



RESEARCH

Open Access

Phylogenetic analysis consistent with a clinical history of sexual transmission of HIV-1 from a single donor reveals transmission of highly distinct variants

Suzanne English¹, Aris Katzourakis², David Bonsall³, Peter Flanagan¹, Anna Duda¹, Sarah Fidler³, Jonathan Weber³, Myra McClure³, SPARTAC Trial Investigators¹, Rodney Phillips^{1,4,5†} and John Frater^{1,4,5*†}

Abstract

Background: To combat the pandemic of human immunodeficiency virus 1 (HIV-1), a successful vaccine will need to cope with the variability of transmissible viruses. Human hosts infected with HIV-1 potentially harbour many viral variants but very little is known about viruses that are likely to be transmitted, or even if there are viral characteristics that predict enhanced transmission *in vivo*. We show for the first time that genetic divergence consistent with a single transmission event *in vivo* can represent several years of pre-transmission evolution.

Results: We describe a highly unusual case consistent with a single donor transmitting highly related but distinct HIV-1 variants to two individuals on the same evening. We confirm that the clustering of viral genetic sequences, present within each recipient, is consistent with the history of a single donor across the viral *env*, *gag* and *pol* genes by maximum likelihood and Bayesian Markov Chain Monte Carlo based phylogenetic analyses. Based on an uncorrelated, lognormal relaxed clock of *env* gene evolution calibrated with other datasets, the time since the most recent common ancestor is estimated as 2.86 years prior to transmission (95% confidence interval 1.28 to 4.54 years).

Conclusion: Our results show that an effective design for a preventative vaccine will need to anticipate extensive HIV-1 diversity within an individual donor as well as diversity at the population level.

Background

A successful HIV-1 vaccine would be designed based upon the antigenicity of transmissible viruses. At the global level, multiple subtypes with evidence of on-going evolution [1] result in a level of diversity that has already frustrated all efforts to synthesize a universal HIV-1 vaccine [2]. Additionally, substantial virus diversity develops within a single host during chronic infection [3], and it is unclear which viral variants are transmissible to a new host. Recent efforts have concentrated on inferring variant transmissibility by

characterizing the precise genetic and antigenic features of viruses found during very early stages of infection [4-9].

Single viral variants are detected in a significant proportion of new HIV-1 infections *in vivo*, indicating a profound genetic bottleneck [6,10]. The degree of genetic bottleneck has been associated with the route of transmission [11-13]. Another factor associated with the number of infecting variants is the presence of genitourinary infections [10]. Together, these data suggest that differences in the degree of genetic bottleneck are related to variations in mucosal defence and its integrity.

However, the actual mechanism of this genetic bottleneck remains unclear, and studies may be confounded by variations in both the risk of transmission among donors and the diversity of transmissible virions within donors [9]. The highest risk of transmission occurs

* Correspondence: john.frater@ndm.ox.ac.uk

† Contributed equally

¹Nuffield Department of Clinical Medicine, Peter Medawar Building for Pathogen Research, Oxford University, South Parks Road, Oxford, OX1 3SY, UK

Full list of author information is available at the end of the article

during primary infection when the population size of infectious virus peaks [14]. However, viral diversity within the acutely-infected donor is limited, potentially making transmitted viruses indistinguishable in the recipient [4-6,11,15].

Furthermore, genetic analysis has also indicated that mucosal defence and integrity are not the only explanations for the apparent genetic bottleneck. Demographic models have been developed that avoid unsupported prior assumptions about the degree of genetic bottleneck [16]. Viral variability was compared [9] in *gag* and *env* genes after transmission in mother-to-child transmission cases and in men who have sex with men (MSM). Viral variability after transmission was not consistently associated with the route of transmission [9]. In addition, a severe genetic bottleneck may be a sufficient, but not a necessary, condition for random transmission of genetic variability [9].

If transmission of viral variability is not random, then transmission may occur by natural selection [17,18]. However, transmissibility has not yet been associated with specific viral characteristics. Most new, sexually-transmitted HIV-1 infections are CCR5-tropic [4,19], but this may reflect biased representation of these variants in genital fluids [20,21]. In eight cases of heterosexual transmission of subtype C [22], transmitted variants tended to have fewer potential N-linked glycosylation sites (PNLGSs) and shorter hypervariable loops than the average variant in the donor host. In addition, recipient *env*-pseudotyped virus was more susceptible to neutralization by donor serum than donor *env*-pseudotyped virus [22]. A study of 35 subtype A cases from Kenya, and 13 subtype B cases from the USA [23] found that recently-infected persons had viruses with shorter, less-glycosylated V1V2 loops compared with a database of viruses [23]. However, studies of subtype B have not shown a consistent decrease in hypervariable loop length or the number of PNLGSs [24,25]. Therefore, there is no firm evidence that natural selection determines transmission of viral variants.

Animal models of HIV infection that use the closely-related simian immunodeficiency virus (SIV) have also demonstrated that many different variants circulating within the host are transmissible. A low-dose, intrarectal inoculum of SIV was given to 18 rhesus macaques [26] to mimic physiological concentrations. Although between one and five variants initiated new infections, the viruses transmitted to all macaques collectively reflected the diversity within the inoculum [26]. Another study [27] demonstrated a stochastic pattern of V1V2 variant transmission from an inoculum. Therefore, a broad range of viruses circulating in a single donor may be potentially transmissible at any one time, consistent

with the hypothesis that transmission of viral variants is a random process.

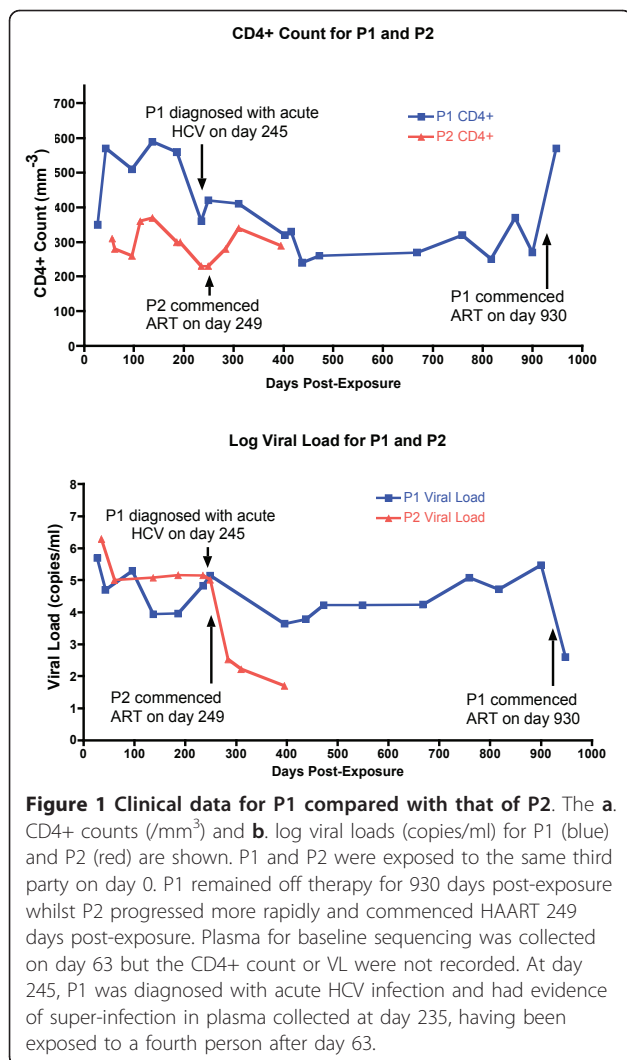
To demonstrate that this lack of predictability is also true for HIV-1 transmission in humans, we present an unusual case consistent with a clinical history of one male having transmitted significantly divergent HIV-1 variants to two recipients on the same evening. We show that, as with macaques, diversity in early infection is limited and compatible with transmission of a single variant to each recipient, but also that a single donor can transmit two very different HIV-1 strains contemporaneously. Furthermore, we do not find any evidence that this between-host genetic divergence is evidence of selection pressure from either humoral or cellular immunity during or since transmission. Finally, if transmission is a random process, we hypothesize that a protective vaccine would need to cover the breadth of transmissible variation within individual donors as well as population-wide diversity.

Results and Discussion

Case history of a single, third party exposure and recent seroconversion

Two adult males, P1 and P2, reported a single sexual encounter each with the same third-party that occurred on day 0 (Figure 1). P1 and P2 reported subsequent exposure only to each other prior to enrolment in the Short Pulse AntiRetroviral Therapy at HIV seroConversion (SPARTAC) trial. Despite repeated efforts, the third-party donor could not be traced. On day 6 post-exposure, P1 presented to his primary care physician with symptoms compatible with HIV seroconversion. On day 25, P1 tested positive for HIV-1 by ELISA with an incident result on a detuned ELISA, suggestive of recent infection [28,29]. P2 had a positive HIV-1 PCR and negative HIV-1 ELISA on day 22, and on day 35 was p24 positive, but negative by Murex ELISA (R&D Systems, UK) [30]. The Murex ELISA was repeated on day 56 and had become clearly positive. Although, the Murex ELISA was positive in P1 earlier than in P2, the result was consistent with reported between-host variability in both the duration of the pre-viraemic phase and the timing of the appearance of markers of seroconversion [30,31]. Therefore, clinical and laboratory evidence supported recent seroconversion in P1 and P2.

P1 and P2 were sampled on the same day when they enrolled in the SPARTAC trial, 63 days post-exposure. Both participants were randomized to receive no therapy. Plasma for sequencing was re-sampled on the same date from both participants, on day 235 post-exposure. P1 reported exposure to a fourth party after day 63 and before day 235. Evidence of HIV-1 super-infection in P1 was seen on plasma collected at day 235 (data not



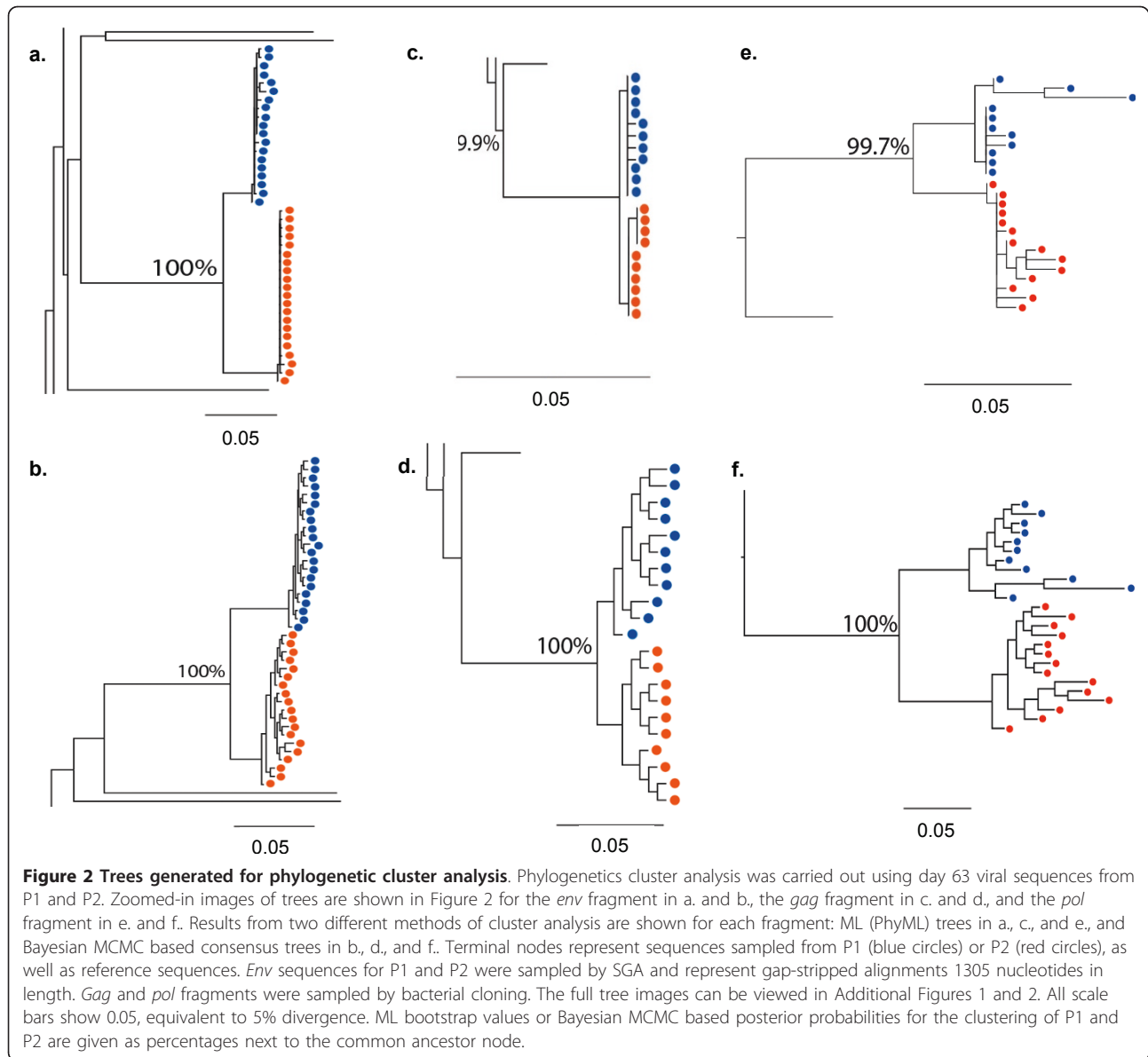
shown). On day 245, P1 was diagnosed with acute hepatitis C virus (HCV) infection (Figure 1) having been negative for HCV by PCR and antibody on day 29. He commenced treatment with ribavirin and interferon after day 245. Therefore, all time-points after day 63 were excluded from further phylogenetic analysis.

The CD4+ count and plasma viral load values for P1 and P2 are shown in Figure 1. Despite the same exposure, P1 and P2 followed different clinical courses. P1 maintained a CD4+ T cell count greater than 350 cells/ mm^3 during the first 310 days of untreated infection compared with P2, who had only two CD4+ readings greater than 350 cells/ mm^3 over the first 249 days of infection. The plasma viral load for P1 was consistently lower than P2 after day 96, with the exception of a second peak reading in P1 taken on day 249, after the detection of HIV-1 super-infection and acute HCV infection. Therefore, P2 appeared to progress more rapidly than P1.

Further clinical laboratory evidence was consistent with the history of a single donor because the time window for one participant to have infected the other was short. Participants P1 and P2 were both positive for p31 antigen on Western Blot on day 63. Therefore, the minimum estimated time since the onset of detectable viraemia (> 50 copies/ml) of approximately 47.4 days [30,31]. Thus, the estimated maximum pre-viraemic phase for either participant was 15 to 16 days. Since, the estimated pre-viraemic phase for HIV-1 lasts between 7 and 25 days [30-33], one participant could have infected the other only between day 7 and day 9 post-exposure to the third party. However, peak viral load in acutely infected subjects is reached 7 or more days after the onset of detectable viraemia [6,12,34] and the infectiousness of a donor MSM is low if his viral load is 400 copies/ml or less [35]. Therefore, while the laboratory evidence did not exclude this alternative scenario, it was unlikely that one participant infected the other.

Sequences for phylogenetic analysis obtained from multiple viral genes

If P1 and P2 had indeed been infected by the same third person on the same night, we expected that viral sequences sampled from one recipient would be highly similar, or even identical, to sequences sampled from the other recipient. We sampled fragments of three different HIV-1 genes, 63 days post-exposure (Figure 2). The gene fragments were located within the *env*, *gag* and *pol* genes. We sampled an *env* fragment from the start of the *gp160* coding region to the end of the *gp120* coding region (HXB2 nucleotide position 6225 to 7757) by single genome amplification (SGA)[4-6,12,13,36]. After 5% gap-stripping with GapStreeze, the *env* gene fragment alignment was 1305 base pairs in length. The more conserved *gag p24 to p6* (HXB2 1471 to 1976) and *pol Reverse Transcriptase (RT)*, (HXB2 2643 to 3428) gene fragments were sampled by bacterial cloning [37]. We included reference sequences from individuals in the same geographical area and demographic risk group, drawn from the SPARTAC trial and the St Mary's Hospital Acute Infection Cohort [38], as well as the LANL UK reference database. Trees were rooted with outlier sequences from different HIV-1 subtypes and non-M groups in the LANL database. Sequences from both participants clustered with subtype B reference sequences in phylogenetic analyses of all three genes. GenBank accession numbers for sequences from the SPARTAC trial UK cohort and the St Mary's Hospital Acute Infection Cohort in this study are FJ645274 to FJ5645360, JF440652 to JF440693, JF499738 to JF499786, JF506093 to JF506179, and JF692885 to JF693023.



Between-host phylogenetic analysis supports the clinical history of a single donor

By both maximum likelihood (ML) and Bayesian MCMC based analyses, sequences from P1 and P2 were highly related and clustered to the exclusion of all other sequences, consistent with a common donor (Figure 2, Additional Files 1 and 2). We demonstrated the statistical support for the robustness of the cluster by both methods (Figure 2 - ML bootstrap values for three genes were: *env* 100%, *gag* 99.9% and *pol* 99.3%, and Bayesian MCMC based posterior probabilities were: 100% for *env*, *gag* and *pol*). We could not use phylogenetic inference to exclude the possibility that one participant infected the other, since such techniques cannot prove the direction of transmission in a forensic sense

[39]. For example, we could not exclude the possibility that two strains were transmitted to one participant and that an initially infectious strain was out-competed to extinction prior to day 63. However, results from other studies suggested this was unlikely [5,6,13,40,41]. Therefore, phylogenetic analyses were consistent with the clinical history that a single, third party contemporaneously transmitted the divergent strains that infected P1 and P2.

Significant between-host divergence observed in transmitted HIV-1 *env* and *pol* genes

We measured the inter-host distance for stem branches, which are the internal branches separating the within-patient sequences. For the *gag* gene fragment, which we

expected to be the most conserved fragment, the inter-host distance was 0.54% by ML analysis (Figure 2c). The inter-host distance for the *env* fragment, which we expected to be the least conserved of the three, was 3.81%(Figure 2a). For the *pol* fragment, the inter-host distance was 1.93% (Figure 2e). The inter-host distance for *env* contrasts with the smaller mean distance within each participant. For *env*, the mean within-patient distance was 0.54% by ML analysis in both participants across the gap-stripped 1305 nucleotide alignment, consistent with the history of recent infection (Figure 2a). In addition, sequence analysis of day 235 plasma also failed to detect *env* or *pol* sequences from P1 in P2 and vice versa (data not shown). Therefore, despite sharing highly similar *gag* genes, consistent with the clinical history of a common origin, P1 and P2 appeared to be infected with remarkably different *env* variants and, to a lesser extent, *pol* variants.

Current implementations of ML and Bayesian tree analysis do not model gaps or non-aligned regions informatively [42]. As phylogenetic analysis of the *env* region meant removing gaps and non-aligned portions, we compared full-fragment, non-stripped *env* sequences from P1 and P2 with the baseline consensus sequence for P1 in a Highlighter plot (Figure 3). There was sequence homogeneity within both P1 and P2, compatible with a single strain initiating a recent infection for each. However, there were multiple sites of variation when P1 was compared with P2. Secondly, we quantified the percentage phylogenetic signal-to-noise (STN)[43] in *env*. We compared our full *env* fragment with gaps to the same fragment with 5% gap-stripping. The percentage STN between P1 and P2 was 70.7% to 24.3% in the unstripped *env* fragment and 62.0% to 30.7% for the stripped *env*. Nevertheless, the percentage STN in the stripped alignment between hosts was greater than in previous studies of multiple-variant transmissions in this genomic region [6,12]. Our analyses indicated that there was a small loss of between-host phylogenetic signal in *env* by stripping gaps or poorly aligned regions. However, stripped *env* fragment alignments contained a higher percentage STN than either the shorter *gag* alignment (49.4% to 50.5%) or shorter *pol* alignment (4.2% to 35.5%). The *gag* and *pol* fragment alignments did not require stripping. Noise $\geq 30\%$ was consistent with a phylogenetic cluster [44,45], but we needed to quantify between-host evolution prior to transmission by another method.

Env divergence quantified by estimating the tMRCA

To quantify pre-transmission evolution, we estimated the time since divergence of the two *env* variants infecting P1 and P2 by calibrating the sequence evolution rate for the *env* C2V5 region of *gp120* against another

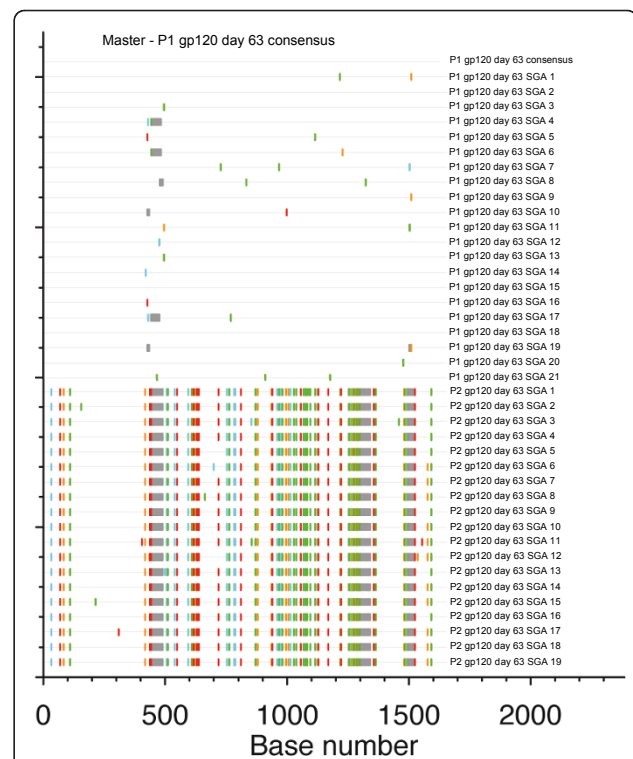
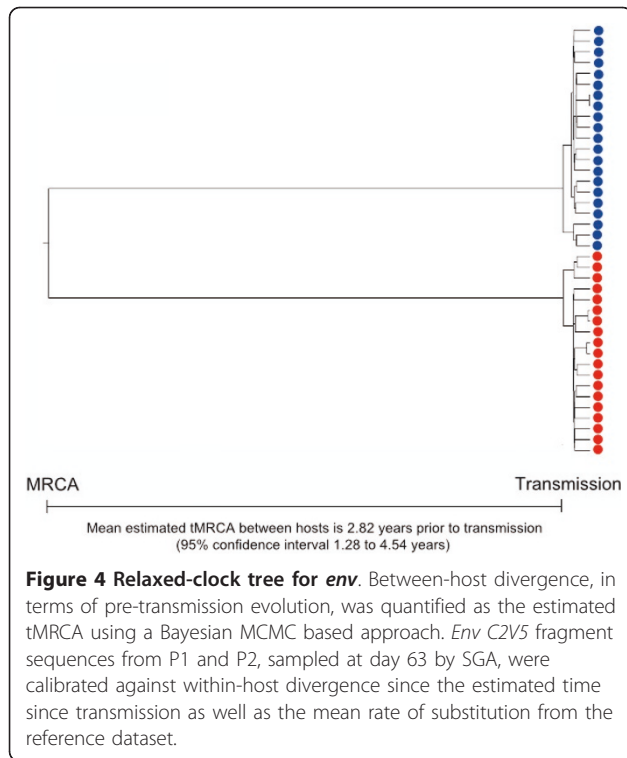


Figure 3 Highlighter plot of *env* *gp120* nucleotide sequences. Full-length *env* *gp120* sequences from day 63 were sampled by SGA. The Highlighter plot shows gaps in grey and nucleotide substitutions (A = green, T = red, G = orange, C = light blue), revealing difficult-to-align regions. The master sequence against which all other sequences are compared is the majority-rule P1 consensus sequence at day 63, shown as the top sequence.

dataset and by measuring the degree of within-host diversification since transmission [3,15]. Using Bayesian MCMC based inference, we estimated the inter-host distance as the time to the most recent common ancestor (tMRCA) which was 2.82 years (95% confidence interval: 1.28 to 4.54 years) of viral evolution (Figure 4). We repeated this analysis with different priors (Additional File 3). All of these results were consistent, and the common ancestor of the HIV-1 *env* genes infecting P1 and P2 was estimated to have existed at least 1.14 years prior to transmission, either in a chronically infected donor or in a recent previous host. These estimates were again consistent with the clinical history of a single, third party having infected both P1 and P2, and that highly divergent sequences could be transmitted by a single donor within a very short period of time.

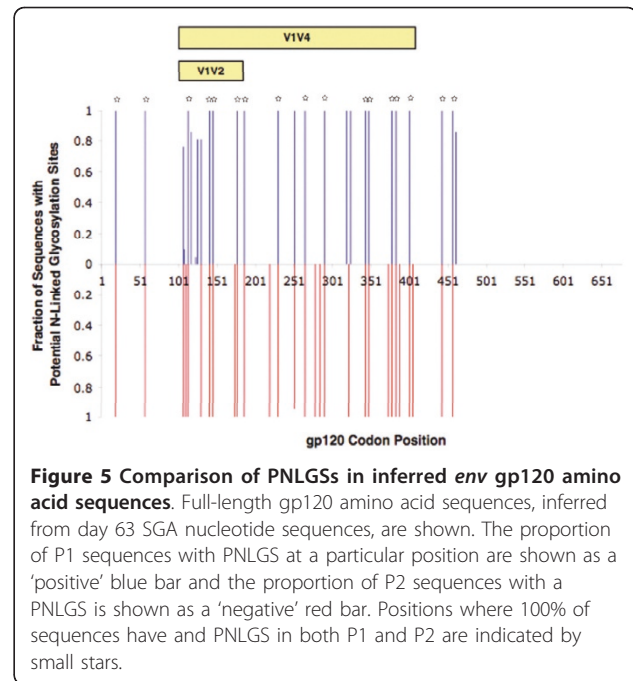
Potential antigenic variation in the *gp120* proteins of transmitted viruses

However, demonstrating a high level of divergence did not answer whether each patient received divergent variants at random or whether there was selection at



transmission. Transmission of divergent *env gp120* variants could be due to hard selection for differences in antigenicity in each recipient. Hard selection involves selective mortality of variants [46]. In rhesus macaques, SIV envelope proteins appear to be under hard selection at transmission due to neutralizing antibodies [47]. Attempts have been made to infer the antigenicity of HIV-1 envelope proteins to neutralizing antibodies from the number of potential N-linked glycosylation sites (PNLGSS) in *gp120* [22,48]. Therefore, we hypothesized that differences in the number of PNLGSS in *gp120* would indicate potential between-host differences in viral antigenicity.

We compared PNLGSS within inferred amino acid sequences for *gp120* from P1 and P2 using N-Glycosite (Figure 5). P1 had a mean of 24 PNLGSS (range 23 to 25). P2 had a mean of 29 PNLGSS (range 28 to 29). Firstly, we looked for positions where P1 and P2 were identical. P1 and P2 shared PNLGSS in 100% of sequences at 17 positions. To demonstrate that this degree of identity was consistent with a phylogenetic cluster, we compared these sequences with 242 unrelated sequences. We studied 87 full-length, inferred amino acid sequences for *gp120* sampled from other SPARTAC participants at trial baseline by population sequencing, as well as 155 subtype B sequences from the LANL database sampled during acute infection. The combined SPARTAC/LANL reference sequences had



100% PNLGSS predictions at only one site, located in C1. Greater than 90% of the reference sequences had a PNLGSS at only seven positions. We concluded that the degree of similarity between P1 and P2 was consistent with a phylogenetic cluster due to transmission from a single donor.

We then looked at the positions that were not 100% identical, to see if there was any evidence of potential hard selection in each recipient during transmission. In particular, we focussed on the V1V4 region that is implicated in susceptibility to neutralizing antibodies. Previous studies of this region have suggested that fewer PNLGSS in this region increases the susceptibility of highly related strains to neutralizing antibody [22,24,25,49]. We found a higher mean number of PNLGSS across V1V4 in P2 (24 sites, range 23-25) than P1 (19 sites, range 18-20; $p < 0.0001$, unpaired T-test). These data indicated that there could be a difference in susceptibility to neutralizing antibodies between these two strains, consistent with a non-random model of transmission.

No autologous or cross-neutralization observed despite potential antigenic variation

We hypothesized that differences at PNLGSS might equate to differences in neutralization that would explain the transmission of divergent *env* variants [22,24,25,49]. Therefore, we investigated whether the viral isolates from P1 and P2 had different neutralization profiles. Viruses pseudotyped with full-length day 63 *env* sequences from P1 and P2 were tested against

autologous or heterologous serum from each participant sampled at day 186 post-exposure. However, the *env* pseudotypes for both P1 and P2 were only poorly neutralized or cross-neutralized (half maximal inhibitory concentration, IC_{50} , of serum $\leq 1:20$, Additional File 4). Therefore, it seemed unlikely that a humoral response was responsible for the detection of different *env* variants in P1 and P2, consistent with transmission being a random process.

However, envelope proteins are not only potentially under immune selection at transmission but also might be selected for an increased ability to enter cells. We used the data from our neutralization assay to estimate the infectivity of the *env* pseudotyped viruses *in vitro*. Pseudoviruses derived from P1 sequences were approximately 2.5 times ($P < 0.05$) more infectious *in vitro* than pseudoviruses from P2, after normalization to reverse transcriptase levels (Additional File 5). We noted between-host diversity in *C2C4*, including differences in glycosylation. *C2C4* encodes discontinuous regions involved in CD4 and co-receptor binding [50-52]. Inferred gp120 protein sequences were analysed with several algorithms that were evaluated by Low and colleagues [53], to detect differences in predicted co-receptor usage and minimize the possibility of missing CXCR4/CCR5 dual-use variants. However, these algorithms predicted that all sampled viruses from P1 and P2 would use CCR5. Our experiment was not specifically set up to test infectivity so all these results must be interpreted with caution. In addition, potential differences in infectivity do not explain why both viruses were able to cause productive infection in different individuals. Therefore, we found no evidence to reject a random model of transmission.

HLA Class 1 restricted responses and potential selection pressure around transmission

We also investigated HIV-1 specific cellular immune responses, to exclude another potential source of hard selection in each participant that might influence our results. Clinical progression and viral load have been associated with host HLA Class I type in chronic infection [54-56]. HLA Class I restricts the ability of host cytotoxic T lymphocytes (CTLs) to recognize and destroy infected cells. Furthermore, sequencing studies have detected evidence consistent with escape from CTL responses within weeks of HIV-1 infection [57]. The role that CTLs play in preventing established viral infection in humans remains unclear. However, vaccination of rhesus macaques to produce detectable CTL responses is associated with partial protection from infection [58], and HIV-1 specific CTL responses have been detected in persons who remain PCR/ELISA

negative despite high-risk exposure [59-61]. Therefore, we hypothesized that CTL responses during and after transmission were a potential source of hard selection in P1 and P2.

Firstly, we compared the Class I HLA type of P1 and P2 with the clinical data to see if there was evidence of selection. P1 possessed HLA-A*0201, A*2402, HLA-B*1402, B*3543, Cw*0102, Cw*0802; P2 possessed HLA-A*0101, A*2901, B*0801, B*5001, Cw*0602, Cw*0701. Neither participant possessed HLA types that are strongly associated with protection from progression in chronic infection [62,63]. However, P2, who progressed quickest, possessed the HLA-A*0101 B*0801 haplotype that is associated with more rapid progression [64]. Therefore, we hypothesize that host factors contribute to the different clinical outcome in these participants and that the viruses had been under different selection pressures since transmission.

Detectable CTL responses do not explain between-host divergence in *env*

We investigated whether different CTL responses could have influenced detection of divergent variants in our study. Phylogenetic analysis assumes neutral evolution rather than natural selection [44]. Therefore, we compared viral sequence data and γ -interferon ELISpot data from each participant to see if cytotoxic T lymphocyte responses since transmission may have accounted for observed between-host divergence in *env* [65,66]. Sequence data were available for the two *env* gp120 optimal peptides against which P2 had a significant response: TVYYGVPVWK (HXB2 gp160 30-46) and SFPEPIHY (HXB2 gp160 202-221). The inferred amino acid sequences for P1 were identical to the wild-type peptides at these epitopes: TVYYGVPVWR and SFPEPIHY. P2 was also infected with wild-type TVYYGVPVWR, as well as both wild-type and mutant SFPEPIHK sequences. Therefore, between-host genetic differences in *env* could not be attributed to detectable, *env*-directed CTL responses, and our data were still consistent with transmission of *env* variants being a random process.

Conclusions

We have quantified for the first time significant, between-host genetic divergence in HIV-1 variants that are likely to have been transmitted by a single donor to two recipients on the same night. Furthermore, these data indicate that currently it is not possible to predict which of the many HIV-1 variants circulating at the time of transmission will successfully seed a new infection. If transmission is a random process, then this represents a major hurdle that any HIV-1 vaccine design will need to overcome.

Methods

Participants

360 participants, 151 of whom were from the UK or Ireland, were recruited to the Short Pulse AntiRetroviral Therapy at HIV seroConversion (SPARTAC) trial (ISRCTN number 76742797; EudraCT number 2004-000446-20). Two male individuals from the UK cohort, P1 and P2, were identified on clinical history as having epidemiologically-linked infections: they were partners and had shared a sexual encounter with a single, third male on the same night. P1 and P2 were enrolled in the trial on the same day and followed up at the Jefferiss Trust Clinic, St. Mary's Hospital, Paddington, London, UK. They were both randomized to receive no therapy.

Ethics Statement

This study has been approved by the Multicentre Research Ethics Committee (MREC). All participants provided written informed consent before participating in this study.

HLA typing

Participant HLA type was determined to the oligo-allelic level using Dynal RELITM Reverse Sequence-Specific Oligonucleotide kits for the HLA-A, -B and -C loci (Dynal Biotech). To obtain four-digit typing, Dynal Biotech Sequence-Specific priming kits were used, in conjunction with the Sequence-Specific Oligonucleotide type.

Separation of PBMCs and plasma

Peripheral blood mononucleocyte (PBMC) and plasma samples were separated from fresh EDTA blood by Ficoll/Hypaque density gradient centrifugation. For PBMC collection, blood was diluted with R10 solution: RPMI 1640 (Sigma UK) with 10% fetal calf serum (FCS; Sigma, UK), 50 units/ml penicillin/streptomycin mix and 2 μ M L-glutamine. The mixture was then layered over Lymphoprep separation medium (Gibco, UK). Samples were centrifuged at 100 \times g at room temperature. The resultant layer of PBMC was removed and washed. 1 ml aliquots containing 5 \times 10⁶ cells were stored in cryotubes in liquid nitrogen at -180 \pm C. For plasma collection, blood samples were prepared as above with dilution with R10, and the resulting plasma was collected in 1 ml aliquots and stored at -80 \pm C.

Viral RNA extraction

1 ml aliquots of frozen plasma were used for each extraction. The plasma was centrifuged at 1600 \times g and 4 \pm C for 1 hour to pellet the virus. Excess plasma was removed and the pellet was resuspended in 140 μ l of remaining plasma. RNA was then extracted with the QIAamp Viral RNA Minikit (Qiagen, UK) according to the manufacturer's instructions.

Reverse transcription and polymerase chain reaction (PCR)

For *env*, viral RNA was reverse transcribed using the SuperScript III Kit (Invitrogen, UK) to produce cDNA. 15 μ l of viral RNA was added to 1.5 μ l dH₂O, 1.5 μ l primer OFM19 [6] (concentration 20 μ M) and 1.5 μ l dNTPs (concentration 10 mM). The mix was heated to 65 $^{\circ}$ C for 5 min followed by 4 $^{\circ}$ C for 1 min to anneal the primers to the RNA. The reverse transcription (RT) reaction mix (5 \times Buffer: 6 μ l, DTT: 1.5 μ l; RNaseOUT 1.5 μ l; SuperScript III 1.5 μ l) was then added to make a final volume of 29 μ l. The reaction mix was heated to 50 $^{\circ}$ C for 60 min, followed by 55 $^{\circ}$ C for 60 min and finally 75 $^{\circ}$ C for 10 minutes. For *gag* and *pol*, viral RNA was reverse transcribed using the Reverse-iT 1st Strand Synthesis Kit (Abgene, UK). 18 μ l of viral RNA was added to 1.5 μ l primer (random decamers and oligodT supplied with the kit, concentration 20 μ M). The mix was heated to 75 $^{\circ}$ C for 5 min followed by 4 $^{\circ}$ C for 2 min to anneal the primers to the RNA. The RT reaction mix (5 \times Buffer: 6 μ l; dNTPs: 3 μ l concentration 10 mM; RTase Blend 1.5 μ l) was then added to make a final volume of 30 μ l. The reaction mixture was heated to 42 $^{\circ}$ C for 60 min followed by 75 $^{\circ}$ C for 10 min. The HIV *gag* and *pol* genes were amplified by separate PCR reactions as described in detail elsewhere [67]. The HIV *env* genes were amplified by PCR using a protocol for single genome amplification as described in detail elsewhere [5,6].

Single genome amplification

Single genome amplification (SGA) of *env* was carried as described elsewhere [5,6]. A 30% cut-off for positive wells was used [5,6,36].

Bacterial cloning

Bacterial cloning was carried out for *gag* and *pol* using the TOPO TA "One Shot" Cloning Kit for Sequencing (Invitrogen, UK). Purified PCR products were ligated into the pCR4-TOPO vector. *Escherichia coli* were mixed on ice with the ligation mix and then transfected by heat shock at 42 $^{\circ}$ C for 30 s. Cells were immediately removed to ice and then added to SOC medium (Invitrogen, UK) and placed on a shaking incubator at 37 $^{\circ}$ C and < 1 \times g for 1 hour. Cells were then spread on plates of 1 \times lysogeny broth (LB) agar (Sigma, UK) containing 0.1 μ g/ml ampicillin (Sigma, UK) and incubated overnight at 37 $^{\circ}$ C. Negative controls were included. Colonies were then selected and added to individual wells containing 2 \times LB medium (Sigma, UK) with 0.05 μ g/ml kanamycin (Sigma, UK). The wells were incubated on a shaking incubator overnight at 37 $^{\circ}$ C and < 1 \times g. Bacteria were lysed and minipreps of clonal plasmid DNA (pDNA) were prepared using the Montage Miniprep₉₆ Kit (Millipore, US).

Sequencing

Sequencing of population PCR, SGA and bacterial cloning DNA products was performed using BigDye technology in a 96-well plate. For population PCR and SGA products, 3 μ l DNA was added to a mix containing 0.8 μ l BigDye Terminator (Applied Biosystems, UK), 1.5 μ l 5 \times sequencing buffer (Applied Biosystems, UK), 2 μ l of primer (3.3 μ M) and 2.7 μ l dH₂O. For bacteria-cloned pDNA, 4 μ l of miniprep was added to a mix containing 1 μ l BigDye Terminator, 1.5 μ l 5 \times sequencing buffer, 1 μ l of primer (3.3 μ M) and 3.5 μ l dH₂O. The following cycling conditions were used: 96°C for 30 s, then 30 cycles of 96°C for 30 s, 50°C for 15 s and 60°C for 4 min. DNA for sequencing was precipitated on ice with 2 μ l 3M sodium acetate, 10 μ l dH₂O, 50 μ l ice-cold 100% ethanol for 5 min at -20°C, centrifuged at 600 \times *g* for 80 min at 4°C, washed twice with ice-cold 70% ethanol and run on an ABI 3700 sequencer.

Sequence alignment

All sequences were manually edited using Sequencher v4.8 (Gene Codes Corporation, US) and manually aligned using Se-AL v2.0a11 [68,69]. For *env* alignment, sequences were first aligned with MUSCLE v3.7 [70] followed by manual alignment. Sequences containing stop codons or frameshifts were deleted prior to subsequent analysis. Where appropriate, reference sequences were obtained from the Los Alamos National Laboratory (LANL) HIV sequence database [71]. For *env*, which contains many gaps and poorly aligned regions, gap stripping was undertaken first with GapStreeze set to 5% [72]. In GapStreeze, the user sets a gap tolerance between 0% and 100%. A value of 5% will cause all columns in the alignment to be deleted if more than 5% of sequences contain a gap at that position. Sequences were manually edited in Se-AL v2.0a11 before and after gap-stripping.

Between-host phylogenetic analysis

Phylogenetic analysis of viral sequences sampled from P1 and P2 was carried out by several methods across the *env*, *gag* and *pol* gene-fragments. Prior to gap-stripping with GapStreeze, a likelihood mapping [45] analysis was run to ensure phylogenetic signal within *env* was significant. Likelihood mapping was implemented in Tree-Puzzle v5.3.rc7 [73] and the *env* fragment was screened from the beginning of the coding start region to the end of *gp120* (HXB2 nucleotide position 6225 to 7757). Additionally, full nucleotide sequences for the fragments from all three genes were visually screened in Highlighter [74], and the inferred protein sequence were screened visually using Jalview v2.6 [75,76]. Phylogenetic trees were initially constructed using the maximum likelihood (ML) method with PhyML v3.0 software [77],

and visualized in FigTree v1.3.1 [78]. We chose the substitution model that gave the highest likelihood with PAUP*v4.0 [79]: the generalized time reversible (GTR) model incorporating estimates of the proportion of invariant sites (I), and the shape parameter of a gamma distribution [80]. ML branch support values were obtained by non-parametric bootstrapping using PhyML v3.0 (1000 replicates). Finally, phylogenetic analysis using a Bayesian MCMC based method was implemented in Mr Bayes v3.1.2 [81,82]. An unconstrained branch length (exponential) prior was used to avoid enforcing a molecular clock [44]. MrBayes v3.1.2 was run in duplicate for at least 50,000,000 steps for *env* and *pol*, sampling trees every 1,000 steps. MrBayes v3.1.2 was run in duplicate for at least 100,000,000 steps for *gag*, sampling every 10,000 steps. Convergence was assessed with Tracer v1.5 [83] with all parameter estimates having effective sample sizes (ESSs) of > 300, because a high ESS reflects a low degree of correlation among samples [44]. The consensus tree for each gene, with posterior probabilities for branch support, was generated and visualized in FigTree v1.3.1.

Inferring the tMRCA using a relaxed molecular clock

To determine the time to the most recent common ancestor (tMRCA) of the sequences isolated from the two participants, we used a Bayesian MCMC based approach. We tested our assumption that all of the observed evolution in *env* within the viral sequence sets from each participant had occurred within each host by demonstrating a star-like intra-host phylogeny, and confirming that intra-host divergence by ML was consistent with that predicted for early, monophyletic infection against other datasets [3,5,6,12,15]. We used a normal tMRCA prior for the sequences within each participant, calibrated to a mean of 63 days since exposure (standard deviation 1 day). We ran BEAST v1.5.4 [84] for at least 100,000,000 steps, sampling every 10,000 steps, and employing an uncorrelated lognormal relaxed clock to allow for rate variation among branches [15,44,85-87]. Rate variation may occur if the two variants evolved at different rates, before or after transmission [15,44]. The substitution model was the GTR model. The underlying demographic model was the Bayesian skyline plot with 10 steps, and was used as a flexible prior on the distribution of the inter-node intervals on the sampled phylogenetic topologies [15,44,85-87]. Convergence was assessed with Tracer v1.5, and all parameter estimates had ESSs of > 300 [15,44,85-87]. Convergence was not achieved when using estimated transmission time as the only prior; the ESSs for the prior and posterior probabilities remained < 100 after 300,000,000 steps [15,44,85-87]. To deal with this issue, a posterior mean rate of substitution prior was estimated from the

posterior mean rate of another dataset, for a fragment of the *env* C2V5 region [15]. This mean rate prior was normally distributed, with a mean of 8.18×10^{-3} substitutions per site per year (standard deviation of 1.15×10^{-3} substitutions per site per year) [15]. The hypervariable regions were cut to be consistent with the original dataset after consultation with the authors [88]. To achieve convergence, our relaxed-clock analysis also required the full-length C2V5 fragment, rather than the part-fragment used in the reference dataset that was missing the 5' end of the C2 region [15].

To determine the sensitivity of our results to the choice of prior, we also analysed the data under a strict molecular clock, calibrating the time of transmission to the same prior as under the relaxed molecular clock, but not enforcing a strong prior on the rate [15,44]. We performed this analysis for C2V5 and our entire 1305 stripped *env* fragment. The mean rate prior from the reference dataset was necessary for the, relaxed clock analysis to converge, but our tMRCA estimate was robust to this choice of prior, as the most important prior for the tMRCA estimate was the time of transmission. Although calibration to the time since transmission may lead to an overestimate of the posterior substitution rate estimate [15], other studies have found that this effect is small for monophyletic infections [6,12]. Both strict-clock and relaxed-clock analyses using the *gag* and *pol* fragments failed to achieve convergence after 300,000,000 steps, and no reference datasets were available for calibration of evolution in these fragments.

Potential N-linked glycosylation site analysis

We compared potential N-linked glycosylation sites (PNLGs) between inferred amino acid sequences for the SGA samples *env* in from P1 and P2, using N-Glycosite [89].

Neutralization and infectivity assays

HIV *env* genes were amplified from reverse transcribed viral RNA, restriction-cloned in pcDNA3.1 (Invitrogen, UK) and co-transfected into 293T cells with an *env* deficient backbone, nl4.3Δ*env* (Dr M. Pizzato, University of Geneva). Virus-containing supernatants were harvested, assayed for reverse transcriptase activity [90], and titrated onto the HIV permissible cell-line, TZM-BL (also known as JC53-BL) using previously described techniques [91] with the following modifications: cell monolayers were fixed with 0.2% glutaraldehyde, stained with an X-gal substrate and air dried. Infected cells were counted with an AID v2.9 EliSpot plate-counter (AID GmbH, Germany). To test serum-mediated neutralizing responses, 400 focus forming units (FFUs) of titrated-pseudovirus were incubated with serial dilutions of heat inactivated autologous sera from participants.

Neutralization was calculated as the percentage-reduction of FFUs compared to virus-only controls.

IFN-γ ELISpot assay

100 μl of 0.5 μg/ml mouse anti-human IFN-γ monoclonal antibody solution (Mabtech, Sweden) was added to each well on an ELISpot plate (Millipore, US). Frozen PBMCs were rapidly defrosted and then pipetted into 10 ml of a solution containing RPMI 1640 and pig skin gelatine (PSG) with added DNase (Sigma, UK). The solution was centrifuged at $300 \times g$ for 5 min. The PBMCs were resuspended in 20 ml of R10 solution and incubated overnight at $37 \pm C$. Cells were then counted and resuspended in a volume of R10 solution to give a final concentration of 5×10^5 cells per 100 μl. The ELISpot plate was washed three times with 200 μl per well of phosphate buffered solution (PBS; Gibco, US) containing 1% FCS. Peptides were added to the appropriate wells, with a final concentration of each peptide being 10 μM. We used overlapping 15 mer peptides covering HIV-1 proteins gag p17 and gag p24 as well as optimal epitopes covering gag, pol, nef and env proteins. 100 μl of PBMC suspension was then added to each well. Duplicate negative controls were prepared, containing R10. Duplicate positive controls were prepared, containing 5 μg/ml PHA-P (Sigma, UK). The plate was incubated for 16 hours at $37 \pm C$. The PBMCs were then discarded and the plate was then washed seven times with PBS. 100 μl of 0.5 μg/ml biotinylated anti-human IFN-γ monoclonal antibody (Mabtech, Sweden) was added to each well. The plate was incubated for 90 min at room temperature. The antibody was then discarded and the plate washed seven times with PBS. 100 μl of 0.5 μg/ml streptavidin-conjugated alkaline phosphatase (ALP; Mabtech, Sweden) was added. The plate was incubated at room temperature for 40 min. The streptavidin-ALP was then discarded and the plate washed seven times with PBS. 100 μl of substrate solution from the ALP conjugate substrate kit (Bio-Rad, US) was added to each well. The plate was incubated at room temperature for 10 min, or until a colour change was noted in the positive control well. The plate was then washed with ordinary tap water and dried. Spots were counted on the AID version 2.9 EliSpot plate-reader. The normalized magnitude of the response (NMOR) was calculated as follows [92]:

$$NMOR = M_{\text{exp}} - (\bar{x}_{\text{neg}} + 3 \times SD_{\text{neg}}) - 50$$

Where M_{exp} is the number of spots in the experimental well, \bar{x}_{neg} is the mean number of spots in the negative control wells, and SD_{neg} is the standard deviation of the negative control wells. NMOR is always a positive integer and all negative values are set to 0.

Additional material

Additional file 1: Images of the entire ML (PhyML) trees for a. env, b. gag and c. pol. Terminal nodes representing day 63 sequences sampled from P1 (blue circles) and P2 (red circles), as well as reference sequences are shown. *Env* sequences for P1 and P2 were sampled by SGA and represent gap-stripped alignments of full-length *gp120*. *Gag* and *pol* fragment sequences were sampled by bacterial cloning.

Additional file 2: Images of the entire Bayesian MCMC based consensus trees for a. env, b. gag and c. pol. Terminal nodes representing day 63 sequences sampled from P1 (blue circles) and P2 (red circles), as well as reference sequences are shown. *Env* sequences for P1 and P2 were sampled by SGA and represent gap-stripped alignments of full-length *gp120*. *Gag* and *pol* fragment sequences were sampled by bacterial cloning.

Additional file 3: Robustness analysis for Bayesian MCMC based approach for a. env C2V5 and b. the entire env fragment. The tMRC estimation analysis was repeated for *env* SGA sequences using a strict molecular clock. The estimated time since transmission was the only prior.

Additional file 4: Neutralization assay results. Neutralization assay results are shown for day 186 post-exposure sera from P1 and P2 against pseudoviruses typed with day 63 P1 and P2 envelopes. Results for two clones from each participant are shown for both autologous and cross-neutralization assays at two serum dilutions, 1:20 and 1:60.

Additional file 5: Results of the infectivity assays. Infectivity assays were used to titre pseudoviruses prior to infection for the neutralization assay. Fold virus dilutions are shown in the legend. The results for the two clones used in the assay shown in Additional File 4 are shown but these results were consistent for the nine clones screened for each participant. Infectivity is corrected against viral reverse transcriptase expression.

Acknowledgements

The authors would like to thank the participants and staff of the St. Mary's Hospital Jefferiss Wing clinic and the participants and staff involved in the SPARTAC trial. The authors would like to thank Mr David English for computing assistance in the rapid processing of MrBayes analyses. The authors would also like to thank the anonymous reviewers of this manuscript for their comments and suggestions. JF is supported by the Medical Research Council. JW and RP are supported by the Wellcome Trust (UK). REP is a NIHR Senior Investigator. SE is supported by the Commonwealth Scholarship Commission. AK is supported by the Royal Society.

SPARTAC Investigators. Trial Steering Committee A Breckenridge (Chair), C Conlon, D Cooper, F Conradie, J Kaldor, M Schechter, P Claydon, P Kaleebu, G Ramjee, F Ssali, G Tambussi, J Weber. **Trial Physician** Sarah Fidler. **Trial Statistician** Abdel Babiker. **Data and Safety Monitoring Committee** A McLaren (in memoriam), V Beral, G Chene, J Hakim. **Central Virology Laboratories and Repositories** Jefferiss Trust Laboratories, Imperial College, London, UK (M McClure, D Muir, I Blain, A Helander, O Erlwien, S Kaye). **Clinical Endpoint Review Committee** N Paton, S Fidler. **Co-ordinating Trial Centres** *Australia*: National Centre in HIV Epidemiology and Clinical Research, University of New South Wales, Sydney (P Grey, D Cooper, T Kelleher, M Law). *UK and Ireland*: MRC Clinical Trials Unit, London (A Babiker, K Porter, P Kelleher, K Boyd, D Johnson, D Nock) **Investigators and Staff at Participating Sites** *Australia*: St Vincent's Hospital, Sydney (D Cooper), Carlton Clinic, Melbourne (J Anderson), 407 Doctors, Sydney, (R McFarlane), Prahran Market Clinic, Melbourne (N Roth), Taylor Square Private Clinic, Sydney (R Finlayson), The Centre Clinic, Melbourne (B Kiem Tee), Sexual Health Centre, Melbourne (T Read), AIDS Medical Unit, Brisbane (M Kelly), Centre for Immunology, Sydney (P Cunningham). *Brazil*: Projeto Praça Onze, Hospital Escola São Francisco de Assis, Universidade federal do Rio de Janeiro, Rio de Janeiro. (M Schechter, R Zajdenverg, M Merçon). *Italy*: Ospedale San Raffaele, Milan (G Tambussi, C Tassan Din, C Ronchetti, G Travi, V Rusconi, G de Bartolo), Ospedale Lazzaro Spallanzani, Roma (G D'Offizi, C Vlassi, A Corpolongo). *South Africa*: Capetown: Desmond Tutu HIV Centre,

Institute of Infectious Diseases, Capetown (R Wood, J Pitt, L-G Bekker, J Aploon, L Fielder, N Killa, T Buhler) *Johannesburg*: Reproductive Health and HIV Research Unit, Bara Clinic, Chris Hani Baragwanath Hospital, Johannesburg (H Rees, J Moyes, S Walaza, K Moitse), Contract Laboratory Services, Johannesburg Hospital, Johannesburg (W Stevens, C Wallis, C Ingram, M Majam) *Kwazulu-Natal*: HIV Prevention Unit, Medical Research Council, Durban (G Ramjee, D Singh, T Mtambo, S Gappoo, H Somaroo, J Moodley, M Mills, A Premraj, N Nozulu, K Naidoo). *Uganda*: MRC/UGanda Virus Research Institute, Entebbe (H Grosskurth, A Kamali, P Kaleebu, J Mugisha, U Bahemuka, F Lyagoba, P Tabuga). *Spain*: Hospital Clinic-IDIBAPS. Univ. of Barcelona. Barcelona (J M Miro, M López-Diequez, F. Agüero, JA Arnaiz, T. Pumarola, M. Plana, M. Tuset, MC Ligeró, C. Gil, T. Gallart, JM Gatell) *UK and Ireland*: Royal Sussex County Hospital, Brighton (M Fisher, L Heald, N Perry, D Pao, D Maitland), St James's Hospital, Dublin (F Mulcahy, G Courtney, D Reidy), Regional Infectious Diseases Unit, Western General Hospital and Genitourinary Dept, Royal Infirmary of Edinburgh, Edinburgh (C Leen, G Scott, L Ellis, S Morris, P Simmonds, T Shaw), Chelsea and Westminster Hospital, London (B Gazzard, D Hawkins, C Higgs, C Mahuma), Homerton Hospital, London (J Anderson, L Muromba), Mortimer Market Centre, London (I Williams, J Turner, D Mullan, D Aldam), North Middlesex Hospital (J Ainsworth, A Waters), Royal Free Hospital (M Johnson, S Kinloch, A Carroll, P Byrne, Z Cuthbertson), St Bartholomew's Hospital, London (C Orkin, J Hand, C De Souza), St Mary's Hospital, London (J Weber, S Fidler, E Thomson, J Fox, K Legg, S Mullaney, A Winston, N Poulter, S Wilson) **Trial Secretariat** D Winogron, S Keeling.

Author details

¹Nuffield Department of Clinical Medicine, Peter Medawar Building for Pathogen Research, Oxford University, South Parks Road, Oxford, OX1 3SY, UK. ²Department of Zoology, Oxford University, South Parks Road, Oxford, OX1 3PS, UK. ³Division of Medicine, Wright Fleming Institute, Imperial College, St. Mary's Hospital, Norfolk Place, Paddington, London W2 1PG, UK. ⁴The James Martin 21st Century School, Peter Medawar Building for Pathogen Research, South Parks Road, Oxford, OX1 3SY, UK. ⁵Oxford NIHR Biomedical Research Centre, Oxford, UK.

Authors' contributions

RP and JF conceived the study; JF, SE and AK designed the study; SE, DB, PF and AD performed the experiments; SE, AK and DB analysed the data; SE and JF wrote the paper; RP, MM, JW, SF and STSC contributed participant information and samples; all authors were involved in drafting this paper; all authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 15 January 2011 Accepted: 7 July 2011 Published: 7 July 2011

References

1. Kawashima Y, Pfafferoth K, Frater J, Matthews P, Payne R, Addo M, Gatanaga H, Fujiwara M, Hachiya A, Koizumi H, *et al*: **Adaptation of HIV-1 to human leukocyte antigen class I.** *Nature* 2009, **458**:641-645.
2. Gaschen B, Taylor J, Yusim K, Foley B, Gao F, Lang D, Novitsky V, Haynes B, Hahn BH, Bhattacharya T, Korber B: **Diversity considerations in HIV-1 vaccine selection.** *Science* 2002, **296**:2354-2360.
3. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, Gupta P, Rinaldo CR, Learn GH, He X, *et al*: **Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection.** *J Virol* 1999, **73**:10489-10502.
4. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, Decker JM, Wang S, Baalwa J, Kraus MH, *et al*: **Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection.** *J Exp Med* 2009, **206**:1273-1289.
5. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, Derdeyn CA, Farmer P, Hunter E, Allen S, *et al*: **Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing.** *J Virol* 2008, **82**:3952-3970.
6. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, *et al*: **Identification and characterization of**

- transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci USA* 2008, **105**:7552-7557.
7. Derdeyn CA, Hunter E: **Viral characteristics of transmitted HIV.** *Curr Opin HIV AIDS* 2008, **3**:16-21.
 8. **The Global HIV/AIDS Vaccine Enterprise: Scientific Strategic Plan.** *PLoS Med* 2005, **2**:111-121.
 9. Edwards CT, Holmes EC, Wilson DJ, Viscidi RP, Abrams EJ, Phillips RE, Drummond AJ: **Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1.** *BMC Evol Biol* 2006, **6**:28.
 10. Haaland RE, Hawkins PA, Salazar-Gonzalez J, Johnson A, Tichacek A, Karita E, Manigart O, Mulenga J, Keele BF, Shaw GM, *et al*: **Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1.** *PLoS Pathog* 2009, **5**:e1000274.
 11. Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping LH, Athreya GS, Treurnicht FK, Keele BF, Wood N, *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**:3556-3567.
 12. Li H, Bar KJ, Wang S, Decker JM, Chen Y, Sun C, Salazar-Gonzalez JF, Salazar MG, Learn GH, Morgan CJ, *et al*: **High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men.** *PLoS Pathog* 2010, **6**:e1000890.
 13. Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, Grayson T, Sun C, Wang S, Learn GH, Morgan CJ, *et al*: **Wide variation in the multiplicity of HIV-1 infection among injection drug users.** *J Virol* 2010, **84**:6241-6247.
 14. Hollingsworth TD, Anderson RM, Fraser C: **HIV-1 transmission, by stage of infection.** *J Infect Dis* 2008, **198**:687-693.
 15. Lemey P, Rambaut A, Pybus OG: **HIV evolutionary dynamics within and among hosts.** *AIDS Rev* 2006, **8**:125-140.
 16. Drummond AJ, Rambaut A, Shapiro B, Pybus OG: **Bayesian coalescent inference of past population dynamics from molecular sequences.** *Mol Biol Evol* 2005, **22**:1185-1192.
 17. Zhang LQ, MacKenzie P, Cleland A, Holmes EC, Brown AJ, Simmonds P: **Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection.** *J Virol* 1993, **67**:3345-3356.
 18. Zhu T, Mo H, Wang N, Nam DS, Cao Y, Koup RA, Ho DD: **Genotypic and phenotypic characterization of HIV-1 patients with primary infection.** *Science* 1993, **261**:1179-1181.
 19. Butler DM, Delpont W, Kosakovsky Pond SL, Lakdawala MK, Cheng PM, Little SJ, Richman DD, Smith DM: **The origins of sexually transmitted HIV among men who have sex with men.** *Sci Transl Med* 2011, **3**:18re11.
 20. Bull M, Learn G, Genowati I, McKernan J, Hitti J, Lockhart D, Tapia K, Holte S, Dragavon J, Coombs R, *et al*: **Compartmentalization of HIV-1 within the female genital tract is due to monotypic and low-diversity variants not distinct viral populations.** *PLoS One* 2009, **4**:e7122.
 21. Diem K, Nickle DC, Motoshige A, Fox A, Ross S, Mullins JL, Corey L, Coombs RW, Krieger N: **Male genital tract compartmentalization of human immunodeficiency virus type 1 (HIV).** *AIDS Res Hum Retroviruses* 2008, **24**:561-571.
 22. Derdeyn CA, Decker JM, Bibollet-Ruche F, Mokili JL, Muldoon M, Denham SA, Heil ML, Kasolo F, Musonda R, Hahn BH, *et al*: **Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission.** *Science* 2004, **303**:2019-2022.
 23. Chohan B, Lang D, Sagar M, Korber B, Lavreys L, Richardson B, Overbaugh J: **Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels.** *J Virol* 2005, **79**:6528-6531.
 24. Liu Y, Curlin ME, Diem K, Zhao H, Ghosh AK, Zhu H, Woodward AS, Maenza J, Stevens CE, Stekler J, *et al*: **Env length and N-linked glycosylation following transmission of human immunodeficiency virus Type 1 subtype B viruses.** *Virology* 2008, **374**:229-233.
 25. Frost SD, Liu Y, Pond SL, Chappey C, Wrin T, Petropoulos CJ, Little SJ, Richman DD: **Characterization of human immunodeficiency virus type 1 (HIV-1) envelope variation and neutralizing antibody responses during transmission of HIV-1 subtype B.** *J Virol* 2005, **79**:6523-6527.
 26. Keele BF, Li H, Learn GH, Hraber P, Giorgi EE, Grayson T, Sun C, Chen Y, Yeh WW, Letvin NL, *et al*: **Low-dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1.** *J Exp Med* 2009, **206**:1117-1134.
 27. Rybarczyk BJ, Montefiori D, Johnson PR, West A, Johnston RE, Swanstrom R: **Correlation between env V1/V2 region diversification and neutralizing antibodies during primary infection by simian immunodeficiency virus sm in rhesus macaques.** *J Virol* 2004, **78**:3561-3571.
 28. Highleyman L: **Detuned assay used to track recent infections.** *BETA* 1999, **12**:6-78.
 29. Kothe D, Byers RH, Caudill SP, Satten GA, Janssen RS, Hannon WH, Mei JV: **Performance characteristics of a new less sensitive HIV-1 enzyme immunoassay for use in estimating HIV seroprevalence.** *J Acquir Immune Defic Syndr* 2003, **33**:625-634.
 30. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, Heldebrandt C, Smith R, Conrad A, Kleinman SH, Busch MP: **Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection.** *Aids* 2003, **17**:1871-1879.
 31. Fiebig EW, Heldebrandt CM, Smith RL, Conrad AJ, Delwart EL, Busch MP: **Intermittent low-level viremia in very early primary HIV-1 infection.** *J Acquir Immune Defic Syndr* 2005, **39**:133-137.
 32. Little SJ, McLean AR, Spina CA, Richman DD, Havlir DV: **Viral dynamics of acute HIV-1 infection.** *J Exp Med* 1999, **190**:841-850.
 33. Clark SJ, Saag MS, Decker WD, Campbell-Hill S, Roberson JL, Veldkamp PJ, Kappes JC, Hahn BH, Shaw GM: **High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection.** *N Engl J Med* 1991, **324**:954-960.
 34. Stafford MA, Corey L, Cao Y, Daar ES, Ho DD, Perelson AS: **Modeling plasma virus concentration during primary HIV infection.** *J Theor Biol* 2000, **203**:285-301.
 35. Wilson DP, Law MG, Grulich AE, Cooper DA, Kaldor JM: **Relation between HIV viral load and infectiousness: a model-based analysis.** *Lancet* 2008, **372**:314-320.
 36. Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, Rock D, Falloon J, Davey RT, Dewar RL, *et al*: **Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis.** *J Clin Microbiol* 2005, **43**:406-413.
 37. Jordan MR, Kearney M, Palmer S, Shao W, Maldarelli F, Coakley EP, Chappey C, Wanke C, Coffin JM: **Comparison of standard PCR/cloning to single genome sequencing for analysis of HIV-1 populations.** *J Virol Methods* 2010, **168**:114-120.
 38. Fidler S, Oxenius A, Brady M, Clarke J, Cropley I, Babiker A, Zhang HT, Price D, Phillips R, Weber J: **Virological and immunological effects of short-course antiretroviral therapy in primary HIV infection.** *Aids* 2002, **16**:2049-2054.
 39. Pillay D, Rambaut A, Geretti AM, Brown AJ: **HIV phylogenetics.** *BMJ* 2007, **335**:460-461.
 40. Liu J, Keele BF, Li H, Keating S, Norris PJ, Carville A, Mansfield KG, Tomaras GD, Haynes BF, Kolodkin-Gal D, *et al*: **Low-dose mucosal simian immunodeficiency virus infection restricts early replication kinetics and transmitted virus variants in rhesus monkeys.** *J Virol* 2010, **84**:10406-10412.
 41. Liu W, Worobey M, Li Y, Keele BF, Bibollet-Ruche F, Guo Y, Goepfert PA, Santiago ML, Ndjango JB, Neel C, *et al*: **Molecular ecology and natural history of simian foamy virus infection in wild-living chimpanzees.** *PLoS Pathog* 2008, **4**:e1000097.
 42. Salemi M, Burkhardt BR, Gray RR, Ghaffari G, Sleasman JW, Goodenow MM: **Phylogenetics of HIV-1 in lymphoid and non-lymphoid tissues reveals a central role for the thymus in emergence of CXCR4-using quasisppecies.** *PLoS One* 2007, **2**:e950.
 43. Hillis DM, Huelsenbeck JP: **Signal, noise, and reliability in molecular phylogenetic analyses.** *J Hered* 1992, **83**:189-195.
 44. Salemi M, Vandamme A-M, Lemey P: *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing.* 2 edition. Cambridge, UK; New York: Cambridge University Press; 2009.
 45. Strimmer K, von Haeseler A: **Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment.** *Proc Natl Acad Sci USA* 1997, **94**:6815-6819.
 46. Hurst LD: **Fundamental concepts in genetics: genetics and the understanding of selection.** *Nat Rev Genet* 2009, **10**:83-93.
 47. Mascola JR, Stiegler G, VanCott TC, Katinger H, Carpenter CB, Hanson CE, Beary H, Hayes D, Frankel SS, Bix DL, Lewis MG: **Protection of macaques against vaginal transmission of a pathogenic HIV-1/SIV chimeric virus by passive infusion of neutralizing antibodies.** *Nat Med* 2000, **6**:207-210.

48. Simmonds P, Balfe P, Ludlam CA, Bishop JO, Brown AJ: **Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1.** *J Virol* 1990, **64**:5840-5850.
49. Zhang M, Gaschen B, Blay W, Foley B, Haigwood N, Kuiken C, Korber B: **Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin.** *Glycobiology* 2004, **14**:1229-1246.
50. Valenzuela A, Blanco J, Krust B, Franco R, Hovanessian AG: **Neutralizing antibodies against the V3 loop of human immunodeficiency virus type 1 gp120 block the CD4-dependent and -independent binding of virus to cells.** *J Virol* 1997, **71**:8289-8298.
51. Rizzuto CD, Wyatt R, Hernandez-Ramos N, Sun Y, Kwong PD, Hendrickson WA, Sodroski J: **A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding.** *Science* 1998, **280**:1949-1953.
52. Thali M, Olshevsky U, Furman C, Gabuzda D, Posner M, Sodroski J: **Characterization of a discontinuous human immunodeficiency virus type 1 gp120 epitope recognized by a broadly reactive neutralizing human monoclonal antibody.** *J Virol* 1991, **65**:6188-6193.
53. Low AJ, Dong W, Chan D, Sing T, Swanstrom R, Jensen M, Pillai S, Good B, Harrigan PR: **Current V3 genotyping algorithms are inadequate for predicting X4 co-receptor usage in clinical isolates.** *AIDS* 2007, **21**:F17-24.
54. O'Brien SJ, Gao X, Carrington M: **HLA and AIDS: a cautionary tale.** *Trends Mol Med* 2001, **7**:379-381.
55. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ: **HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage.** *Science* 1999, **283**:1748-1752.
56. Brumme ZL, Tao I, Szeto S, Brumme CJ, Carlson JM, Chan D, Kadie C, Frahm N, Brander C, Walker B, *et al*: **Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection.** *AIDS* 2008, **22**:1277-1286.
57. Fischer W, Ganusov VV, Giorgi EE, Hraber PT, Keele BF, Leitner T, Han CS, Gleasner CD, Green L, Lo CC, *et al*: **Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing.** *PLoS One* 2010, **5**:e12303.
58. Hansen SG, Vieville C, Whizin N, Coyne-Johnson L, Siess DC, Drummond DD, Legasse AW, Axthelm MK, Oswald K, Trubey CM, *et al*: **Effector memory T cell responses are associated with protection of rhesus monkeys from mucosal simian immunodeficiency virus challenge.** *Nat Med* 2009, **15**:293-299.
59. Langlade-Demoyen P, Ngo-Giang-Huong N, Ferchal F, Oksenhendler E: **Human immunodeficiency virus (HIV) nef-specific cytotoxic T lymphocytes in noninfected heterosexual contact of HIV-infected patients.** *J Clin Invest* 1994, **93**:1293-1297.
60. Clerici M, Giorgi JV, Chou CC, Gudeman VK, Zack JA, Gupta P, Ho HN, Nishanian PG, Berzofsky JA, Shearer GM: **Cell-mediated immune response to human immunodeficiency virus (HIV) type 1 in seronegative homosexual men with recent sexual exposure to HIV-1.** *J Infect Dis* 1992, **165**:1012-1019.
61. Rowland-Jones S, Sutton J, Ariyoshi K, Dong T, Gotch F, McAdam S, Whitby D, Sabally S, Gallimore A, Corrah T, *et al*: **HIV-specific cytotoxic T-cells in HIV-exposed but uninfected Gambian women.** *Nat Med* 1995, **1**:59-64.
62. Gao X, Nelson GW, Karacki P, Martin MP, Phair J, Kaslow R, Goedert JJ, Buchbinder S, Hoots K, Vlahov D, *et al*: **Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS.** *N Engl J Med* 2001, **344**:1668-1675.
63. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferoth KJ, Hilton L, *et al*: **Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA.** *Nature* 2004, **432**:769-775.
64. Kaslow RA, Duquesnoy R, VanRaden M, Kingsley L, Marrari M, Friedman H, Su S, Saah AJ, Detels R, Phair J, *et al*: **A1, Cw7, B8, DR3 HLA antigen combination associated with rapid decline of T-helper lymphocytes in HIV-1 infection. A report from the Multicenter AIDS Cohort Study.** *Lancet* 1990, **335**:927-930.
65. Frater AJ, Edwards CT, McCarthy N, Fox J, Brown H, Milicic A, Mackie N, Pillay T, Drijfhout JW, Dustan S, *et al*: **Passive sexual transmission of human immunodeficiency virus type 1 variants and adaptation in new hosts.** *J Virol* 2006, **80**:7226-7234.
66. Streeck H, Jolin JS, Qi Y, Yassine-Diab B, Johnson RC, Kwon DS, Addo MM, Brumme C, Routy JP, Little S, *et al*: **Human immunodeficiency virus type 1-specific CD8+ T-cell responses during primary infection are major determinants of the viral set point and loss of CD4+ T cells.** *J Virol* 2009, **83**:7641-7648.
67. Frater AJ, Brown H, Oxenius A, Gunthard HF, Hirschel B, Robinson N, Leslie AJ, Payne R, Crawford H, Prendergast A, *et al*: **Effective T-cell responses select human immunodeficiency virus mutants and slow disease progression.** *J Virol* 2007, **81**:6742-6751.
68. Rambaut A: **Se-Al. Sequence Alignment Editor v2.0a11.** *Book Se-Al Sequence Alignment Editor v2.0a11 (Editor ed. ^eds.)* City; 2002.
69. Helseth E, Kowalski M, Gabuzda D, Olshevsky U, Haseltine W, Sodroski J: **Rapid complementation assays measuring replicative potential of human immunodeficiency virus type 1 envelope glycoprotein mutants.** *J Virol* 1990, **64**:2416-2420.
70. Wyatt R, Sullivan N, Thali M, Repke H, Ho D, Robinson J, Posner M, Sodroski J: **Functional and immunologic characterization of human immunodeficiency virus type 1 envelope glycoproteins containing deletions of the major variable regions.** *J Virol* 1993, **67**:4557-4565.
71. Truckenmiller ME, Kulaga H, Coggiano M, Wyatt R, Snyder SH, Sweetnam PM: **Human cortical neuronal cell line: a model for HIV-1 infection in an immature neuronal system.** *AIDS Res Hum Retroviruses* 1993, **9**:445-453.
72. Williams MF, Eisele DW, Wyatt SH: **Neck needle foreign bodies in intravenous drug abusers.** *Laryngoscope* 1993, **103**:59-63.
73. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**:502-504.
74. Jansson M, Popovic M, Karlsson A, Cocchi F, Rossi P, Albert J, Wigzell H: **Sensitivity to inhibition by beta-chemokines correlates with biological phenotypes of primary HIV-1 isolates.** *Proc Natl Acad Sci USA* 1996, **93**:15382-15387.
75. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ: **Jalview Version 2—a multiple sequence alignment editor and analysis workbench.** *Bioinformatics* 2009, **25**:1189-1191.
76. Clamp M, Cuff J, Searle SM, Barton GJ: **The Jalview Java alignment editor.** *Bioinformatics* 2004, **20**:426-427.
77. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**:696-704.
78. FigTree v1.3.1. [<http://tree.bio.ed.ac.uk/software/figtree>].
79. Valenzuela A, Blanco J, Callebaut C, Jacotot E, Lluís C, Hovanessian AG, Franco R: **Adenosine deaminase binding to human CD26 is inhibited by HIV-1 envelope glycoprotein gp120 and viral particles.** *J Immunol* 1997, **158**:3721-3729.
80. Büchen-Osmond CE: **00.061.1.06.001. Human immunodeficiency virus 1.** *ICTVdB - The Universal Virus Database, version 3* 2004.
81. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17**:754-755.
82. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
83. Tracer v1.5. [<http://tree.bio.ed.ac.uk/software/tracer>].
84. Drummond AJ, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees.** *BMC Evol Biol* 2007, **7**:214.
85. Drummond AJ, Ho SY, Phillips MJ, Rambaut A: **Relaxed phylogenetics and dating with confidence.** *PLoS Biol* 2006, **4**:e88.
86. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ: **Episodic sexual transmission of HIV revealed by molecular phylodynamics.** *PLoS Med* 2008, **5**:e50.
87. Drummond A, Pybus OG, Rambaut A: **Inference of viral evolutionary rates from molecular sequences.** *Adv Parasitol* 2003, **54**:331-358.
88. Lemy P: *Personal Communication* April 9, 2011.
89. Seminari E, Cocchi L, Antoniazzi E, Giacchino R, Maserati R: **Clinical significance of diplopia in HIV infection. Assessment of a personal case-load and review of the literature.** *Minerva Med* 1996, **87**:515-523.

90. Pizzato M, Erlwein O, Bonsall D, Kaye S, Muir D, McClure MO: **A one-step SYBR Green I-based product-enhanced reverse transcriptase assay for the quantitation of retroviruses in cell culture supernatants.** *J Virol Methods* 2009, **156**:1-7.
91. Montefiori DC: **Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays.** *Curr Protoc Immunol* 2005, **Chapter 12**; Unit 12.11.
92. Duda A, Lee-Turner L, Fox J, Robinson N, Dustan S, Kaye S, Fryer H, Carrington M, McClure M, McLean AR, *et al*: **HLA-associated clinical progression correlates with epitope reversion rates in early human immunodeficiency virus infection.** *J Virol* 2009, **83**:1228-1239.

doi:10.1186/1742-4690-8-54

Cite this article as: English *et al*: Phylogenetic analysis consistent with a clinical history of sexual transmission of HIV-1 from a single donor reveals transmission of highly distinct variants. *Retrovirology* 2011 **8**:54.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

