but has been reported to have limited impact on the occurrence of clinical events [22].

In this study we focus on transmission of multiple HIV lineages from the same source. The goal of the study was to identify transmission of multiple virus lineages from the same source based on the inferred phylogeny and distribution of viral pairwise distances of viral sequences representing intra-host HIV quasispecies. Better understanding of HIV transmission and the ability to distinguish between transmissions of multiple virus variants from a single source and those from multiple sources should assist in the analysis of HIV transmission networks and their dynamics.

## Materials and Methods

### Ethics statement

The study on primary HIV-1C infection in Botswana, the *Tshedimoso* study [4, 38–45], was conducted according to the principles expressed in the Declaration of Helsinki. The study was approved by the Health Research and Development Committee (HRDC) of the Republic of Botswana, and the Office of Human Research Administration (OHRA) of the Harvard T.H. Chan School of Public Health. All adult study subjects provided written informed consent for participation in the study; all minor study subjects provided written informed assent, and each minor's guardian provided written informed consent, for their participation in the study.

### HIV-1C sequences

Viral sequences were generated within the *Tshedimoso* study of primary HIV-1C infection in Botswana [4, 38–45]. Briefly, 42 individuals with primary HIV-1C infection (including 8 acute and 34 recent cases) were longitudinally sampled over a period of about 500 days post-seroconversion. Viral sequences were generated by single-genome amplification and sequencing and spanned the V1C5 region of HIV-1C *env* gp120, about 1,200 bp in length (HXB2 nucleotide positions 6,615 to 7,757). The initial set of viral sequences included 225 time points; eight time points with fewer than four sequences each were excluded (16 sequences total). The total number of 217 time points analyzed in this study represented 42 patients, median 5 (IQR: 4–6) time points per patient, range from 2 to 12 time points per patient. The analyzed time points were represented by a total of 2,524 sequences, approximately 12 sequences per patient per time point. Participant characteristics are described elsewhere [40, 43, 44]. All participants were infected with HIV-1C, and were predominantly female (76%), with a median age of 27 (IQR 25–33) years at enrollment. Both viral RNA and proviral DNA were used as templates for amplification and sequencing. The GenBank accession numbers of the viral sequences used in this study are KC628761–KC630726 and KX644184–KX644757.

### Multiple sequence alignment

Codon-based multiple sequence alignment of viral sequences was performed by Muscle [46, 47] with default setting for gap penalty and gap extension. Minor manual adjustments across the multiple sequence alignment were performed in BioEdit [48].

### Phylogenetic analysis

The Maximum Likelihood (ML) phylogeny was reconstructed for each time point of sampling ( $n = 217$ ; 42 patients; median 5 (IQR: 4–6) time points per patient, range from 2 to 12 time points per patient) using PhyML [49, 50]. The resulting phylogeny was visualized in SeaView [51].

### Pairwise distances

The distribution of pairwise raw distances of viral sequences per subject per time point was analyzed by `dist.dna` (`ape` [52] package in R) using multiple sequence nucleotide alignment.

### Testing for sub-clusters

The goal of this analysis was to identify (or reject) the presence of potential sub-clusters within each set of viral sequences representing intra-host HIV-1C quasispecies at a single time point. In this paper the term “sub-clusters” indicates clusters within the pool of HIV sequences

representing intra-patient viral quasispecies. Sub-clusters were defined by a specific topology in the inferred ML phylogenetic tree: presence of a monophyletic patient-specific lineage with sub-clusters, which was evident by a combination of relatively long branches separating sub-clusters of viral sequences and short branches within each cluster. Such a topology was considered to be associated with transmission of multiple viral variants from the same (or a closely related) source of presumably established (chronic) HIV infection.

To standardize identification of sub-clusters based on phylogeny of viral sequences representing intra-host HIV-1C quasispecies, we developed a simple test using R packages *ape* [52] and *stats* [53]. The pairwise distance matrix was generated by *dist.dna* (*ape* [52] R package). To identify (or reject) potential sub-clusters, *kmeans* (*stat* R package) was utilized to partition the pairwise distance matrix into two groups ( $k = 2$ ). The ratio of *withinss* (vector of within-cluster sum of squares, one per cluster) to *betweenss* (the between-cluster sum of squares) was used to determine the validity of partitioning. The clustering was considered valid if the ratio values (*withinss* to *betweenss*) for both sub-clusters were greater than zero and less than a particular threshold. To inform the choice of the threshold, we performed simulation studies to calculate the sensitivity, specificity and predictive values for the ratio *withinss* to *betweenss* thresholds set at 0.1, 0.15, 0.20, 0.25 and 0.30 using the R package *caret* [54]. The clustering estimates at different ratio thresholds were compared with the reference data. The reference data with and without sub-clusters were generated by evaluation of ML phylogeny and distribution of pairwise distances for 217 time points. For each threshold examined, sensitivity was defined as the proportion of clustered cases with this threshold out of the number of clustered cases in the reference data. Specificity was defined as the proportion of non-clustered cases with the specified threshold out of the number of non-clustered cases in the reference data. Positive predictive value was defined as the proportion of predicted time points with sub-clusters out of clustered reference data:  $\text{sensitivity} * \text{Prevalence} / ((\text{sensitivity} * \text{Prevalence}) + ((1 - \text{specificity}) * (1 - \text{Prevalence})))$ . Negative predictive value was defined as the proportion of non-clustered time points out of non-clustered reference data:  $(\text{specificity} * (1 - \text{Prevalence})) / (((1 - \text{sensitivity}) * \text{Prevalence}) + ((\text{specificity}) * (1 - \text{Prevalence})))$ . The sensitivity and specificity of different values that were estimated from the simulation studies are presented in Table 1.

Based on the results of simulation studies, the value of 0.20 has been chosen as the threshold for ratio *withinss* to *betweenss*, although the value of 0.15 could also be considered. In this study, within each set of viral sequences representing intra-host HIV-1C quasispecies at a given time point, sub-clusters were considered present if the ratio values for both sub-clusters were greater than zero and less than 0.20.

### Statistical analysis

The statistical analysis was performed in R version 3.3.1 [53]. The proportions and the associated 95% confidence intervals (CIs) of transmitted multiple viral variants were estimated based on binomial distributions (*prop.test()* in R). McNemar’s test [55] was used to compare

**Table 1. Simulation studies for the ratio *withinss* to *betweenss* threshold values.**

Threshold values	Sensitivity	Specificity	Positive predictive value	Negative predictive value
0.10	0.989	0.667	0.949	0.909
0.15	0.989	0.700	0.954	0.913
0.20	0.977	0.767	0.963	0.852
0.25	0.947	0.767	0.962	0.697
0.30	0.909	0.767	0.960	0.575

doi:10.1371/journal.pone.0166746.t001

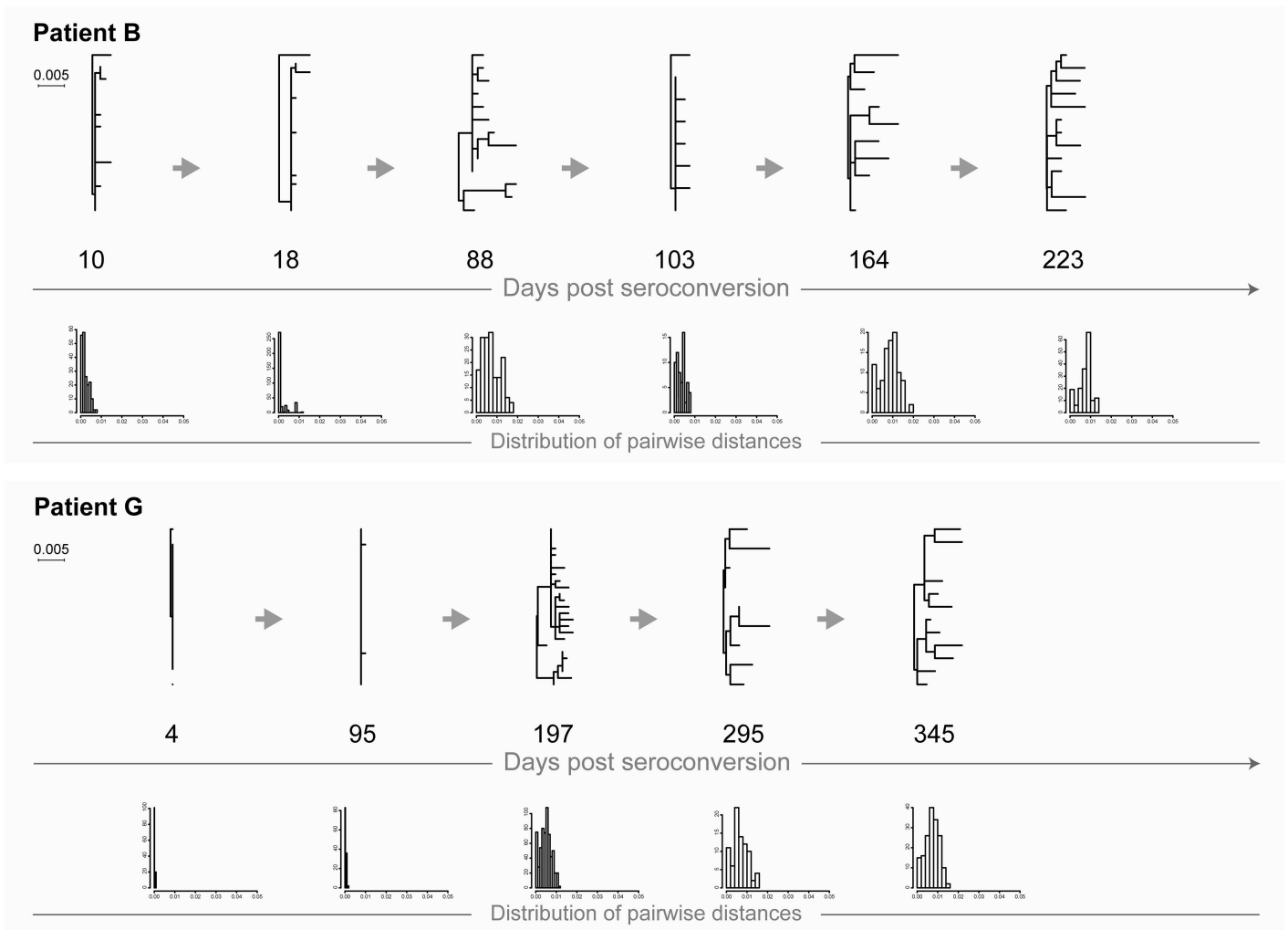
the proportions of two dichotomous traits from the same group of subjects. The rate of HIV super-infection was estimated by using the participants' maximum follow-up time and was expressed as the number of events per 100 person-years. P-values less than 0.05 were considered statistically significant. The reported p-values are 2-sided.

## Results

HIV-1C evolutionary dynamics are exemplified by the following four scenarios: (1) transmission of a single founder virus, (2) transmission of multiple viruses from the same source partner, (3) super-infection, and (4) transmission of two founder viruses followed by an HIV super-infection. The inferred ML trees and distribution of pairwise distances are presented at the specified time points of sampling expressed as days post-seroconversion.

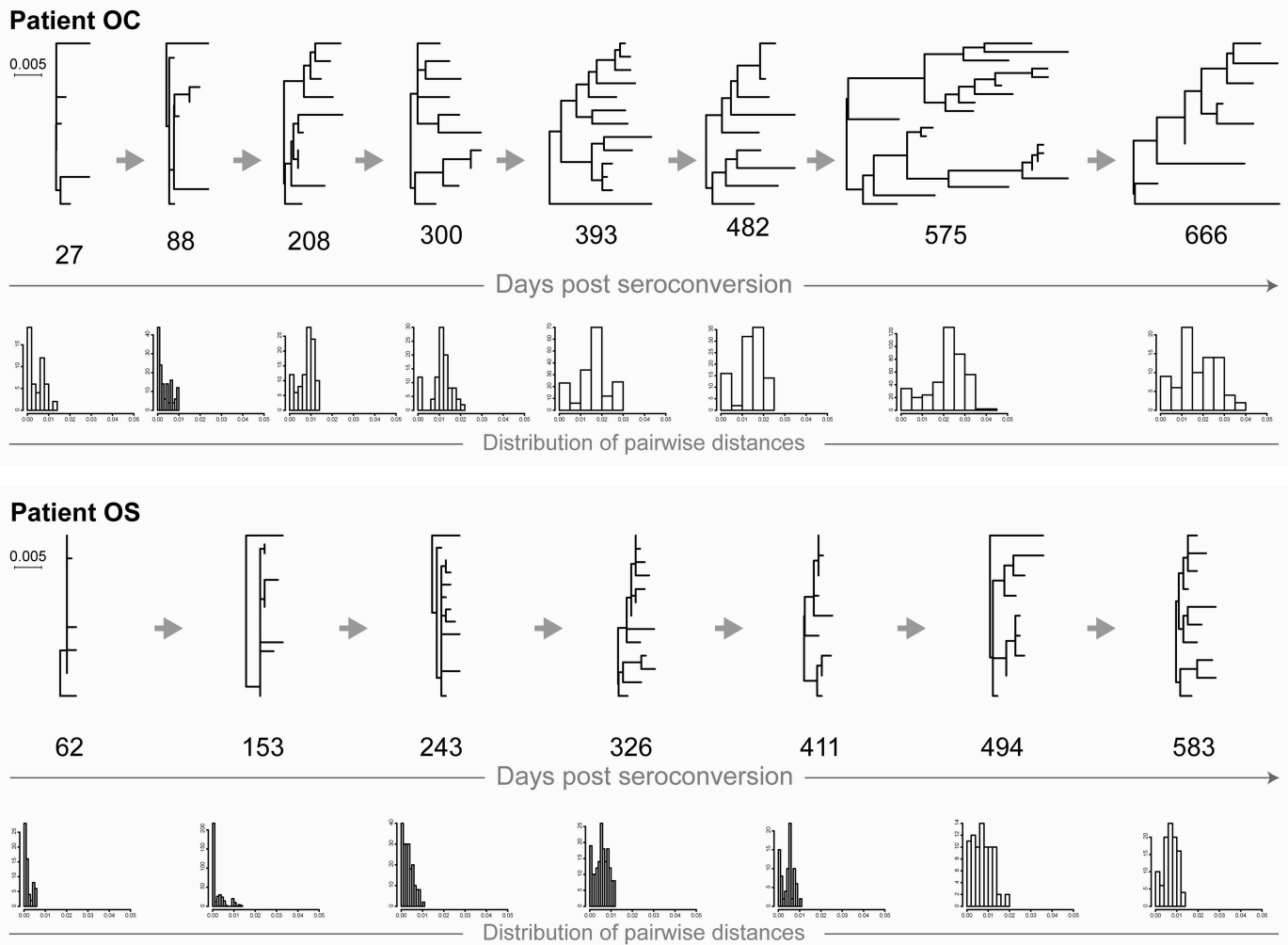
### Transmission of a single founder virus

Extreme homogeneity of viral quaspecies in most cases is a hallmark of transmission of a single viral variant (Fig 1: patients B and G; Fig 2: patients OC and OS). It is possible that



**Fig 1. HIV transmission of single viral variants in acutely infected patients B and G.** Maximum likelihood trees inferred by PhyML and distribution of raw pairwise distances are shown. Numbers below each phylogenetic tree indicate time of sampling in days post-seroconversion. Both phylogenetic trees and histograms of pairwise distances are drawn to the patient-specific scale shown at left for each patient.

doi:10.1371/journal.pone.0166746.g001



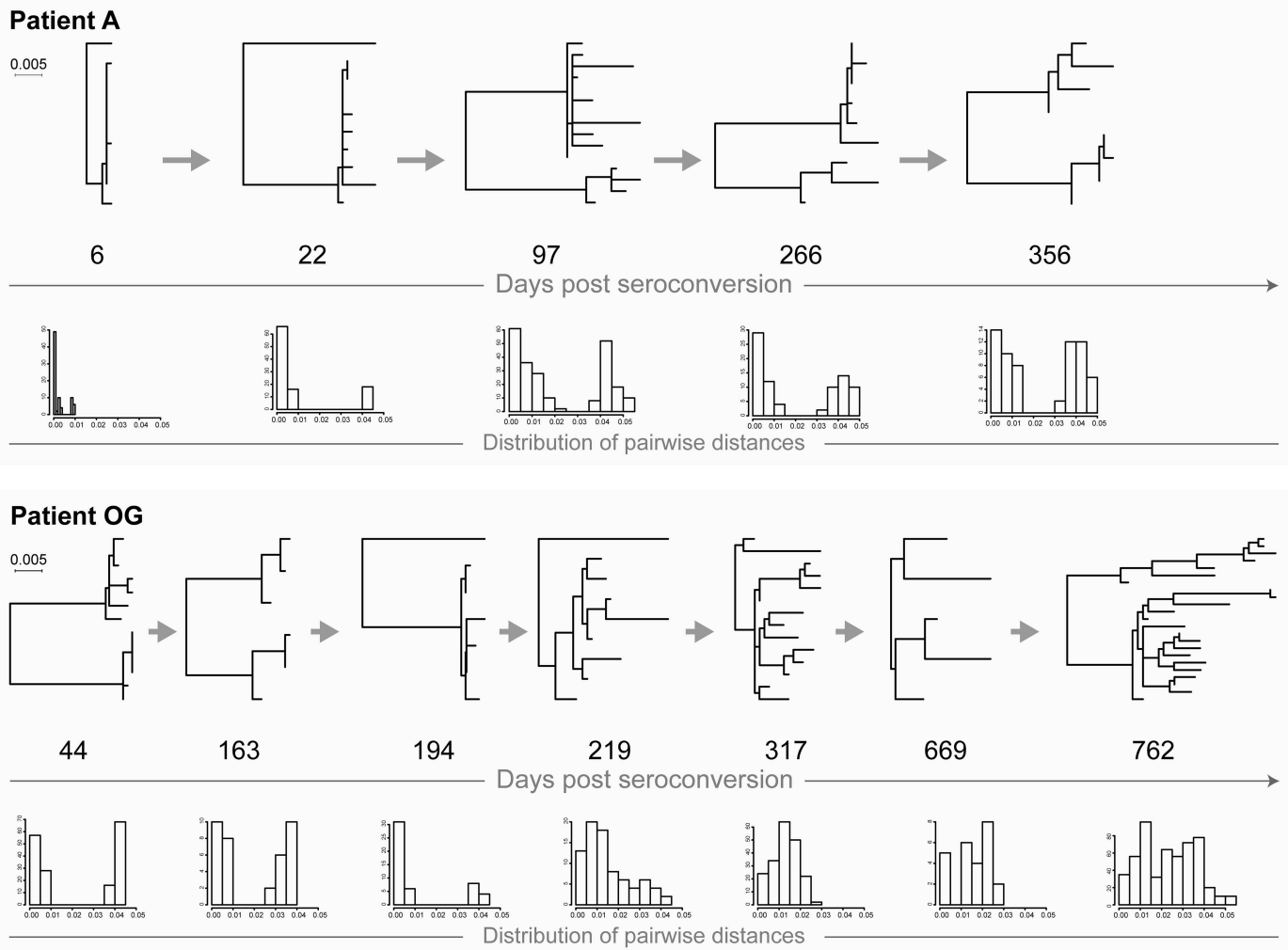
**Fig 2. HIV transmission of single viral variants in recently infected patients OC and OS.** For explanation of phylogenetic trees, pairwise distances, and time points of sampling, please see Fig 1 legend.

doi:10.1371/journal.pone.0166746.g002

homogeneity of viral quasispecies could also occur during transmission of multiple viruses followed by a single variant outgrowth. Viral diversity increases gradually over time (with or without fluctuations) which is evident from the increasing branch lengths in the phylogenetic tree. The close to normal distribution of pairwise distances reflects gradual increase of viral diversity over time. The majority of HIV infections follow this scenario.

### Transmission of multiple (i.e., two) viruses from the same source

This is an uncommon scenario of HIV transmission that is made evident by the specifics of the tree topology and the distribution of pairwise distances. In the phylogenetic tree, viral sequences representing intra-host HIV quasispecies form and can be identified as a distinct monophyletic clade with specific structure. At the earlier stage (e.g., within weeks or a few months after HIV transmission and seroconversion), the structure of the monophyletic clade includes multiple (i.e., two) sub-clusters with relatively low diversity within each sub-cluster (Fig 3: patient A at days 22 and 97, and patient OG up to day 194; Fig 4: patient D up to days 301/393, and patient PK up to days 108/135). The corresponding distribution of pairwise



**Fig 3. HIV transmission of multiple viral variants in patients A (acute HIV infection) and OG (recent infection).** For explanation of phylogenetic trees, pairwise distances, and time points of sampling, please see Fig 1 legend.

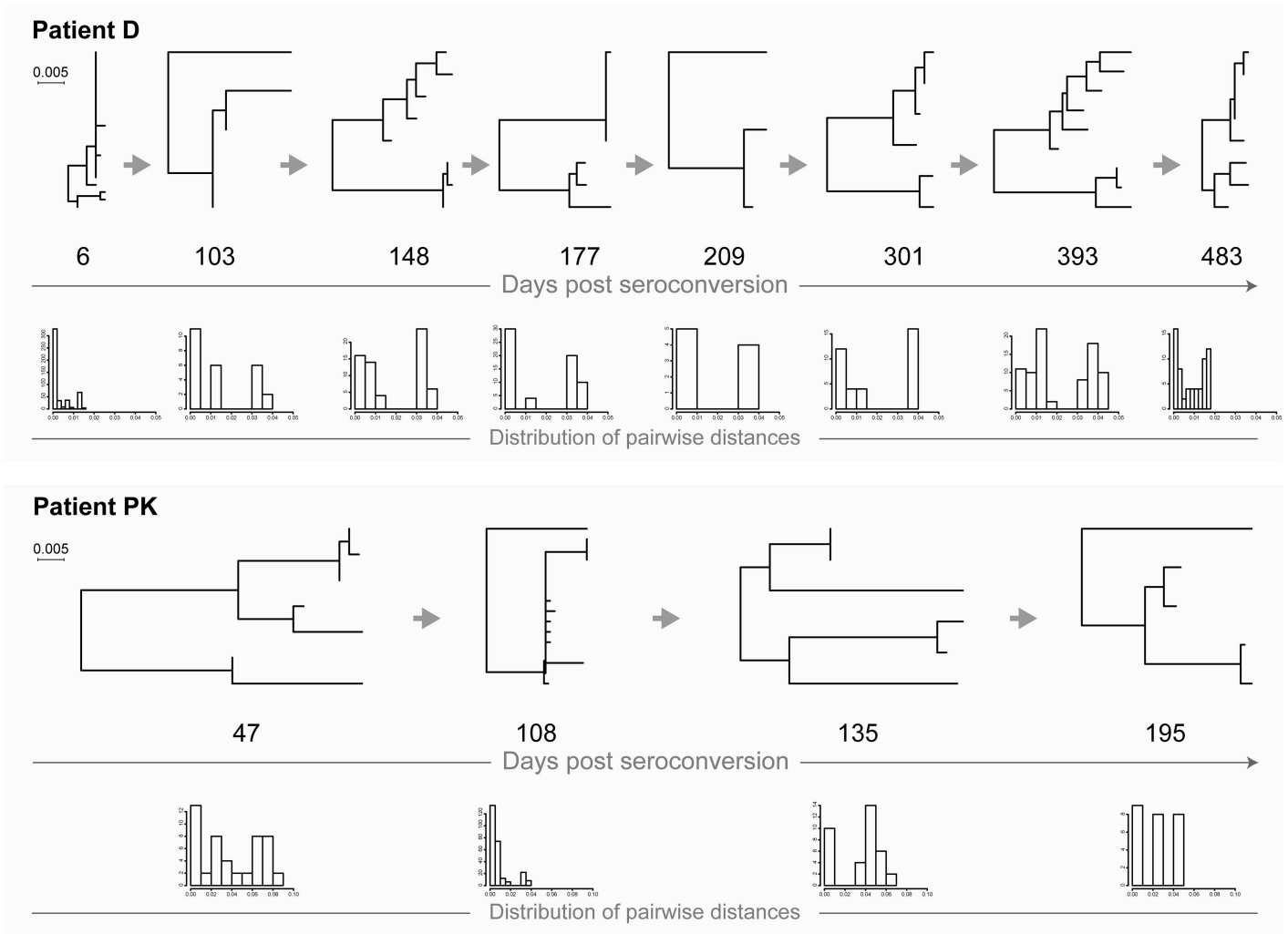
doi:10.1371/journal.pone.0166746.g003

distances is characterized by two distinct peaks on the histogram indicating low levels of diversity within each sub-cluster and sizable pairwise diversity associated with pairwise distances between sub-clusters.

The typical tree topology upon transmission of multiple viral variants from the same source is transient, and therefore can be easily overlooked. The distinction between sub-clusters disappears over time, apparently due to *de novo* generated recombinants that can fill the gap between sub-clusters in the phylogenetic tree (as shown in our previous analysis [42]) and convergence of distinct peaks in the histogram with pairwise distances (Fig 3: patient OG at day 219 and later time points; Fig 4: patient D at day 483, and patient PK at day 195). Note that in patient A (Fig 3), virus sequences did not close the gap between sub-clusters by day 356.

### HIV super-infection

Transmission of distinct HIVs can occur over the course of HIV infection. In contrast to a monophyletic clade, viral sequences representing super-infection fall into different parts of the



**Fig 4. HIV transmission of multiple viral variants in patients D (acute HIV infection) and PK (recent infection).** For explanation of phylogenetic trees, pairwise distances, and time points of sampling, please see Fig 1 legend.

doi:10.1371/journal.pone.0166746.g004

inferred phylogenetic tree and are separated by HIV sequences from other patients. Upon HIV super-infection, long branches separate distinct viral variants. Multiple (i.e., two) peaks can be found on the histogram of pairwise distances reflecting distances within and between two viral variants (Fig 5: patient NA at days 231, 354, and 448). Note that at earlier time points, days 79 and 169, patient NA was infected with a single viral variant.

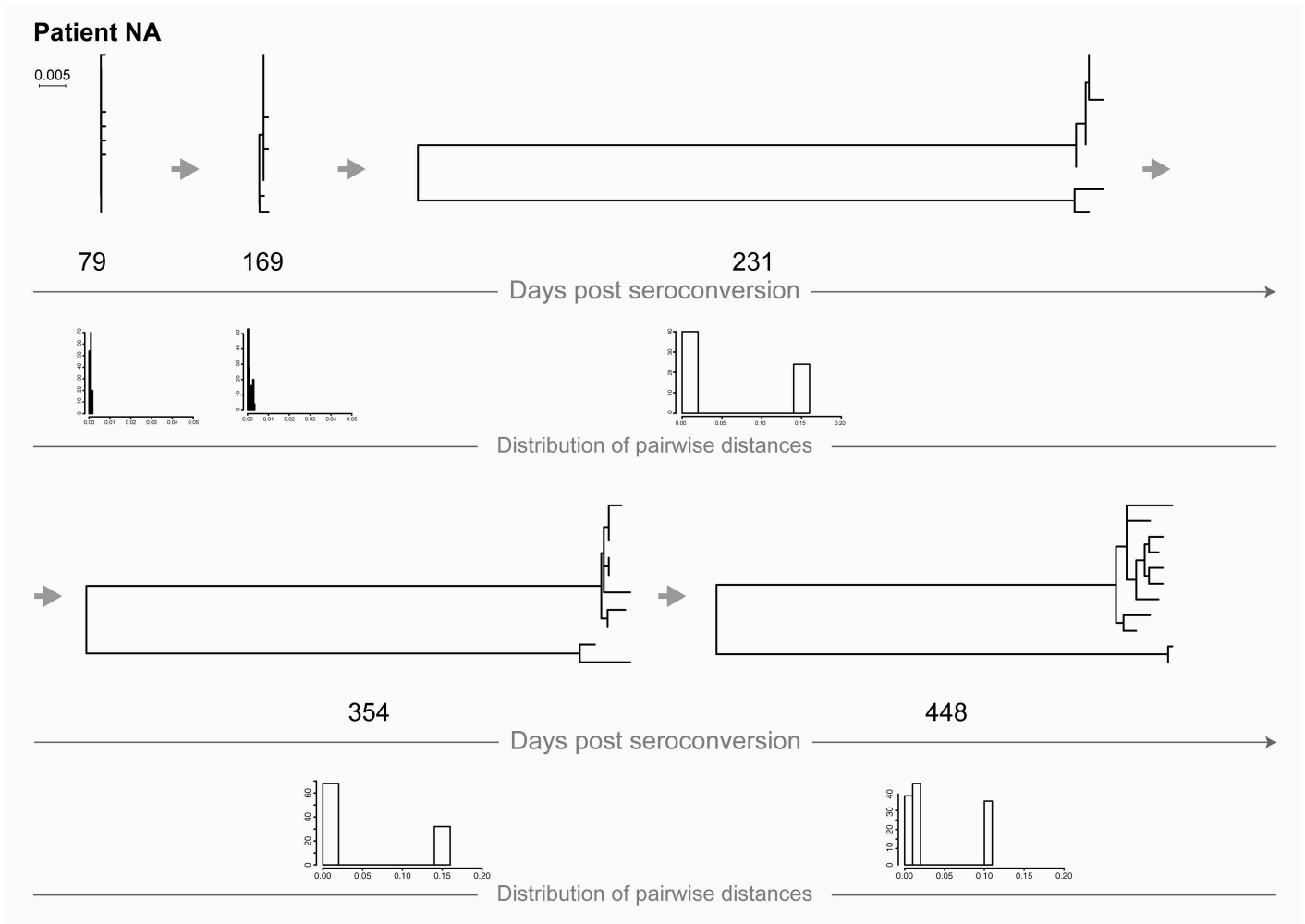
### Transmission of two founder viruses followed by a super-infection

As shown in Fig 6, patient OW was infected with two viral variants, which was evident from the inferred phylogenetic tree from days at earlier time points. Then, by day 469, this patient acquired a distinct HIV, constituting a super-infection with a distinct virus.

### Frequency of HIV transmission

The frequencies of different HIV transmissions within the *Tshedimoso* study cohort [38–40] are presented in Table 2.





**Fig 5. HIV transmission of single viral variant in patient NA (recent infection) followed by a super-infection.** For explanation of phylogenetic trees, pairwise distances, and time points of sampling, please see Fig 1 legend.

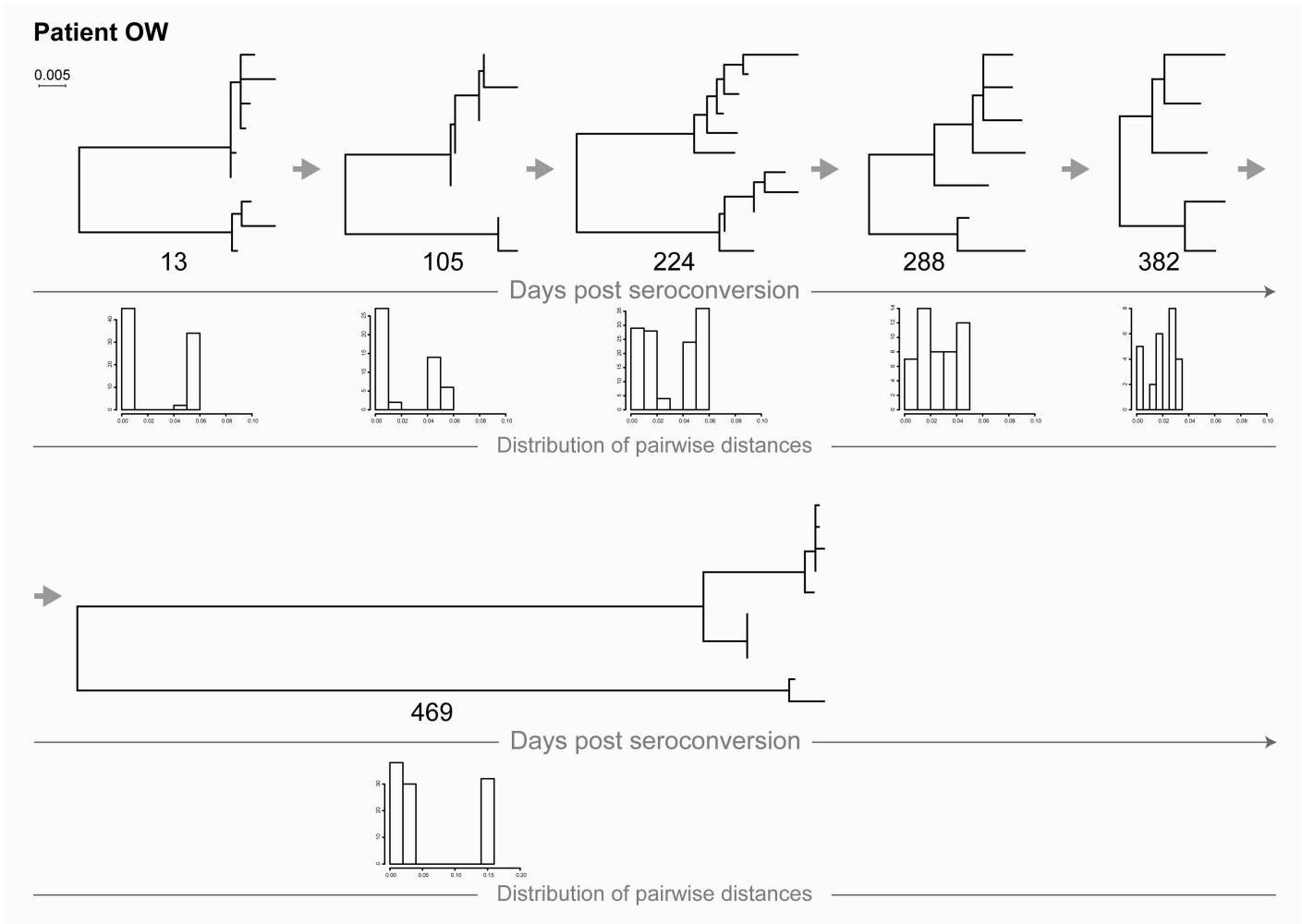
doi:10.1371/journal.pone.0166746.g005

Based on phylogeny and pairwise distance analysis, transmission of a single viral variant was evident in 33 cases (79%; 95% CI 63–90%). Transmission of multiple viral variants from the same source was evident in 9 (21%; 95% CI 10–37%) cases. HIV-1 super-infection was identified in 2 cases (5%; 95% CI 1–17%). The estimated rate of HIV-1C super-infection is 3.9 per 100 person-years. Transmission of multiple viruses combined with HIV super-infection at a later time point was observed once. The population frequency of HIV transmission as multiple variants from the same source remains unclear and warrants further studies.

The frequency of multiple-variant transmission from the same source (21%; 95% CI 10–37%) appeared to be larger than the frequency of HIV superinfection (5%; 95% CI 1–17%), the difference reached statistical significance at the 5% level ( $p = 0.04$ ; McNemar’s test).

## Discussion

Multiplicity of HIV transmission has important implications for design and development of treatment and prevention strategies, and particularly for advancing HIV vaccine research. The extreme cases in multiplicity of HIV transmission, such as transmission of a single founder



**Fig 6. HIV transmission of multiple viral variants in patient OW (recent infection) followed by a super-infection.** For explanation of phylogenetic trees, pairwise distances, and time points of sampling, please see Fig 1 legend.

doi:10.1371/journal.pone.0166746.g006

**Table 2. Sensitivity analysis.** Frequency and rate of different types of HIV transmission in a cohort of 42 individuals with primary HIV-1C infection: phylogeny and pairwise distance analysis is compared with ratio *withinss* to *betweenss* threshold values.

Types of HIV transmission	Phylogeny & pairwise distances		Ratio <i>withinss</i> to <i>betweenss</i> threshold values									
			0.10		0.15		0.20		0.25		0.30	
	n	prop*	n	prop*	n	prop*	n	prop*	n	prop*	n	prop*
Single founder virus	33	0.79 (0.63–0.90)	35	0.83 (0.67–0.93)	34	0.81 (0.66–0.91)	31	0.74 (0.60–0.86)	28	0.67 (0.50–0.80)	21	0.50 (0.34–0.66)
Multiple viral variants from the same source partner	9	0.21 (0.10–0.37)	7	0.17 (0.07–0.31)	8	0.19 (0.09–0.34)	11	0.26 (0.14–0.42)	14	0.33 (0.20–0.50)	20	0.48 (0.32–0.64)
HIV super-infection	2	0.05 (0.01–0.17)	2	0.05 (0.01–0.17)	2	0.05 (0.01–0.17)	2	0.05 (0.01–0.17)	2	0.05 (0.01–0.17)	2	0.05 (0.01–0.17)
Multiple HIV variants from the same source partner followed by a super-infection	1	0.02 (0.001–0.14)	1	0.02 (0.001–0.14)	1	0.02 (0.001–0.14)	1	0.02 (0.001–0.14)	1	0.02 (0.001–0.14)	1	0.02 (0.001–0.14)

prop\*: proportion (95% CI)

doi:10.1371/journal.pone.0166746.t002

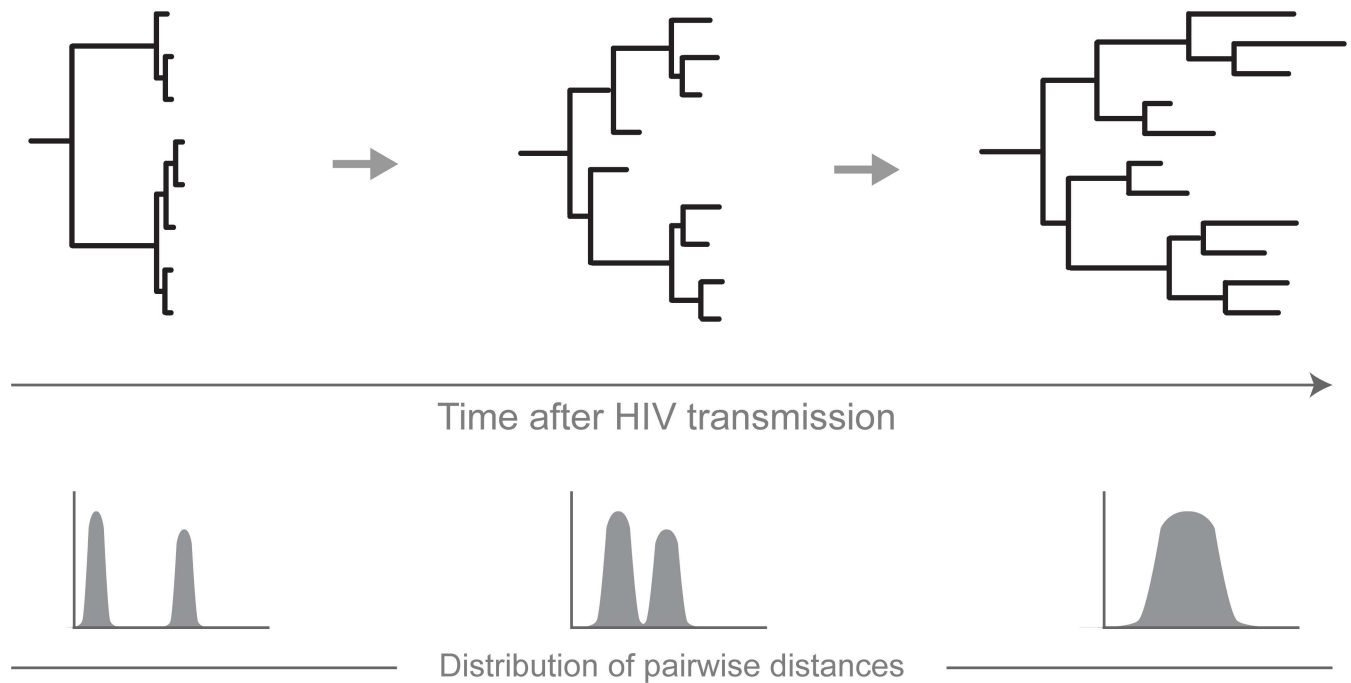
virus, or of substantially distinct multiple viral variants, are well defined. These cases can be detected relatively easily, e.g., by phylogenetic inference of viral sequences representing intra-host HIV quasispecies. However, HIV transmission of closely related multiple viral variants remains uncertain, and the criteria for identification of such multiplicity have not been defined, nor have its clinical or epidemiological relevance.

In this study we utilized HIV sequences representing intra-host viral quasispecies from a prospectively sampled cohort of 42 individuals in Botswana who were enrolled in a primary HIV-1C infection project, the *Tshedimoso* study [4, 38–45]. Viral sequences were generated by single-genome amplification and sequencing and spanned the V1C5 region of the HIV-1C *env* gp120. Transmission of at least two distinct HIV variants from the same source partner was demonstrated in 17% (7 of 42) of cases. The frequency of HIV super-infection was 5% (2 of 42) of cases, similar to the rate in MSM [56].

The identification of closely related multiple viral variants might be challenging. The transient nature of distinct sub-clusters requires sampling during the early stage of HIV infection. If the early time points of sampling are missed, the topology and branch length in the phylogenetic tree and distribution of pairwise distance might not be informative. Moreover, within a short time after transmission of multiple viral variants, the elevated branch lengths and the extended pairwise distances could be interpreted as evidence for an established (chronic) HIV infection, leading to a misclassified recent HIV infection. This phenomenon and sub-optimal sampling could complicate the use of viral diversity as a marker of HIV recency in population studies. However, knowledge of the pattern of multivariant HIV transmission from the same source and its frequency in different populations could help to refine the estimation of HIV recency. If analysis of HIV recency relies on viral diversity, an adjustment for transmission of multiple viral variants could improve accuracy and result in more precise estimation of HIV recency.

Sub-clustering of viral sequences could be defined by a topology of the inferred ML phylogenetic tree—presence of a monophyletic patient-specific lineage with sub-clusters, accompanied by a specific distribution of virus pairwise distances. However, such identification could be subjective, as the criteria for identification of sub-clusters are not well defined. To alleviate this problem and reduce subjectivity in identification of sub-clusters within the pool of HIV sequences representing intra-host viral quasispecies, we suggested a simple method based on the ratio withinness to betweenness. Our intention was to assess the extent to which the ratio withinness to betweenness can be used as a more objective surrogate for subjective interpretation of phylogeny plus pairwise distance distribution. We performed simulation studies (see Table 1), and found that the ratio values (withinness to betweenness) for both sub-clusters greater than zero and less than 0.20 are associated with high sensitivity (0.97) and moderate specificity (0.77), and were accompanied by acceptable positive and negative predictive values. A potential clonal expansion of viral variants, or bias in the sequencing system, may affect or even mislead identification of sub-clusters. This limitation of sub-clusters analysis provides a rationale for developing new methodologies and warrants further studies.

A simplistic diagram in Fig 7 outlines the concept for transmission of multiple HIV variants from a single source partner. The diagram highlights only some key processes occurring during transmission of multiple viruses and does not intend to represent the complexity of HIV evolution. A monophyletic clade evident by a long, patient-specific branch separates viral sequences that represent intra-host HIV quasispecies from other patients' sequences or reference sequences. At the early stage of HIV infection, the internal structure of the clade shows at least two distinct sub-clusters with low diversity of viral quasispecies within each sub-cluster. The histogram of pairwise distances has multiple (at least two) distinct peaks that could represent pairwise distances within and between sub-clusters. This is a transient phase. The



**Fig 7. A simplistic model for HIV transmission of closely related multiple viral variants.** The dynamics of phylogenetic trees and pairwise distances are presented over time of HIV infection.

doi:10.1371/journal.pone.0166746.g007

duration of this phase could reflect complex virus-host interactions and is patient-specific. The branching pattern of the phylogenetic tree changes over time. The dynamic process of filling the gap between originally distinct sub-clusters deserves a separate investigation. Over time, the peaks of pairwise distances in the histogram could converge. The presented diagram does not reflect all possible scenarios, such as overlapping of peaks, or multiple peaks originating from alternative processes.

In summary, the results of this study suggest that upon HIV infection, transmission of closely related multiple viral variants from the same source can be distinguished from transmission of viral variants from different sources. The proposed simplistic model highlights the dynamics of multivariant HIV transmission from the same source. The frequency of this transmission in different populations needs to be addressed in future studies.

### Conclusions

Multiple HIV lineages transmitted from the same source produce a monophyletic clade in the inferred phylogenetic tree. Such a clade has transiently distinct sub-clusters in the early stage of HIV infection, and follows a predictable evolutionary pathway. Over time, the gap between initially distinct viral lineages fills in and initially distinct sub-clusters converge. Identification of cases with transmission of multiple viral lineages from the same source needs to be taken into account in cross-sectional estimation of HIV recency in epidemiological and population studies.

### Acknowledgments

We thank the *Tshedimoso* study participants in Botswana. We also thank Lendsey Melton for excellent editorial assistance.

## Author Contributions

**Conceptualization:** VN ME.

**Formal analysis:** VN SM RW SG.

**Funding acquisition:** VN ME.

**Writing – original draft:** VN.

**Writing – review & editing:** VN RW ME.

## References

1. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A*. 2008; 105(21):7552–7. doi: [10.1073/pnas.0802203105](https://doi.org/10.1073/pnas.0802203105) PMID: [18490657](https://pubmed.ncbi.nlm.nih.gov/18490657/)
2. Haaland RE, Hawkins PA, Salazar-Gonzalez J, Johnson A, Tichacek A, Karita E, et al. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS Pathog*. 2009; 5(1):e1000274. Epub 2009/01/24. doi: [10.1371/journal.ppat.1000274](https://doi.org/10.1371/journal.ppat.1000274) PMID: [19165325](https://pubmed.ncbi.nlm.nih.gov/19165325/)
3. Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping LH, et al. Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J Virol*. 2009; 83(8):3556–67. Epub 2009/02/06. doi: [10.1128/JVI.02132-08](https://doi.org/10.1128/JVI.02132-08) PMID: [19193811](https://pubmed.ncbi.nlm.nih.gov/19193811/)
4. Novitsky V, Wang R, Margolin L, Baca J, Rossen Khan R, Moyo S, et al. Transmission of single and multiple viral variants in primary HIV-1 subtype C infection. *PLoS One*. 2011; 6(2):e16714. Epub 2011/03/19. doi: [10.1371/journal.pone.0016714](https://doi.org/10.1371/journal.pone.0016714) PMID: [21415914](https://pubmed.ncbi.nlm.nih.gov/21415914/)
5. Kiwelu IE, Novitsky V, Margolin L, Baca J, Manongi R, Sam N, et al. HIV-1 subtypes and recombinants in Northern Tanzania: distribution of viral quasispecies. *PLoS One*. 2012; 7(10):e47605. doi: [10.1371/journal.pone.0047605](https://doi.org/10.1371/journal.pone.0047605) PMID: [23118882](https://pubmed.ncbi.nlm.nih.gov/23118882/)
6. Redd AD, Mullis CE, Serwadda D, Kong X, Martens C, Ricklefs SM, et al. The rates of HIV superinfection and primary HIV incidence in a general population in Rakai, Uganda. *J Infect Dis*. 2012; 206(2):267–74. doi: [10.1093/infdis/jis325](https://doi.org/10.1093/infdis/jis325) PMID: [22675216](https://pubmed.ncbi.nlm.nih.gov/22675216/)
7. Redd AD, Mullis CE, Wendel SK, Sheward D, Martens C, Bruno D, et al. Limited HIV-1 superinfection in seroconverters from the CAPRISA 004 Microbicide Trial. *J Clin Microbiol*. 2014; 52(3):844–8. doi: [10.1128/JCM.03143-13](https://doi.org/10.1128/JCM.03143-13) PMID: [24371237](https://pubmed.ncbi.nlm.nih.gov/24371237/)
8. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, et al. Deciphering Human Immunodeficiency Virus Type 1 Transmission and Early Envelope Diversification by Single-Genome Amplification and Sequencing. *J Virol*. 2008; 82(8):3952–70. doi: [10.1128/JVI.02660-07](https://doi.org/10.1128/JVI.02660-07) PMID: [18256145](https://pubmed.ncbi.nlm.nih.gov/18256145/)
9. Joseph SB, Swanstrom R, Kashuba AD, Cohen MS. Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. *Nat Rev Microbiol*. 2015; 13(7):414–25. doi: [10.1038/nrmicro3471](https://doi.org/10.1038/nrmicro3471) PMID: [26052661](https://pubmed.ncbi.nlm.nih.gov/26052661/)
10. Li H, Bar KJ, Wang S, Decker JM, Chen Y, Sun C, et al. High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men. *PLoS Pathog*. 2010; 6(5):e1000890. Epub 2010/05/21. doi: [10.1371/journal.ppat.1000890](https://doi.org/10.1371/journal.ppat.1000890) PMID: [20485520](https://pubmed.ncbi.nlm.nih.gov/20485520/)
11. Sterrett S, Learn GH, Edlefsen PT, Haynes BF, Hahn BH, Shaw GM, et al. Low multiplicity of HIV-1 infection and no vaccine enhancement in VAX003 injection drug users. *Open forum infectious diseases*. 2014; 1(2):ofu056. doi: [10.1093/ofid/ofu056](https://doi.org/10.1093/ofid/ofu056) PMID: [25734126](https://pubmed.ncbi.nlm.nih.gov/25734126/)
12. Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, Grayson T, et al. Wide variation in the multiplicity of HIV-1 infection among injection drug users. *J Virol*. 2010; 84(12):6241–7. Epub 2010/04/09. doi: [10.1128/JVI.00077-10](https://doi.org/10.1128/JVI.00077-10) PMID: [20375173](https://pubmed.ncbi.nlm.nih.gov/20375173/)
13. Masharsky AE, Dukhovlinova EN, Verevchkin SV, Toussova OV, Skochilov RV, Anderson JA, et al. A substantial transmission bottleneck among newly and recently HIV-1-infected injection drug users in St Petersburg, Russia. *J Infect Dis*. 2010; 201(11):1697–702. doi: [10.1086/652702](https://doi.org/10.1086/652702) PMID: [20423223](https://pubmed.ncbi.nlm.nih.gov/20423223/)
14. Tully DC, Ogilvie CB, Batorsky RE, Bean DJ, Power KA, Ghebremichael M, et al. Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus. *PLoS Pathog*. 2016; 12(5):e1005619. doi: [10.1371/journal.ppat.1005619](https://doi.org/10.1371/journal.ppat.1005619) PMID: [27163788](https://pubmed.ncbi.nlm.nih.gov/27163788/)

15. Cornelissen M, Euler Z, van den Kerkhof TL, van Gils MJ, Boeser-Nunnink BD, Kootstra NA, et al. The neutralizing antibody response in an individual with triple HIV-1 infection remains directed at the first infecting subtype. *AIDS research and human retroviruses*. 2016:Epub ahead of print, 2016 Feb 24.
16. Cortez V, Odem-Davis K, McClelland RS, Jaoko W, Overbaugh J. HIV-1 superinfection in women broadens and strengthens the neutralizing antibody response. *PLoS Pathog*. 2012; 8(3):e1002611. doi: [10.1371/journal.ppat.1002611](https://doi.org/10.1371/journal.ppat.1002611) PMID: [22479183](https://pubmed.ncbi.nlm.nih.gov/22479183/)
17. Blish CA, Dogan OC, Derby NR, Nguyen MA, Chohan B, Richardson BA, et al. Human immunodeficiency virus type 1 superinfection occurs despite relatively robust neutralizing antibody responses. *J Virol*. 2008; 82(24):12094–103. doi: [10.1128/JVI.01730-08](https://doi.org/10.1128/JVI.01730-08) PMID: [18842728](https://pubmed.ncbi.nlm.nih.gov/18842728/)
18. Basu D, Kraft CS, Murphy MK, Campbell PJ, Yu T, Hraber PT, et al. HIV-1 subtype C superinfected individuals mount low autologous neutralizing antibody responses prior to intrasubtype superinfection. *Retrovirology*. 2012; 9:76. doi: [10.1186/1742-4690-9-76](https://doi.org/10.1186/1742-4690-9-76) PMID: [22995123](https://pubmed.ncbi.nlm.nih.gov/22995123/)
19. Kraft CS, Basu D, Hawkins PA, Hraber PT, Chomba E, Mulenga J, et al. Timing and source of subtype-C HIV-1 superinfection in the newly infected partner of Zambian couples with disparate viruses. *Retrovirology*. 2012; 9:22. doi: [10.1186/1742-4690-9-22](https://doi.org/10.1186/1742-4690-9-22) PMID: [22433432](https://pubmed.ncbi.nlm.nih.gov/22433432/)
20. Redd AD, Wendel SK, Longosz AF, Fogel JM, Dadabhai S, Kumwenda N, et al. Evaluation of postpartum HIV superinfection and mother-to-child transmission. *AIDS*. 2015; 29(12):1567–73. doi: [10.1097/QAD.0000000000000740](https://doi.org/10.1097/QAD.0000000000000740) PMID: [26244396](https://pubmed.ncbi.nlm.nih.gov/26244396/)
21. Sheward DJ, Ntale R, Garrett NJ, Woodman ZL, Abdool Karim SS, Williamson C. HIV-1 Superinfection Resembles Primary Infection. *J Infect Dis*. 2015; 212(6):904–8. doi: [10.1093/infdis/jiv136](https://doi.org/10.1093/infdis/jiv136) PMID: [25754982](https://pubmed.ncbi.nlm.nih.gov/25754982/)
22. Ronen K, Richardson BA, Graham SM, Jaoko W, Mandaliya K, McClelland RS, et al. HIV-1 superinfection is associated with an accelerated viral load increase but has a limited impact on disease progression. *AIDS*. 2014; 28(15):2281–6. doi: [10.1097/QAD.0000000000000422](https://doi.org/10.1097/QAD.0000000000000422) PMID: [25102090](https://pubmed.ncbi.nlm.nih.gov/25102090/)
23. Castro E, Zhao H, Cavassini M, Mullins JI, Pantaleo G, Bart PA. HIV-1 superinfection with a triple-class drug-resistant strain in a patient successfully controlled with antiretroviral treatment. *AIDS*. 2014; 28(12):1840–4. doi: [10.1097/QAD.0000000000000342](https://doi.org/10.1097/QAD.0000000000000342) PMID: [24911350](https://pubmed.ncbi.nlm.nih.gov/24911350/)
24. Ronen K, McCoy CO, Matsen FA, Boyd DF, Emery S, Odem-Davis K, et al. HIV-1 superinfection occurs less frequently than initial infection in a cohort of high-risk Kenyan women. *PLoS Pathog*. 2013; 9(8):e1003593. doi: [10.1371/journal.ppat.1003593](https://doi.org/10.1371/journal.ppat.1003593) PMID: [24009513](https://pubmed.ncbi.nlm.nih.gov/24009513/)
25. Jost S, Bernard MC, Kaiser L, Yerly S, Hirschel B, Samri A, et al. A patient with HIV-1 superinfection. *N Engl J Med*. 2002; 347(10):731–6. doi: [10.1056/NEJMoa020263](https://doi.org/10.1056/NEJMoa020263) PMID: [12213944](https://pubmed.ncbi.nlm.nih.gov/12213944/)
26. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Antiretroviral therapy for the prevention of HIV-1 transmission. *N Engl J Med*. 2016:Epub ahead of print, 2016 Jul 18.
27. 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(6):493–505. Epub 2011/07/20. doi: [10.1056/NEJMoa1105243](https://doi.org/10.1056/NEJMoa1105243) PMID: [21767103](https://pubmed.ncbi.nlm.nih.gov/21767103/)
28. 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(12):1918–26. Epub 2011/10/13. doi: [10.1093/infdis/jir651](https://doi.org/10.1093/infdis/jir651) PMID: [21990420](https://pubmed.ncbi.nlm.nih.gov/21990420/)
29. Ping LH, Jabara CB, Rodrigo AG, Hudelson SE, Piwowar-Manning E, Wang L, et al. HIV-1 transmission during early antiretroviral therapy: evaluation of two HIV-1 transmission events in the HPTN 052 prevention study. *PLoS One*. 2013; 8(9):e71557. doi: [10.1371/journal.pone.0071557](https://doi.org/10.1371/journal.pone.0071557) PMID: [24086252](https://pubmed.ncbi.nlm.nih.gov/24086252/)
30. Grobler J, Gray CM, Rademeyer C, Seoighe C, Ramjee G, Karim SA, et al. Incidence of HIV-1 dual infection and its association with increased viral load set point in a cohort of HIV-1 subtype C-infected female sex workers. *J Infect Dis*. 2004; 190(7):1355–9. doi: [10.1086/423940](https://doi.org/10.1086/423940) PMID: [15346349](https://pubmed.ncbi.nlm.nih.gov/15346349/)
31. Altfeld M, Allen TM, Yu XG, Johnston MN, Agrawal D, Korber BT, et al. HIV-1 superinfection despite broad CD8+ T-cell responses containing replication of the primary virus. *Nature*. 2002; 420(6914):434–9. doi: [10.1038/nature01200](https://doi.org/10.1038/nature01200) PMID: [12459786](https://pubmed.ncbi.nlm.nih.gov/12459786/)
32. Pacold ME, Pond SL, Wagner GA, Delpont W, Bourque DL, Richman DD, et al. Clinical, virologic, and immunologic correlates of HIV-1 intraclade B dual infection among men who have sex with men. *AIDS*. 2012; 26(2):157–65. doi: [10.1097/QAD.0b013e32834dcd26](https://doi.org/10.1097/QAD.0b013e32834dcd26) PMID: [22045341](https://pubmed.ncbi.nlm.nih.gov/22045341/)
33. Janes H, Herbeck JT, Tovnanubtra S, Thomas R, Frahm N, Duerr A, et al. HIV-1 infections with multiple founders are associated with higher viral loads than infections with single founders. *Nat Med*. 2015; 21(10):1139–41. doi: [10.1038/nm.3932](https://doi.org/10.1038/nm.3932) PMID: [26322580](https://pubmed.ncbi.nlm.nih.gov/26322580/)
34. Gottlieb GS, Nickle DC, Jensen MA, Wong KG, Grobler J, Li F, et al. Dual HIV-1 infection associated with rapid disease progression. *Lancet*. 2004; 363(9409):619–22. doi: [10.1016/S0140-6736\(04\)15596-7](https://doi.org/10.1016/S0140-6736(04)15596-7) PMID: [14987889](https://pubmed.ncbi.nlm.nih.gov/14987889/)

35. Gottlieb GS, Nickle DC, Jensen MA, Wong KG, Kaslow RA, Shepherd JC, et al. HIV type 1 superinfection with a dual-tropic virus and rapid progression to AIDS: a case report. *Clin Infect Dis*. 2007; 45(4):501–9. doi: [10.1086/520024](https://doi.org/10.1086/520024) PMID: [17638203](https://pubmed.ncbi.nlm.nih.gov/17638203/)
36. Sagar M, Lavreys L, Baeten JM, Richardson BA, Mandaliya K, Chohan BH, et al. Infection with multiple human immunodeficiency virus type 1 variants is associated with faster disease progression. *J Virol*. 2003; 77(23):12921–6. doi: [10.1128/JVI.77.23.12921-12926.2003](https://doi.org/10.1128/JVI.77.23.12921-12926.2003) PMID: [14610215](https://pubmed.ncbi.nlm.nih.gov/14610215/)
37. Cornelissen M, Pasternak AO, Grijsen ML, Zorgdrager F, Bakker M, Blom P, et al. HIV-1 dual infection is associated with faster CD4+ T-cell decline in a cohort of men with primary HIV infection. *Clin Infect Dis*. 2012; 54(4):539–47. doi: [10.1093/cid/cir849](https://doi.org/10.1093/cid/cir849) PMID: [22157174](https://pubmed.ncbi.nlm.nih.gov/22157174/)
38. Novitsky V, Wang R, Kebaabetswe L, Greenwald J, Rossenkhan R, Moyo S, et al. Better control of early viral replication is associated with slower rate of elicited antiviral antibodies in the detuned enzyme immunoassay during primary HIV-1C infection. *J Acquir Immune Defic Syndr*. 2009; 52(2):265–72. Epub 2009/06/16. doi: [10.1097/QAI.0b013e3181ab6ef0](https://doi.org/10.1097/QAI.0b013e3181ab6ef0) PMID: [19525854](https://pubmed.ncbi.nlm.nih.gov/19525854/)
39. Novitsky V, Woldegabriel E, Kebaabetswe L, Rossenkhan R, Mlotshwa B, Bonney C, et al. Viral Load and CD4+ T Cell Dynamics in Primary HIV-1 Subtype C Infection. *J Acquir Immune Defic Syndr*. 2009; 50(1):65–76. doi: [10.1097/QAI.0b013e3181900141](https://doi.org/10.1097/QAI.0b013e3181900141) PMID: [19295336](https://pubmed.ncbi.nlm.nih.gov/19295336/)
40. Novitsky V, Woldegabriel E, Wester C, McDonald E, Rossenkhan R, Ketunuti M, et al. Identification of primary HIV-1C infection in Botswana. NIHMSID # 79283. *AIDS Care*. 2008; 20(7):806–11. doi: [10.1080/09540120701694055](https://doi.org/10.1080/09540120701694055) PMID: [18608056](https://pubmed.ncbi.nlm.nih.gov/18608056/)
41. Novitsky V, Gaolathe T, Woldegabriel E, Makhema J, Essex M. A seronegative case of HIV-1 subtype C infection in Botswana. *Clin Infect Dis*. 2007; 45(5):e68–71. Epub 2007/08/09. doi: [10.1086/520683](https://doi.org/10.1086/520683) PMID: [17682982](https://pubmed.ncbi.nlm.nih.gov/17682982/)
42. Novitsky V, Lagakos S, Herzig M, Bonney C, Kebaabetswe L, Rossenkhan R, et al. Evolution of proviral gp120 over the first year of HIV-1 subtype C infection. NIHMSID # 79286. *Virology* 2009; 383(1):47–59. doi: [10.1016/j.virol.2008.09.017](https://doi.org/10.1016/j.virol.2008.09.017) PMID: [18973914](https://pubmed.ncbi.nlm.nih.gov/18973914/)
43. Novitsky V, Wang R, Margolin L, Baca J, Kebaabetswe L, Rossenkhan R, et al. Timing constraints of in vivo gag mutations during primary HIV-1 subtype C infection. *PLoS One*. 2009; 4(11):e7727. Epub 2009/11/06. doi: [10.1371/journal.pone.0007727](https://doi.org/10.1371/journal.pone.0007727) PMID: [19890401](https://pubmed.ncbi.nlm.nih.gov/19890401/)
44. Novitsky V, Wang R, Margolin L, Baca J, Moyo S, Musonda R, et al. Dynamics and timing of in vivo mutations at Gag residue 242 during primary HIV-1 subtype C infection. *Virology*. 2010; 403(1):37–46. Epub 2010/05/07. doi: [10.1016/j.virol.2010.04.001](https://doi.org/10.1016/j.virol.2010.04.001) PMID: [20444482](https://pubmed.ncbi.nlm.nih.gov/20444482/)
45. Novitsky V, Wang R, Rossenkhan R, Moyo S, Essex M. Intra-host evolutionary rates in HIV-1C env and gag during primary infection. *Infect Genet Evol*. 2013; 19:361–8. doi: [10.1016/j.meegid.2013.02.023](https://doi.org/10.1016/j.meegid.2013.02.023) PMID: [23523818](https://pubmed.ncbi.nlm.nih.gov/23523818/)
46. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004; 5:113. doi: [10.1186/1471-2105-5-113](https://doi.org/10.1186/1471-2105-5-113) PMID: [15318951](https://pubmed.ncbi.nlm.nih.gov/15318951/)
47. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004; 32(5):1792–7. doi: [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340) PMID: [15034147](https://pubmed.ncbi.nlm.nih.gov/15034147/)
48. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser*. 1999; 41:95–8.
49. Guindon S, Delsuc F, Dufayard JF, Gascuel O. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol*. 2009; 537:113–37. Epub 2009/04/21. doi: [10.1007/978-1-59745-251-9\\_6](https://doi.org/10.1007/978-1-59745-251-9_6) PMID: [19378142](https://pubmed.ncbi.nlm.nih.gov/19378142/)
50. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010; 59(3):307–21. Epub 2010/06/09. doi: [10.1093/sysbio/syq010](https://doi.org/10.1093/sysbio/syq010) PMID: [20525638](https://pubmed.ncbi.nlm.nih.gov/20525638/)
51. 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(2):221–4. Epub 2009/10/27. doi: [10.1093/molbev/msp259](https://doi.org/10.1093/molbev/msp259) PMID: [19854763](https://pubmed.ncbi.nlm.nih.gov/19854763/)
52. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004; 20:289–90. PMID: [14734327](https://pubmed.ncbi.nlm.nih.gov/14734327/)
53. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
54. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Software*. 2008; 28(5):1–26.
55. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*. 1947; 12(2):153–7. PMID: [20254758](https://pubmed.ncbi.nlm.nih.gov/20254758/)
56. Wagner GA, Pacold ME, Kosakovsky P, Pond SL, Caballero G, Chaillon A, Rudolph AE, et al. Incidence and prevalence of intrasubtype HIV-1 dual infection in at-risk men in the United States. *J Infect Dis*. 2014; 209(7):1032–8. doi: [10.1093/infdis/jit633](https://doi.org/10.1093/infdis/jit633) PMID: [24273040](https://pubmed.ncbi.nlm.nih.gov/24273040/)