

SCIENTIFIC REPORTS



OPEN

Differential evolution of a CXCR4-using HIV-1 strain in CCR5wt/wt and CCR5 Δ 32/ Δ 32 hosts revealed by longitudinal deep sequencing and phylogenetic reconstruction

Anh Q. Le¹, Jeremy Taylor², Winnie Dong², Rosemary McCloskey², Conan Woods², Ryan Danroth¹, Kanna Hayashi^{2,3}, M.-J. Milloy^{2,3}, Art F. Y. Poon^{1,2,3,*} & Zabrina L. Brumme^{1,2,*}

Rare individuals homozygous for a naturally-occurring 32 base pair deletion in the CCR5 gene (CCR5 Δ 32/ Δ 32) are resistant to infection by CCR5-using ("R5") HIV-1 strains but remain susceptible to less common CXCR4-using ("X4") strains. The evolutionary dynamics of X4 infections however, remain incompletely understood. We identified two individuals, one CCR5wt/wt and one CCR5 Δ 32/ Δ 32, within the Vancouver Injection Drug Users Study who were infected with a genetically similar X4 HIV-1 strain. While early-stage plasma viral loads were comparable in the two individuals (~ 4.5 – $5 \log_{10}$ HIV-1 RNA copies/ml), CD4 counts in the CCR5wt/wt individual reached a nadir of < 20 CD4 cells/mm³ within 17 months but remained > 250 cells/mm³ in the CCR5 Δ 32/ Δ 32 individual. Ancestral phylogenetic reconstructions using longitudinal envelope-V3 deep sequences suggested that both individuals were infected by a single transmitted/founder (T/F) X4 virus that differed at only one V3 site (codon 24). While substantial within-host HIV-1 V3 diversification was observed in plasma and PBMC in both individuals, the CCR5wt/wt individual's HIV-1 population gradually reverted from 100% X4 to $\sim 60\%$ R5 over ~ 4 years whereas the CCR5 Δ 32/ Δ 32 individual's remained consistently X4. Our observations illuminate early dynamics of X4 HIV-1 infections and underscore the influence of CCR5 genotype on HIV-1 V3 evolution.

Entry of human immunodeficiency virus type-1 (HIV-1) into target cells occurs via binding of the viral envelope protein gp120 to the host CD4 receptor¹ followed by binding to chemokine coreceptors CCR5 or CXCR4 on the host cell surface^{2,3}. HIV-1 strains that utilize CCR5 or CXCR4 are termed "R5" and "X4" respectively; those capable of utilizing either coreceptor are termed "R5/X4" (or dual-tropic)⁴. As its principal genetic determinants lie within the third variable (V3) loop of envelope gp120^{5,6}, HIV-1 coreceptor usage can be determined phenotypically using cell-culture based assays that express patient-derived envelope proteins⁷ or genotypically using algorithms trained on large linked V3 sequence/phenotype datasets^{8,9}.

R5 strains predominate globally as well as during all infection stages¹⁰. R5 strains are also preferentially transmitted^{10–12}. Before the availability of antiretroviral therapies to treat HIV-1, approximately 50% of individuals who acquired an R5 HIV-1 subtype B strain at transmission would continue to harbor R5 variants throughout their disease course, whereas in the remaining 50%, X4 variants would eventually

¹Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada. ²British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada. ³Department of Medicine, University of British Columbia, Vancouver, Canada.

*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Z.L.B. (email: zbrumme@sfu.ca)

emerge alongside their R5 counterparts over a timeline of years^{13–15}. Referred to as the “coreceptor switch”, this phenomenon is associated with an accelerated clinical progression to AIDS^{13,14,16}, though it remains somewhat unclear whether X4 strains cause, or emerge as a consequence, of immune depletion¹⁷.

In contrast, acquisition of X4 HIV-1 variants at transmission is less common: recent studies estimate that X4/dual tropic strains comprise between 3–23% of primary infections^{18–23}. Though some evidence suggests X4 virus transmission is associated with more rapid clinical progression^{24–26}, such infections remain generally less well understood. Moreover, a naturally-occurring 32 base pair deletion in the CCR5 gene (“CCR5Δ32”) that results in a non-functional CCR5 protein²⁷ also modulates HIV-1 acquisition risk. Specifically, rare individuals homozygous for this deletion (“CCR5Δ32/Δ32”), who comprise approximately 1% of individuals of European descent^{27,28}, are resistant to infection by R5 HIV-1 strains but remain susceptible to infection by X4 or dual tropic strains^{24,27,29–42}. Resistance to HIV-1 by the CCR5Δ32/Δ32 genotype is also demonstrated by the “Berlin patient”, the first (and only) individual cured of HIV-1 infection to date, via a stem cell transplant from a histocompatibility-matched donor who additionally carried the CCR5Δ32/Δ32 mutation⁴³.

Major advances have recently been made in our understanding of HIV-1 transmission^{11,44–46}. We now know that this event is characterized by a severe genetic bottleneck where an estimated 80% of heterosexual infections are productively initiated by a single transmitted/founder (T/F) variant^{44,46}, whereas infection via injection drug use is generally initiated by more than one closely-related T/F virus⁴⁵. Sequence reconstruction of T/F viruses is traditionally performed by computing a consensus sequence from single-template (*e.g.* clonal, deep-sequenced or single-genome amplified) HIV-1 sequences sampled from plasma shortly after infection^{44,45}. Alternatively, phylogenetic ancestral reconstruction techniques can be applied to longitudinal single-template HIV-1 sequence datasets—even those sampled weeks or months following infection—to estimate infection dates, reconstruct T/F virus sequences and study within-host HIV-1 evolution in detail^{11,47–49}. For example, phylogenetic techniques have been applied to longitudinal within-host HIV-1 V3 deep sequence data to reconstruct the timing and emergence of X4 lineages in patients who underwent a coreceptor switch^{48,50}. In the present study, we apply longitudinal next-generation sequencing and phylogenetic approaches to study a far more rare occurrence: a case where two individuals—one CCR5wt/wt and one CCR5Δ32/Δ32—were infected with a highly genetically similar X4 HIV-1 strain.

Methods

Vancouver Injection Drug Users Study (VIDUS). Founded in 1996, the original Vancouver Injection Drug Users Study (VIDUS) was a longitudinal cohort comprised of 1603 active injection drug users 18 years or older recruited from the Greater Vancouver area through self-referral and street outreach⁵¹. At baseline and semi-annual follow-up visits, participants completed a structured interviewer-administered questionnaire and provided a blood sample that was separated into plasma and peripheral blood mononuclear cells (PBMC) and stored at -80°C until use. PBMC pellets were frozen directly (*i.e.* not cryopreserved); as such, cell separation and viral outgrowth assays were not possible. All individuals completed an HIV test at baseline; HIV-seronegative individuals were tested at each biannual study visit until they seroconverted or until the end of the study. Of 1603 VIDUS participants recruited, 325 (20.3%) were HIV-1 positive (seroprevalent) at study entry whereas 141 (8.8%) seroconverted during follow-up; all other participants did not register an HIV-positive test during follow-up. The present study made use of available bulk plasma HIV-1 RNA and/or DNA sequences spanning Gag, Integrase, V3 and Nef from 115 (of 141, 82.3%) seroconverters and 124 (of 325, 38.2%) seroprevalent VIDUS participants (total 239).

Ethics statement. Written informed consent was obtained from all participants of the VIDUS cohort. The study was approved by the institutional review boards at Providence Health Care/University of British Columbia and Simon Fraser University, and the study was carried out in accordance with the approved guidelines.

Amplification and bulk sequencing of HIV-1 RNA and DNA from VIDUS participants. Total nucleic acids were extracted from plasma and PBMC pellets collected from VIDUS participants using standard methods. HIV-1 Gag, Integrase, V3 and Nef were amplified by nested RT-PCR (for HIV-1 RNA) or nested PCR (for HIV-1 DNA) using the Invitrogen SuperScript III One-Step RT-PCR System and/or Roche Expand High Fidelity PCR System respectively, using primers optimized for HIV-1 subtype B. Amplicons were bidirectionally sequenced on a 3130xl or 3730xl automated DNA sequencer (Applied Biosystems). Chromatograms were analyzed using Sequencher v5.0.1 (Genecodes) or custom software RECall⁵² with nucleotide mixtures called if the height of the secondary peak exceeded 25% of the dominant peak height (Sequencher) or 20% of the dominant peak area (RECall). Alignment to the HIV-1 subtype B reference strain HXB2 (for Gag, Integrase and Nef) or a modified subtype B reference sequence (for V3) was done using an in-house alignment tool based on the HyPhy platform⁵³. Maximum likelihood phylogenetic trees were constructed using PhyML 3.0⁵⁴. Patristic (tip-to-tip) genetic distances, expressed in terms of substitutions per nucleotide site (sub/nt site), were extracted from maximum-likelihood Newick treefiles using PATRISTIC⁵⁵. Trees were visualized using Figtree v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Identification and host genetic characterization of the participant pair. Phylogenetic analysis of population-level bulk HIV-1 sequences identified two VIDUS participants whose HIV-1 sequences exhibited the shortest patristic (tip-to-tip) genetic distances in the cohort, for all HIV-1 genes analyzed (see results). CCR5 Δ 32 genotyping of these individuals was performed as described previously⁵⁶. Briefly, a ~172 bp region spanning the deletion site was amplified by nested PCR from plasma and/or PBMC-derived DNA and visualized on a 2% agarose gel. To confirm the genotype, 2nd round amplicons were bidirectionally sequenced and chromatograms were visually assessed for length and the presence of the prolonged mixed-base motif characteristic of heterozygous CCR5wt/ Δ 32 genotypes. In doing so, one individual was identified as homozygous CCR5wt/wt and the other as CCR5 Δ 32/ Δ 32. As the latter is a rare genotype, it was confirmed by testing all specimens collected longitudinally from this individual, all with the same result. High resolution HLA class I typing was performed by sequence-based methods⁵⁷.

Clinical estimated dates of infection were calculated as the midpoint between the last HIV-negative and first positive sample, yielding estimates of March 2000 for the CCR5wt/wt individual and August 2001 for the CCR5 Δ 32/ Δ 32 individual. As this was a longitudinal study, we arbitrarily designated the CCR5 Δ 32/ Δ 32's estimated infection date as "time-zero" and expressed all other timepoints/specimens relative to this date. For the CCR5wt/wt individual, paired plasma and PBMCs were available at -13, -7, -1, and +35 months, while for the CCR5 Δ 32/ Δ 32 individual, plasma samples were available at +5 months, and paired plasma/PBMCs at +6, and +12 months.

The CCR5wt/wt individual, initially antiretroviral naïve, began highly active antiretroviral therapy (HAART) in late August 2001 and remained intermittently on HAART over the course of study followup. The CCR5 Δ 32/ Δ 32 individual remained antiretroviral-naïve over the course of study followup.

Longitudinal deep-sequencing of HIV-1 V3 RNA and DNA. Prior to deep-sequencing, the V3 region was re-amplified in triplicate from all plasma and PBMC-derived nucleic acid extracts obtained from each individual. Nested second round amplification was performed using forward and reverse primers incorporating one of 12 multiplex identifier (MID) tags and a linker sequence at the 5' end and visualized on a 1% agarose gel. Amplicons were quantified with the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) on a DTX 880 Multimode Detector (Beckman Coulter), pooled in equal proportions, purified, re-quantified, and deep-sequenced using the GS Junior Titanium Sequencing Kit on a GS Junior instrument (Roche/454). To avoid low-level, intra-run sequence cross-contamination by genetically similar amplicons, we sequenced each sample on a separate GS-Junior run (as each run typically included 24 V3 amplicons, this means that each run contained one amplicon from the present study and 23 V3 amplicons from patients unrelated to the present study). The one exception was sample +35M^{Plasma/PBMC} from the CCR5wt/wt individual, where data are derived from an initial run that included other study samples. Inclusion of data from this sample was deemed appropriate after quality-control experiments confirmed that HIV-1 sequences and their distributions obtained from separate vs. combined runs were highly concordant (not shown).

Processing of deep sequencing data. Raw sequences were processed, aligned, and trimmed to a modified HIV-1 HXB2 V3 reference standard (HXB2 gp120 codons 296–331) using an iterative process as described previously⁵⁸. Briefly, identical sequences were collapsed and annotated with read counts. Sequences were discarded if the MID or primer sequence was a mismatch to the one assigned to the sample or the sequence did not align to the V3 reference standard. Sequences that were identical except for 1-2 gap characters (attributable to erroneous indels introduced during sequencing) were merged, and read counts updated. The remaining sequences were re-aligned to generate a sample-specific consensus sequence, which was used as the reference standard in subsequent steps.

After realigning all sequences to the specimen-specific consensus, any gap characters followed by ≥ 3 instances of the same nucleotide were replaced with that nucleotide (to correct for the GS-Junior platform's difficulty in sequencing homopolymer repeats), and insertions/deletions (indels) were moved to be in-frame. Identical sequences were again merged and read counts updated. A multiple alignment was performed on all remaining sequences, and sequences observed at frequencies of $< 1\%$ that still contained a single gap character were discarded. To remove any low-level sequence contamination from other patient-derived amplicons sequenced in the same run, an intra-run cross contamination check was performed. To do this, the 5 most frequent sequences within each run (that were observed at a $> 10\%$ overall prevalence) were identified. Every sequence in our sample was then compared against this list and discarded if it represented an exact match. Lastly, nucleotide sequences with read counts of ≤ 2 , those not divisible by 3 after removal of gap characters, those not encoding cysteines (C) as the starting and final V3 residues, and those < 96 or > 189 base pairs were discarded as invalid prior to final analysis⁵⁹. Overall, 8 to 55% of raw sequences were discarded as a result of this processing pipeline.

Ancestral phylogenetic reconstructions. Ancestral phylogenetic reconstructions of intra-host HIV-1 evolution, including the estimation of transmitted/founder (T/F) sequences and dates, were performed using deep sequence data from the three plasma specimens collected closest to "time zero": -13 M, -7 M, and -1 M for the CCR5wt/wt individual and +5 M, +6 M, and +12 M for the CCR5 Δ 32/ Δ 32 individual. To maximize information incorporated into the phylogeny, a ~250 bp sequence encompassing V3 and flanking regions (mapping approximately to HXB2 genomic nucleotides 7086–7336) was used.

Processing of V3 deep sequence data for ancestral reconstructions was done using an in-house pipeline described previously⁴⁸. Briefly, raw sequences were grouped by their unique MID tag, and nucleotides with low quality scores (as reported by Roche GS-Junior software) were trimmed from the 5' and 3' ends. Identical sequences were temporarily collapsed and annotated with read counts. These were subsequently aligned using a custom sequence alignment algorithm in HyPhy⁵³ that adjusts for the high indel rates observed with the GS-Junior platform by aligning all three reading frames to a reference protein standard spanning HXB2 gp120 codons 278–375. This algorithm assumes that a true V3 sequence will encode a single open reading frame, with any frameshifts attributable to erroneous indels introduced during sequencing. Aligned sequences were then re-expanded by their read counts and annotated with sample dates expressed in terms of days elapsed since January 1, 1990.

Time-calibrated phylogenies were reconstructed using Bayesian Evolutionary Analysis Sampling Trees (BEAST) v1.6.1⁶⁰ using parameters described previously with some modifications⁴⁸. Briefly, 100 sequences were randomly sampled from each timepoint, for a total of 600 sequences included in each reconstruction. These 600 sequences were aligned using MUSCLE v3.8.31⁶¹ and alignments were manually curated using Se-AL (<http://tree.bio.ed.ac.uk/software/seal/>). Alignments were converted into a BEAST XML file with the following parameter settings: Tamura-Nei⁶² nucleotide substitution model; uncorrelated log-normal molecular clock; Bayesian skyline model with 5 population size classes; and a chain length of 10^8 with chain states written to log files at intervals of 10^5 with a burn-in period of 2×10^7 (20%). The resulting trees were then thinned down to 100 sampled at regular intervals. Convergence of chain states was assessed using Gelman and Rubin's convergence diagnostic implemented in the R package *coda*⁶³. For each tree, a Muse-Gaut codon substitution model crossed with a general time-reversible model of nucleotide substitution (implemented in HyPhy⁵³) was fit to every tree. Ancestral sequences were generated by sampling 100 character states from the posterior distributions reconstructed at every node of the tree. In total, 10 independent ancestral reconstructions, each randomly sampling 100 sequences per timepoint for a total of 600 sequences, were performed.

Assessing V3 sequence divergence and diversity. Within-host HIV-1 genetic divergence over time was assessed by calculating patristic (tip-to-tip) phylogenetic distances between each host's reconstructed T/F virus and all the sequences observed in their plasma and PBMC specimens thereafter, taking into consideration the frequency of each sequence. Within-host HIV-1 diversity, calculated as per-codon differences in Shannon entropy, were calculated from V3 amino acid alignments from the earliest and latest plasma timepoints using Entropy-Two (<http://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy.html>) using 1000 randomizations with replacement.

Inference of HIV-1 coreceptor usage. HIV-1 coreceptor usage (R5 vs. X4) was predicted from bulk and deep HIV-1 V3 sequences using $\text{geno2pheno}_{[\text{coreceptor}]}$ (g2p)⁸. This algorithm assigns each sequence a false-positive rate (FPR) that represents the probability of classifying an R5-virus falsely as X4. We employed a false positive rate (FPR) of 5.0%, meaning that V3 sequences with $\text{FPR} \leq 5.0\%$ and $>5.0\%$ were classified as X4 and R5, respectively.

Results

Acquisition of a similar X4-using HIV-1 strain in CCR5wt/wt and CCR5 Δ 32/ Δ 32 hosts. Maximum-likelihood phylogenies were constructed using one bulk HIV-1 plasma RNA or PBMC DNA Gag, Integrase, V3, and Nef sequence per individual for 239 VIDUS participants (Fig. 1 and Supplementary Figure S1). Computation of patristic (tip-to-tip) genetic distances within these phylogenies consistently identified a participant pair who exhibited the lowest overall distances for all HIV-1 genes examined: these were 0.0027 substitutions per nucleotide site (sub/nt site) in gag (compared to a cohort median of 0.064 [IQR 0.055–0.070]), 0.0023 for integrase (cohort median of 0.034 [IQR 0.025–0.041]), 0.010 for V3 (cohort median of 0.087 [IQR 0.056–0.12]) and 0.023 for nef (cohort median of 0.10 [IQR 0.081–0.11]). The overall prevalence of X4 HIV-1 among all VIDUS seroconverters and seroprevalent participants studied, inferred from bulk V3 sequences, was 13% (12% among seroconverters sequenced within the first year of infection). Over one-third of these resided in a single large cluster that contained the pair of interest (Fig. 1): their bulk V3 sequences were predicted as X4 with g2p FPR values of 1.7% and 2.8%, respectively. CCR5 genotyping further revealed that one individual of the pair was CCR5wt/wt whereas the other was homozygous CCR5 Δ 32/ Δ 32. Neither individual expressed classical “protective” HLA class I alleles⁶⁴: their types were A*03:01/A*31:01, B*07:02/B*51:01, C*07:02/C*14:02 (CCR5wt/wt) and A*23:01/A*25:01, B*35:08/B*44:02, C*04:01/C*05:01 (CCR5 Δ 32/ Δ 32).

HIV-1 infection dates, estimated as the midpoint between the last HIV-negative and first HIV-positive tests, were March 2000 for the CCR5wt/wt individual (timepoint –17M, see methods) and August 2001 for the CCR5 Δ 32/ Δ 32 individual (timepoint 0M) (Fig. 2a). The timing of their respective infections and the observation that their HIV-1 sequences were nearly identical at the bulk level suggest that they could represent a transmission pair, with the CCR5wt/wt the putative donor and the CCR5 Δ 32/ Δ 32 individual the putative recipient. However, confirmation of transmission (e.g. via participant contact) was not possible due to the retrospective nature of the analysis and ethics guidelines, and involvement of a

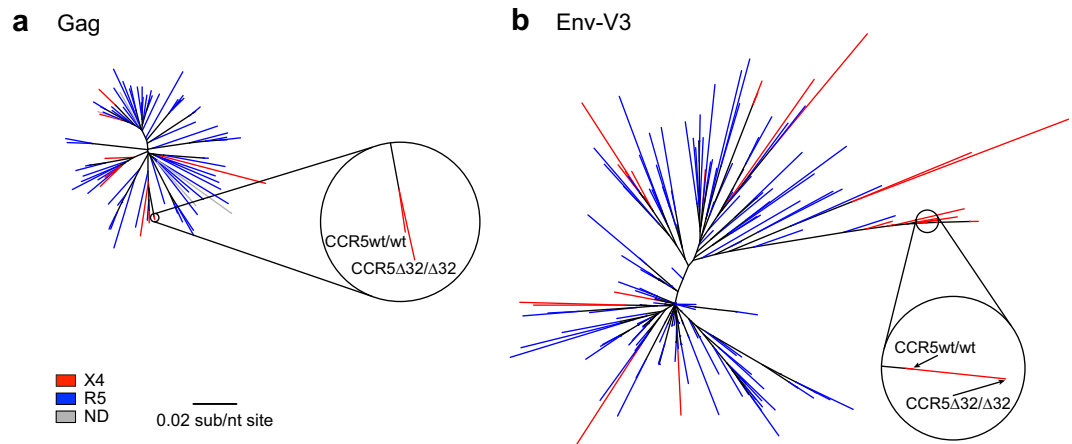


Figure 1. Maximum likelihood phylogenies of bulk HIV-1 Gag and V3 sequences from VIDUS participants. Maximum likelihood phylogenetic trees were constructed using available bulk Gag (*panel a*) and V3 (*panel b*) sequences from acute and chronically infected participants of the Vancouver Injection Drug Users Study. The CCR5wt/wt and CCR5Δ32/Δ32 individuals' sequences are shown in the zoomed-in window. Tree tips are coloured according to coreceptor usage predicted using V3 genotypes: red for X4-using, blue for R5-using sequences and gray for Gag sequences for which no corresponding V3 sequence was available for coreceptor prediction (ND; not determined).

third individual or intermediary host cannot be ruled out. Regardless, this represents a rare opportunity to study the evolutionary dynamics of a near-identical HIV-1 strain in individuals with distinct CCR5 genetics.

Marked differences in nadir CD4 T-cell count. We first analyzed available pre-therapy clinical measurements (Fig. 2b,c). The CCR5wt/wt individual's highest plasma viral load (pVL), $5.1 \log_{10}$ HIV-1 RNA copies/ml, and nadir CD4 T-cell count, 20 cells/mm³, were observed 17 months postinfection. This individual initiated HAART <1 month thereafter. The CCR5Δ32/Δ32 individual's highest pVL, observed 4.5 months postinfection, was $4.7 \log_{10}$ HIV-1 RNA copies/ml whereas their nadir CD4 count, observed 9 months postinfection, was 270 CD4 cells/mm³. CD4 counts in this individual subsequently rebounded to >400 cells/mm³ and this individual remained HAART-naïve throughout follow-up.

Deep sequencing and ancestral reconstruction. Deep sequencing of the HIV-1 V3 region was performed on all plasma and PBMC samples from both individuals using the Roche GS-Junior Platform. A median of 3143 (range 1905–7248) high quality sequences per sample were analyzed. In addition, 10 phylogenetic ancestral reconstructions were performed using 100 randomly sampled sequences from the three CCR5wt/wt and CCR5Δ32/Δ32 individuals' plasma specimens collected closest to "time zero" (the latter's infection date, see methods). Although the genetic similarity of these infections suggests they may represent a transmission pair, our reconstruction of the evolving lineages on a single timeline is not contingent on this being the case. This reconstruction provides a genetic and temporal context for the evolutionary fate of an unusual virus lineage transmitted from one host to another, possibly through an unknowable number of intermediate hosts.

All 10 ancestral reconstructions suggested that HIV-1 infection in both individuals was established by a single T/F viral strain (Fig. 3, Supplementary Figure S2 and S3). Moreover, the phylogenetically-estimated infection time ranges, averaged over all reconstructions (March to April 2000 for the CCR5wt/wt individual, and January to September 2001 for the CCR5Δ32/Δ32 individual) corroborated the clinically estimated infection dates for these persons. The T/F virus sequences for both individuals (estimated as the consensus of all 10 ancestral reconstructions performed) were predicted as X4 (median g2p FPR 1.7% [range 1.7–3.2%] for the CCR5wt/wt individual and 2.6% [range 1.7–3.8%] for the CCR5Δ32/Δ32 individual) (Figs 3,4, Supplementary Figure S2, and S3). Their reconstructed T/F V3 sequences differed by only one amino acid at V3 codon 24: the CCR5wt/wt T/F virus harbored arginine (R) whereas the CCR5Δ32/Δ32 T/F virus harbored lysine (K) (Fig. 4b).

Reconstruction of T/F virus sequences allowed us to track their frequencies in both hosts over time. After infection, the CCR5wt/wt individual's T/F virus dominated in plasma (86.9%) at their earliest studied timepoint (–13 M), continued to co-dominate for at least a year thereafter (42.4% at –7 M and 34.1% at –1 M), but was undetectable in plasma at +35 M. Concomitantly, the frequency of this sequence steadily decreased in this individual's PBMCs, from 51.7% at –13 M, to 17.2% at –7 M and then to low/undetectable levels thereafter (0% at –1 M and 1.0% at +35 M). Similarly, 6 months following infection

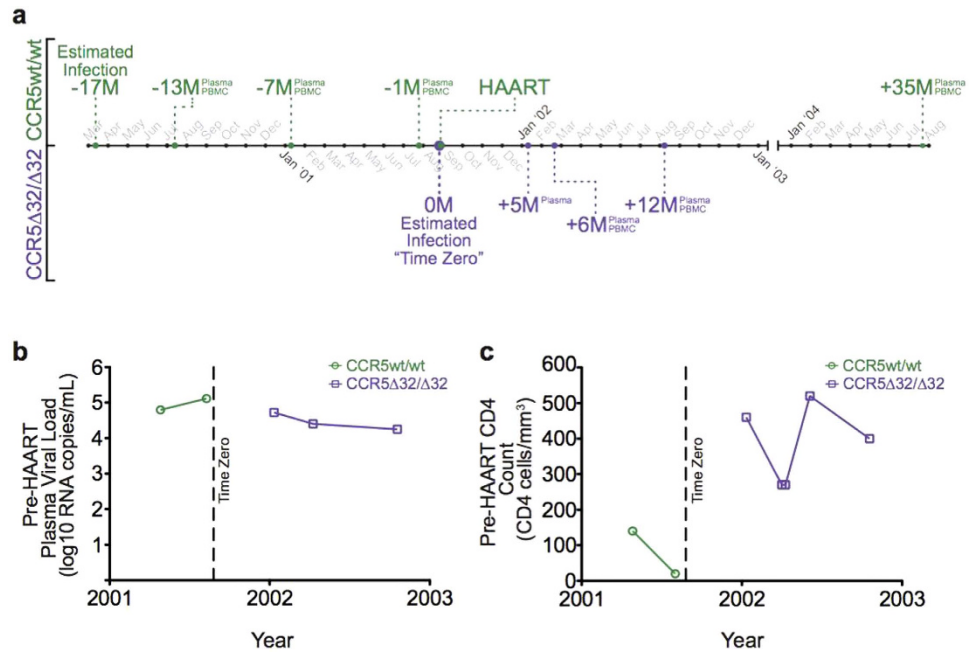


Figure 2. Sampling timeline and clinical histories for the CCR5wt/wt and CCR5 Δ 32/ Δ 32 hosts. *Panel (a):* Timelines for the CCR5wt/wt (above, green) and CCR5 Δ 32/ Δ 32 (below, purple) individuals are shown. Putative infection dates (March 2000 for CCR5wt/wt and August 2001 for CCR5 Δ 32/ Δ 32 individuals) were estimated as the midpoint between their last HIV-negative and first HIV-positive tests. The CCR5 Δ 32/ Δ 32 individual's estimated date of infection was arbitrarily set as "time zero" (0M); all other timepoints were expressed relative to this date. For the CCR5wt/wt individual plasma and PBMC were available -13 , -7 , -1 , and $+35$ months relative to time zero; for the CCR5 Δ 32/ Δ 32 individual, plasma samples were available $+5$, $+6$, and $+12$ months and PBMC $+6$ and $+12$ months relative to time zero. *Panel (b):* Available pre-HAART plasma viral loads (pVL) for CCR5wt/wt (green) and CCR5 Δ 32/ Δ 32 (purple) individuals. Maximum pVL for the CCR5wt/wt and CCR5 Δ 32/ Δ 32 individuals were 5.1 and 4.7 \log_{10} RNA copies/mL, respectively. *Panel (c):* Available pre-HAART CD4 counts for CCR5wt/wt (green) and CCR5 Δ 32/ Δ 32 (purple) individuals. Nadir CD4 counts for CCR5wt/wt and CCR5 Δ 32/ Δ 32 individuals were 20 and 270 CD4 cells/ mm^3 respectively.

the CCR5 Δ 32/ Δ 32 individual's T/F virus sequence remained co-dominant in plasma (36.2% at $+5$ M; 50.0% at $+6$ M) but was no longer detected in plasma at $+12$ M (Fig. 5c,d). By contrast, it remained co-dominant in PBMCs over the entire course of follow-up (35.6% at $+6$ M and 32.5% at $+12$ M).

Acknowledging the possibility that the two individuals could represent a transmission pair, we also investigated the frequency of the CCR5 Δ 32/ Δ 32 individual (putative recipient)'s T/F virus sequence in the CCR5wt/wt individual (putative donor)'s plasma and PBMC (Fig. 5a,b). The CCR5 Δ 32/ Δ 32 individual's T/F sequence was first detectable in the CCR5wt/wt individual at very low frequencies in plasma and PBMCs ($<1\%$ and 1.1% respectively) seven months prior to the CCR5 Δ 32/ Δ 32 individual's estimated infection date. One month prior to the CCR5 Δ 32/ Δ 32 individual's estimated infection date, the CCR5 Δ 32/ Δ 32 individual's T/F sequence remained low frequency ($<1\%$) in the CCR5wt/wt individual's plasma but co-dominated (33.5%) in PBMC. By timepoint $+35$ M, this sequence was no longer detected in the CCR5wt/wt individual's plasma and was observed at $<1\%$ in PBMC. While detection of this sequence within the CCR5wt/wt individual's HIV-1 variant pool is intriguing, this observation does not in itself constitute proof of transmission (for example, transmission via one or more intermediary host(s) could have introduced additional unseen genetic bottleneck(s) between our CCR5wt/wt and CCR5 Δ 32/ Δ 32 individuals).

Divergence from the reconstructed T/F viruses. We next wished to compare the extent to which plasma HIV-1 RNA V3 sequences in both individuals initially diverged from their respective T/F viruses (Fig. 6). To eliminate HAART as a confounder, analysis was restricted to the CCR5wt/wt individual's first 10 months of infection (pre-HAART period) and a comparable follow-up time for the CCR5 Δ 32/ Δ 32 individual. During this time, the mean divergence from the CCR5wt/wt individual's T/F virus was 0.0027 sub/nt site, an average rate of divergence of 0.00065 sub/nt site per month. In contrast, mean initial divergence of plasma V3 sequences from the T/F virus in the CCR5 Δ 32/ Δ 32 individual was 0.031 sub/nt site, an average rate of 0.0036 sub/nt site per month, a value that was 5.5-fold higher than that observed in the CCR5wt/wt individual.

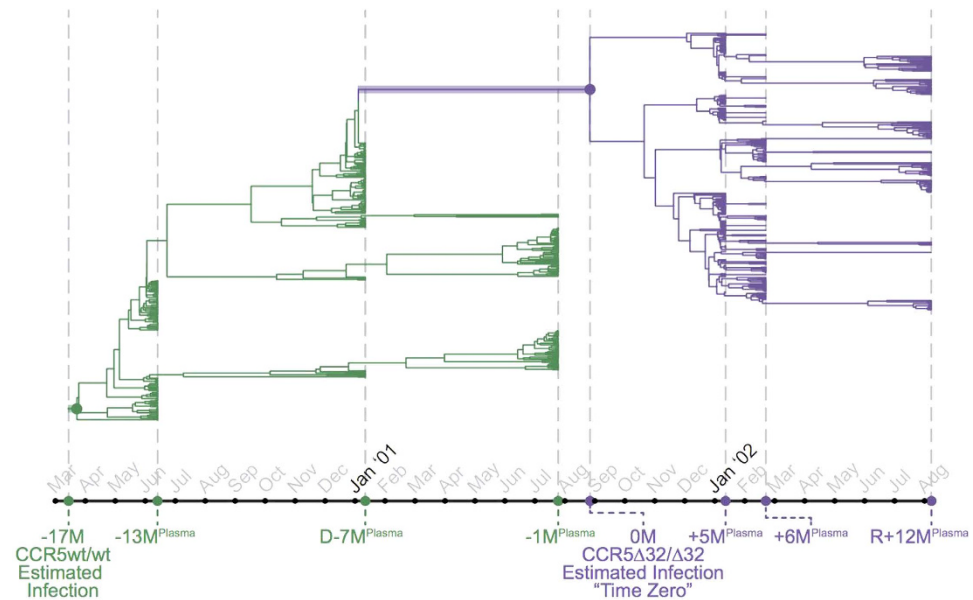


Figure 3. Ancestral phylogenetic reconstruction of HIV-1 V3 transmission/evolution. $N = 10$ ancestral phylogenetic reconstructions were performed by sampling 100 *plasma* HIV RNA-derived ultradeep sequences per timepoint for the three CCR5wt/wt (green) and three CCR5 Δ 32/ Δ 32 (purple) timepoints closest to time zero. A representative reconstructed phylogeny is shown. Shaded branches and their associated internal nodes represent phylogenetically-inferred date ranges for the CCR5wt/wt (green) and CCR5 Δ 32/ Δ 32 (purple) individuals. This ancestral phylogenetic reconstruction indicates that both CCR5wt/wt and CCR5 Δ 32/ Δ 32 individuals were productively infected by a single X4 virus, within a time period that coincides with the clinical estimated dates of infection. The remaining nine phylogenetic reconstructions were also consistent with transmission of a single T/F virus; in addition, 9 of 10 reconstructions yielded estimated infection date ranges that coincided with clinical estimates. Two additional reconstructions are shown in supplementary figures S2 and S3.

Differential HIV-1 coreceptor usage evolution. We next investigated HIV-1 coreceptor usage evolution in both hosts. In the CCR5wt/wt individual, over a total of 52 months followup, R5 V3 sequences gradually emerged alongside their X4 counterparts (Figs. 5a,b,7a). The first R5 variants were detected 4 months following infection in PBMC (timepoint $-13M$): these early R5 variants comprised $<1\%$ of all sequences in this sample and exhibited g2p FPRs in the marginal range (5.1–10.8%). No R5 variants were detected in plasma at this timepoint. By 10 months following infection (timepoint $-7M$), R5 variants were detected at $<1\%$ frequency in plasma and 1.7% in PBMC, again with marginal FPRs (5.3–10.8%). By 16 months after infection (timepoint $-1M$), R5 variants reached frequencies of 41.3% in plasma and 18.1% in PBMC, though FPRs remained marginal (median 8.7% in both compartments). By timepoint $+35M$, R5 sequences dominated in both plasma (62.4%) and PBMCs (74.6%), with median FPRs of 18.9% in both compartments (Fig. 7a).

In contrast, in the CCR5 Δ 32/ Δ 32 individual, essentially all (14721 of 14809; 99.4%) plasma and PBMC HIV-1 sequences remained X4 throughout followup (median FPR 2.6% in both compartments) (Figs. 5c,d,7b). The remaining minority (88 of 14809; 0.6%) of sequences were technically R5, but these exhibited marginal g2p FPRs (range 5.3–8.7%). Moreover, unlike in the CCR5wt/wt individual, the frequencies of sequences with FPRs in this range did not increase over time in the CCR5 Δ 32/ Δ 32 individual.

We also investigated V3 codon substitutions over time (Fig. 8). For the CCR5wt/wt individual, comparison of the earliest ($-13M$) and latest ($+35M$) plasma V3 sequences identified five codons (5, 24, 25, 27, and 34) that diversified significantly and three (8, 18, and 26) that contracted modestly during this time (Fig. 8a) ($p < 0.001$). Codon 25 diversified to the greatest extent, with the X4-associated arginine (R) decreasing from 99.8% to 38.7%. In the CCR5 Δ 32/ Δ 32 individual, comparison of earliest ($+5M$) and latest ($+12M$) plasma V3 sequences identified nine diversifying (4, 9, 24–27, 29, 30, 32) and three contracting codons (10, 13 and 18) ($p < 0.001$) (Fig. 8b). Codon 25 also ranked among the most highly diversifying in this individual, with the dominant X4-associated arginine (R) giving way to a 63.5/35.7% mixture of lysine (K)/arginine (R).

Discussion

We performed longitudinal HIV-1 env-V3 deep sequencing and ancestral phylogenetic reconstruction to study infection (presumably via injection drug use) and subsequent diversification of a highly genetically similar X4 HIV-1 strain in CCR5wt/wt and CCR5 Δ 32/ Δ 32 hosts. A total of 10 independent ancestral

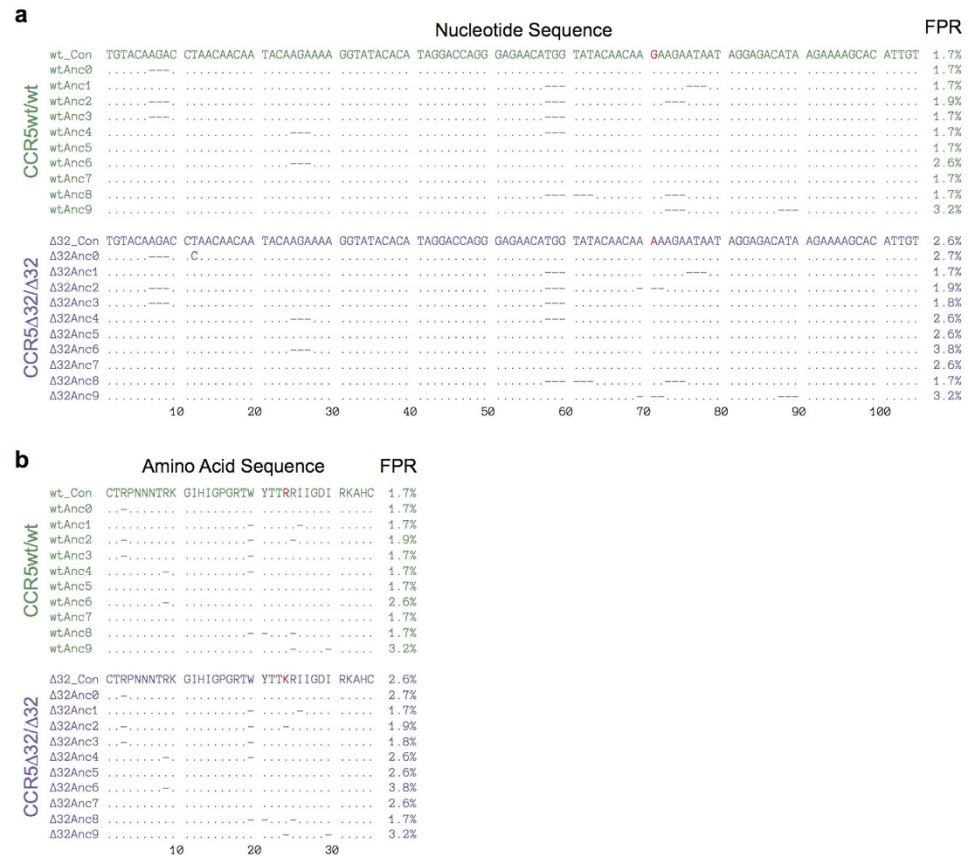


Figure 4. Nucleotide and protein alignments of reconstructed transmitted/founder viruses. Nucleotide and amino acid sequence alignments of the reconstructed T/F virus sequence that infected the CCR5wt/wt (green) and CCR5Δ32/Δ32 (purple) individuals. The consensus sequence of all 10 ancestral reconstructions (labeled “wt_Con” and “Δ32_Con” for the CCR5wt/wt and CCR5Δ32/Δ32 individuals respectively) is used as the reference. Periods (“.”) indicate positions where the sequence is the same as the reference and dashes (“-”) indicate deletions. The “FPR” value following each reconstructed T/F sequence denotes its false-positive rate assigned by $\text{geno2pheno}_{\text{coreceptor}}^8$; sequences with $\text{FPR} \leq 5.0\%$ are considered X4. **Panel (a):** CCR5wt/wt (top, green) and CCR5Δ32/Δ32 (bottom, purple) nucleotide acid alignments. Consensus nucleotide differences between the CCR5wt/wt and CCR5Δ32/Δ32 individuals are shown in red. **Panel (b):** CCR5wt/wt (top, green) and CCR5Δ32/Δ32 (bottom, purple) amino acid alignments. The single amino acid difference between the two T/F viruses (at codon 24) is shown in red.

reconstructions performed using plasma-derived deep sequences uniformly suggested that both individuals were productively infected with a single, nearly-identical X4 T/F virus, a multiplicity of infection (MOI) which is lower than the reported average for infection via injection drug use ($N = 3$)⁴⁵. The lower-than-average MOI observed in both individuals is consistent with reports of poorer “transmission fitness” of X4 viruses^{11,14,65}, though larger studies would be required to confirm a relationship between HIV-1 coreceptor usage and MOI in different risk groups. It also argues against the CCR5Δ32/Δ32 genotype acting as a major additional transmission bottleneck in this case. Consistent with initial rapid diversification of the T/F virus followed by extinction of many viral lineages by immune selection⁶⁶, within-host phylogenies for both individuals were initially starlike but later exhibited a more asymmetrical appearance. Indeed, evidence of selection by neutralizing antibodies was observed in both hosts in the form of diversity loss at V3 codon 18, the final residue of the GPGR “crown motif” within a neutralizing antibody epitope⁶⁷. That this occurred in both hosts is notable as it supports reproducible pathways and timecourse of antibody-driven escape in individuals acquiring genetically-similar viral strains⁶⁸.

Our observations may also shed some insight on the longstanding debate regarding whether X4 viruses cause immune decline (i.e. whether they are inherently more pathogenic than their R5 counterparts), or whether they arise as a consequence of it (i.e. whether R5 and X4 viruses are inherently “fitter” under conditions of relatively preserved immune function versus immunodeficiency, respectively). Our observations indicate that both may be true to some extent. The markedly different nadir CD4 counts observed in the studied individuals supports X4 viruses as inherently more pathogenic when acquired at transmission - but only in CCR5wt/wt hosts. Indeed, this is consistent with previous reports of rapid

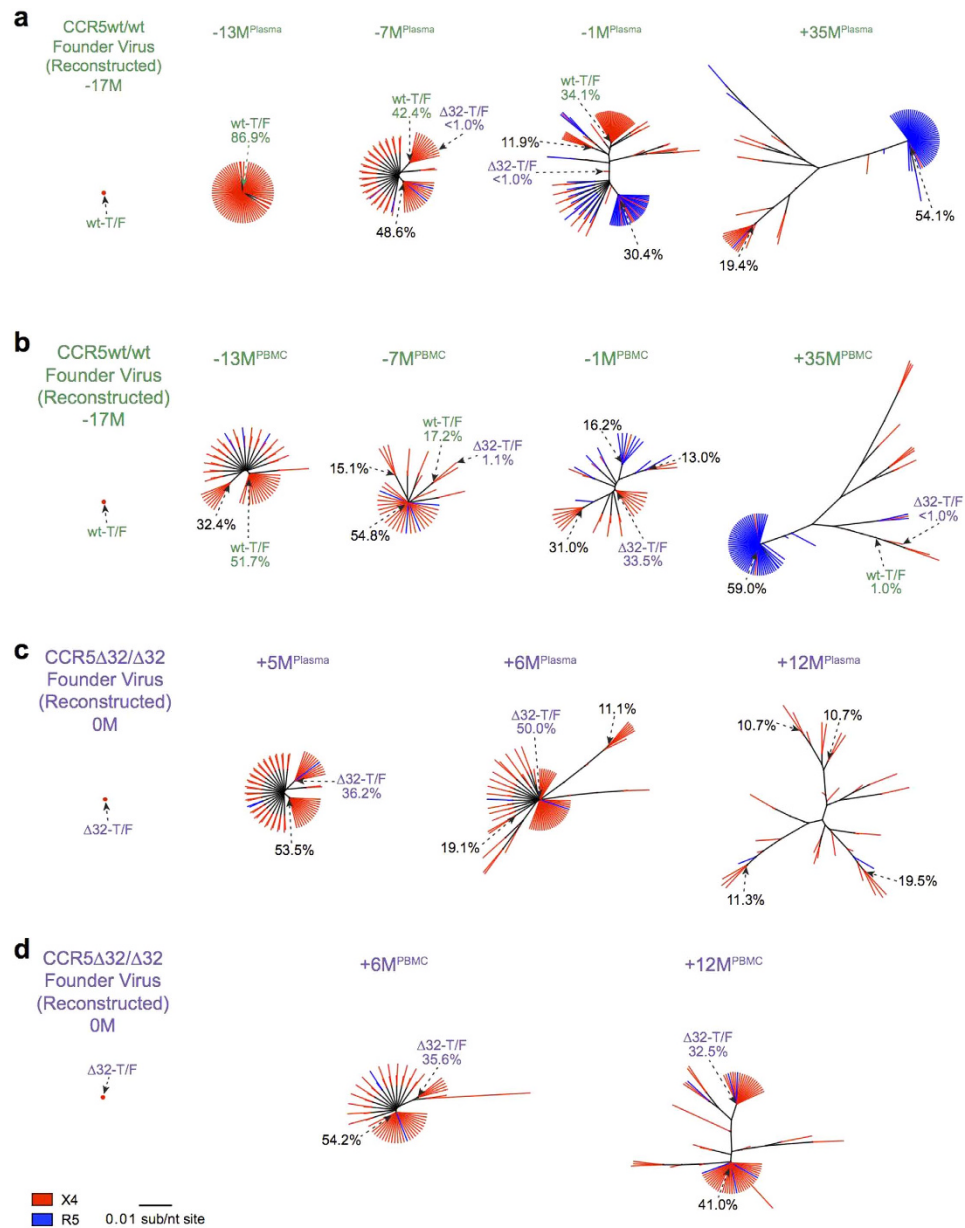


Figure 5. HIV-1 V3 diversification over time. Maximum likelihood phylogenetic trees constructed from unique plasma and PBMC deep V3 sequences from the CCR5wt/wt (panel (a) and (b)) and CCR5Δ32/Δ32 (panel (c) and (d)) individuals. Branches are colored by predicted coreceptor usage: red for X4; blue for R5. Prevalent sequences are labeled with their observed frequencies. “wt-T/F” (green) denotes the transmitted/founder virus acquired by the CCR5wt/wt individual; its presence and frequency is tracked throughout the CCR5wt/wt trees. “Δ32-T/F” (purple) denotes the transmitted/founder virus acquired by the CCR5Δ32/Δ32 individual; its presence and frequency is tracked throughout both individuals’ trees. All phylogenies are drawn on the same genetic distance scale.

untreated HIV-1 progression following X4 virus infection in CCR5wt/wt^{24,25}-but not CCR5Δ32/Δ32⁴¹-individuals. On the other hand, the gradual emergence of R5 variants alongside their X4 counterparts in the CCR5wt/wt host over time is consistent with the hypothesis that, under conditions of relatively preserved immune function, R5 viruses have some advantages over their X4 counterparts¹²—possibly in terms of decreased sensitivity to neutralizing antibodies⁶⁹ and/or ability to infect CCR5-expressing CD4 cell subsets^{17,70,71}. Indeed, such X4-to-R5 “reversions” have previously been reported in similar contexts^{50,72}. Of note, the lengthy timeline and incomplete nature of this process^{50,72} suggest that *in vivo* advantages of R5 over X4 viruses are modest in magnitude when the former arise in this context.

Our results also highlight host CCR5 genotype as a major determinant of the evolutionary dynamics of X4 infections. In contrast to the CCR5wt/wt individual who gradually and incompletely reverted to

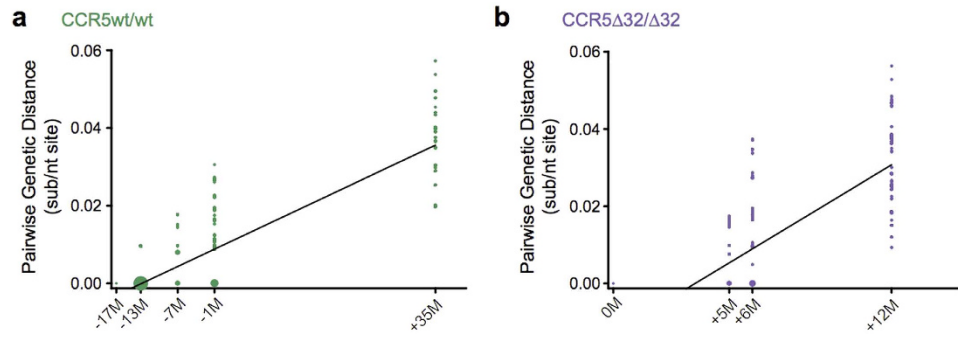


Figure 6. Divergence from the transmitted/founder HIV-1 V3 sequence. Panel (a) Pairwise genetic distances between the CCR5wt/wt individual's estimated T/F V3 sequence and all subsequently-observed plasma HIV RNA sequences, measured in substitutions per nucleotide site (sub/nt site). Datapoint sizes reflect observed sequence frequencies, with the largest point representing ~5000 sequences. Panel (b): Corresponding genetic distances between the CCR5 Δ 32/ Δ 32 individual's T/F V3 and subsequent plasma V3 sequences.

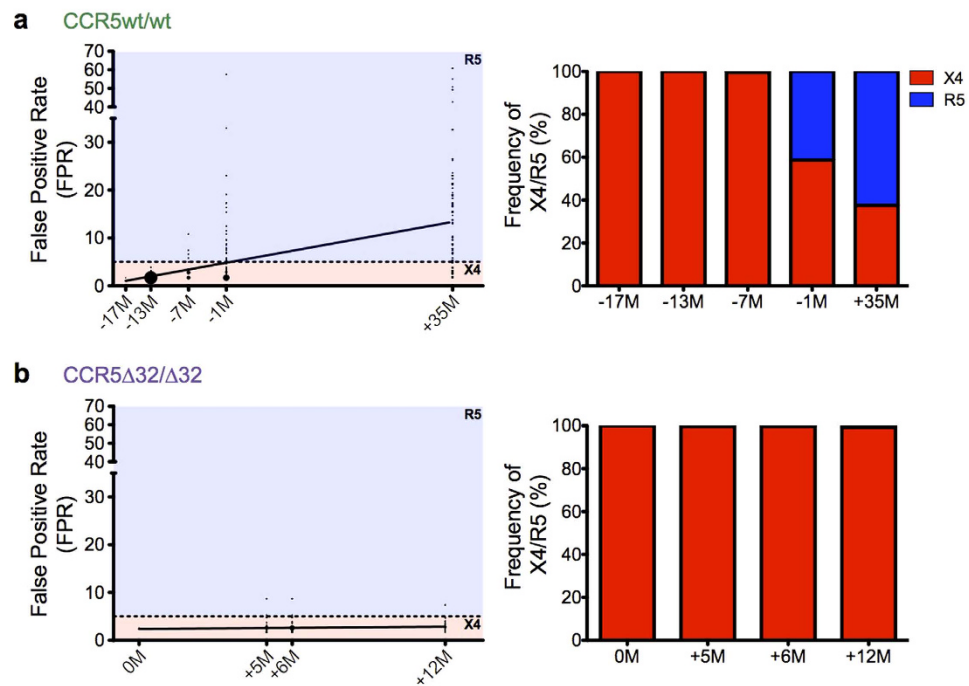


Figure 7. Marked differences in coreceptor usage evolution in CCR5wt/wt vs. CCR5 Δ 32/ Δ 32 hosts. Panel (a) left: The false-positive rate (FPR) of HIV-1 coreceptor usage prediction for the CCR5wt/wt individual's T/F virus (−17M) and each unique plasma HIV RNA sequence collected thereafter. The horizontal dotted line denotes FPR = 5.0%; sequences with values at or below this threshold are considered X4. Panel (a) right: summarizes the CCR5wt/wt individual's data in terms of the % of total sequences displaying X4 (red) vs. R5 (blue) usage at each timepoint. Panel (b): Corresponding analyses for the CCR5 Δ 32/ Δ 32 individual.

R5, in the CCR5 Δ 32/ Δ 32 individual V3 evolution “explored” genetic space while maintaining a strong X4 phenotype. The CCR5 Δ 32/ Δ 32 genotype did not appear to constrain V3 evolution: initial within-host rates of sequence divergence from the T/F strain in this host were greater than in the CCR5wt/wt individual.

It is not possible to prove the source of HIV-1 transmission using clinical and phylogenetic data alone. Nevertheless, the observation that the two individuals' patristic (tip-to-tip) HIV-1 genetic distances were the shortest among all VIDUS participants examined (the individual harboring the next closest HIV-1 sequence to the CCR5 Δ 32/ Δ 32 individual exhibited mean genetic distances >5.6-fold [range 1.4–107] greater than those separating the studied pair) and the observation that the CCR5 Δ 32/ Δ 32's HIV-1

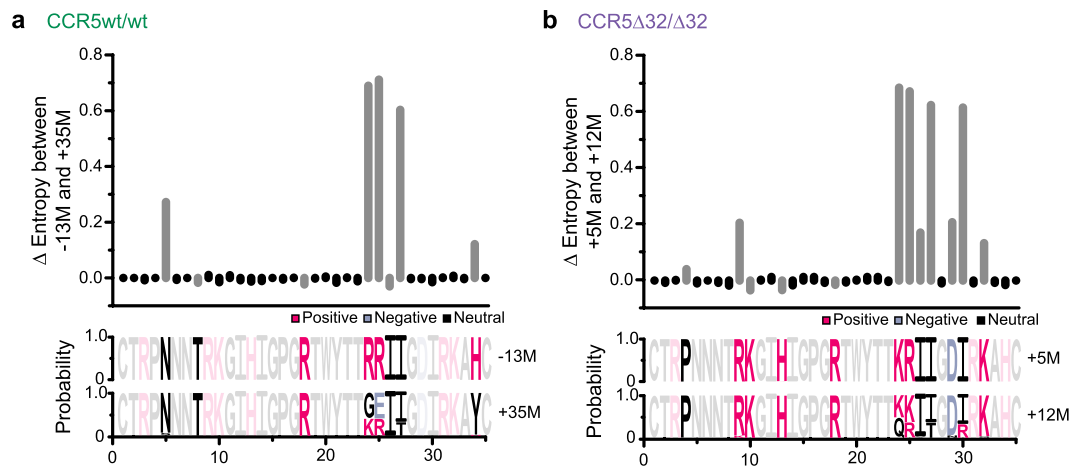


Figure 8. Amino acid diversification at key coreceptor tropism determining sites. *Panel (a) Top:* Differences in Shannon Entropy (Δ Entropy) between the latest (+35M) and earliest (−13M) V3 plasma HIV-1 amino acid alignments from the CCR5wt/wt individual. Positive values denote residues that exhibit higher entropy in the later vs. the earlier timepoint (negative values denote the opposite). Significant ($p < 0.001$) values are shown in Gray. *Panel (a) bottom:* Corresponding plasma V3 amino acid frequencies at these two timepoints. Positive, negative and neutrally-charged residues are in pink, grey and black respectively, with significantly-changing sites in bright colors and non-significantly-changing sites in dull colors. *Panel (b):* Corresponding analysis for the latest (+12M) vs. earliest (+5M) plasma V3 sequences from the CCR5 Δ 32/ Δ 32 individual.

sequences formed a monophyletic lineage nested within a more diverse group of sequences from the CCR5wt/wt individual support the possibility that CCR5wt/wt and CCR5 Δ 32/ Δ 32 individuals represent donor and recipient respectively. Hypothesizing that this is the case, it is interesting to note that at the timepoint closest to the putative transmission event (−1M), the CCR5 Δ 32/ Δ 32's inferred T/F strain represented a minority (<1%) variant in the CCR5wt/wt individual's plasma but co-dominated (33.5%) in PBMC. This in turn raises the intriguing possibility that, if this were a transmission pair, infection of the CCR5 Δ 32/ Δ 32 individual could have occurred via transfer of an infected cell (or via transfer of a minority plasma variant) from the CCR5wt/wt individual. If so, this would corroborate the published finding that T/F viruses cannot be directly predicted from analysis of donor HIV-1 sequences only⁷³. Moreover, if the individuals represented a transmission pair the observation that their T/F viruses differed by only one V3 amino acid corroborates the published finding that transmitted viruses tend to be genetically closer to “ancestral” donor viruses than those present in plasma at the time of transmission^{49,74}. Indeed, the patristic distance separating the individuals' T/F viruses (0.01704 sub/nt site) was marginally yet significantly lower than that separating the CCR5 Δ 32/ Δ 32 individuals' T/F virus and plasma viruses present within the CCR5wt/wt individual at the former's time of infection (0.01711 sub/nt site) ($p < 0.0001$, Wilcoxon one-sample test).

Several limitations of this study merit mention. First, we infer that HIV-1 infection in both individuals occurred via injection drug use, but sexual transmission cannot be ruled out. Second, as viral templates were not directly quantified (e.g. using “primer ID” techniques⁷⁵), possible PCR amplification and/or template resampling biases must be acknowledged, though our approach of triplicate amplifying each extract, quantifying DNA and pooling resulting amplicons equally prior to deep sequencing reduces this concern to some extent⁴⁸. Nevertheless, such biases could conceivably influence our phylogenetic reconstruction of T/F virus sequences, though several lines of evidence indicate this is not the case. Firstly, the composition (in terms of unique sequences and their frequencies) was consistent across plasma and PBMC compartments over time in both hosts. Second, within-host plasma and PBMC phylogenies exhibited characteristic shapes over time. Thirdly, ancestral reconstructions were performed using *plasma* sequences only – but despite this, we detected the reconstructed T/F sequences in PBMCs as well. Finally, replicate sequence data available from 8 of 13 analyzed samples were concordant with the original results in terms of both sequence identity and variant frequencies. Lastly, as this is a descriptive study of a rare event our ability to draw broad conclusions is limited. Nevertheless, our observations highlight the utility of deep sequencing paired with phylogenetic ancestral reconstruction to illuminate early dynamics of X4 HIV-1 infections and underscore the influence of CCR5 genotype on HIV-1 V3 evolution.

References

1. Klatzmann, D. *et al.* T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* **312**, 767–768 (1984).

2. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872–877 (1996).
3. Samson, M., Labbe, O., Mollereau, C., Vassart, G. & Parmentier, M. Molecular cloning and functional expression of a new human CC-chemokine receptor gene. *Biochemistry* **35**, 3362–3367, doi: 10.1021/bi952950g (1996).
4. Berger, E. A. *et al.* A new classification for HIV-1. *Nature* **391**, 240, doi: 10.1038/34571 (1998).
5. De Jong, J. J., De Ronde, A., Keulen, W., Tersmette, M. & Goudsmit, J. Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. *J Virol* **66**, 6777–6780 (1992).
6. Fouchier, R. A. *et al.* Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol* **66**, 3183–3187 (1992).
7. Coakley, E. *et al.* Comparison of human immunodeficiency virus type 1 tropism profiles in clinical samples by the Trofile and MT-2 assays. *Antimicrob Agents Chemother* **53**, 4686–4693, doi: 10.1128/AAC.00229-09 (2009).
8. Lengauer, T., Sander, O., Sierra, S., Thielen, A. & Kaiser, R. Bioinformatics prediction of HIV coreceptor usage. *Nat Biotechnol* **25**, 1407–1410, doi: 10.1038/nbt1371 (2007).
9. Jensen, M. A. *et al.* Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *J Virol* **77**, 13376–13388 (2003).
10. Berger, E. A., Murphy, P. M. & Farber, J. M. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol* **17**, 657–700, doi: 10.1146/annurev.immunol.17.1.657 (1999).
11. Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci USA* **105**, 7552–7557, doi: 10.1073/pnas.0802203105 (2008).
12. Grivel, J. C., Shattock, R. J. & Margolis, L. B. Selective transmission of R5 HIV-1 variants: where is the gatekeeper? *J Transl Med* **9 Suppl 1**, S6, doi: 10.1186/1479-5876-9-S1-S6 (2011).
13. Koot, M. *et al.* Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS. *Ann Intern Med* **118**, 681–688 (1993).
14. Schuitemaker, H. *et al.* Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus population. *J Virol* **66**, 1354–1360 (1992).
15. Brumme, Z. L. *et al.* Molecular and clinical epidemiology of CXCR4-using HIV-1 in a large population of antiretroviral-naive individuals. *J Infect Dis* **192**, 466–474, doi: 10.1086/431519 (2005).
16. Richman, D. D. & Bozzette, S. A. The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression. *J Infect Dis* **169**, 968–974 (1994).
17. Schuitemaker, H., van 't Wout, A. B. & Lusso, P. Clinical significance of HIV-1 coreceptor usage. *J Transl Med* **9 Suppl 1**, S5, doi: 10.1186/1479-5876-9-S1-S5 (2011).
18. Ceresola, E. R. *et al.* Performance of commonly used genotypic assays and comparison with phenotypic assays of HIV-1 coreceptor tropism in acutely HIV-1-infected patients. *J Antimicrob Chemother*, doi: 10.1093/jac/dku573 (2015).
19. Raymond, S. *et al.* Characterization of CXCR4-using HIV-1 during primary infection by ultra-deep pyrosequencing. *J Antimicrob Chemother* **68**, 2875–2881, doi: 10.1093/jac/dkt290 (2013).
20. Frange, P. *et al.* High frequency of X4/DM-tropic viruses in PBMC samples from patients with primary HIV-1 subtype-B infection in 1996–2007: the French ANRS CO06 PRIMO Cohort Study. *J Antimicrob Chemother* **64**, 135–141, doi: 10.1093/jac/dkp151 (2009).
21. Frange, P. *et al.* Prevalence of CXCR4-tropic viruses in clustered transmission chains at the time of primary HIV-1 infection. *Clin Microbiol Infect* **19**, E252–255, doi: 10.1111/1469-0691.12127 (2013).
22. Sierra-Enguita, R. *et al.* X4 tropic viruses are on the rise in recent HIV-1 seroconverters in Spain. *AIDS* **28**, 1603–1609, doi: 10.1097/QAD.0000000000000269 (2014).
23. Wagner, G. A. *et al.* Using ultradeep pyrosequencing to study HIV-1 coreceptor usage in primary and dual infection. *J Infect Dis* **208**, 271–274, doi: 10.1093/infdis/jit168 (2013).
24. Sheppard, H. W. *et al.* HIV-1 infection in individuals with the CCR5-Delta32/Delta32 genotype: acquisition of syncytium-inducing virus at seroconversion. *J Acquir Immune Defic Syndr* **29**, 307–313 (2002).
25. Raymond, S. *et al.* CXCR4-using viruses in plasma and peripheral blood mononuclear cells during primary HIV-1 infection and impact on disease progression. *AIDS* **24**, 2305–2312, doi: 10.1097/QAD.0b013e32833e50bb (2010).
26. Supupira, M. C. *et al.* Faster HIV-1 disease progression among Brazilian individuals recently infected with CXCR4-utilizing strains. *PLoS One* **7**, e30292, doi: 10.1371/journal.pone.0030292 (2012).
27. Liu, R. *et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367–377 (1996).
28. Martinson, J. J., Chapman, N. H., Rees, D. C., Liu, Y. T. & Clegg, J. B. Global distribution of the CCR5 gene 32-basepair deletion. *Nat Genet* **16**, 100–103, doi: 10.1038/ng0597-100 (1997).
29. Huang, Y. *et al.* The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat Med* **2**, 1240–1243 (1996).
30. Marmor, M. *et al.* Homozygous and heterozygous CCR5-Delta32 genotypes are associated with resistance to HIV infection. *J Acquir Immune Defic Syndr* **27**, 472–481 (2001).
31. Biti, R. *et al.* HIV-1 infection in an individual homozygous for the CCR5 deletion allele. *Nat Med* **3**, 252–253 (1997).
32. O'Brien, T. R. *et al.* HIV-1 infection in a man homozygous for CCR5 delta 32. *Lancet* **349**, 1219 (1997).
33. Theodorou, I., Meyer, L., Magierowska, M., Katlama, C. & Rouzioux, C. HIV-1 infection in an individual homozygous for CCR5 delta 32. Seroco Study Group. *Lancet* **349**, 1219–1220 (1997).
34. Balotta, C. *et al.* Homozygous delta 32 deletion of the CCR-5 chemokine receptor gene in an HIV-1-infected patient. *AIDS* **11**, F67–71 (1997).
35. Heiken, H., Becker, S., Bastisch, I. & Schmidt, R. E. HIV-1 infection in a heterosexual man homozygous for CCR-5 delta32. *AIDS* **13**, 529–530 (1999).
36. Kuipers, H. *et al.* An HIV-1-infected individual homozygous for the CCR-5 delta32 allele and the SDF-1 3'A allele. *AIDS* **13**, 433–434 (1999).
37. Gorry, P. R. *et al.* Persistence of dual-tropic HIV-1 in an individual homozygous for the CCR5 Delta 32 allele. *Lancet* **359**, 1832–1834, doi: 10.1016/S0140-6736(02)08681-6 (2002).
38. Iversen, A. K. *et al.* Limited protective effect of the CCR5Delta32/CCR5Delta32 genotype on human immunodeficiency virus infection incidence in a cohort of patients with hemophilia and selection for genotypic X4 virus. *J Infect Dis* **187**, 215–225, doi: 10.1086/345881 (2003).
39. Gray, L. *et al.* Genetic and functional analysis of R5X4 human immunodeficiency virus type 1 envelope glycoproteins derived from two individuals homozygous for the CCR5delta32 allele. *J Virol* **80**, 3684–3691, doi: 10.1128/JVI.80.7.3684-3691.2006 (2006).
40. Oh, D. Y. *et al.* CCR5Delta32 genotypes in a German HIV-1 seroconverter cohort and report of HIV-1 infection in a CCR5Delta32 homozygous individual. *PLoS One* **3**, e2747, doi: 10.1371/journal.pone.0002747 (2008).

41. Henrich, T. J. *et al.* Viremic control and viral coreceptor usage in two HIV-1-infected persons homozygous for CCR5 Delta32. *AIDS*, doi: 10.1097/QAD.0000000000000629 (2015).
42. Bhaskaran, K., Brettle, R., Porter, K., Walker, A. S. & Collaboration, C. Systemic non-Hodgkin lymphoma in individuals with known dates of HIV seroconversion: incidence and predictors. *AIDS* **18**, 673–681 (2004).
43. Hutter, G. *et al.* Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N Engl J Med* **360**, 692–698, doi: 10.1056/NEJMoa0802905 (2009).
44. Salazar-Gonzalez, J. F. *et al.* Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J Exp Med* **206**, 1273–1289, doi: 10.1084/jem.20090378 (2009).
45. Bar, K. J. *et al.* Wide variation in the multiplicity of HIV-1 infection among injection drug users. *J Virol* **84**, 6241–6247, doi: 10.1128/JVI.00077-10 (2010).
46. Derdeyn, C. A. *et al.* Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* **303**, 2019–2022, doi: 10.1126/science.1093137 (2004).
47. McCloskey, R. M., Liang, R. H., Harrigan, P. R., Brumme, Z. L. & Poon, A. F. An evaluation of phylogenetic methods for reconstructing transmitted HIV variants using longitudinal clonal HIV sequence data. *J Virol* **88**, 6181–6194, doi: 10.1128/JVI.00483-14 (2014).
48. Poon, A. F. *et al.* Reconstructing the dynamics of HIV evolution within hosts from serial deep sequence data. *PLoS Comput Biol* **8**, e1002753, doi: 10.1371/journal.pcbi.1002753 (2012).
49. Herbeck, J. T. *et al.* Human immunodeficiency virus type 1 env evolves toward ancestral states upon transmission to a new host. *J Virol* **80**, 1637–1644, doi: 10.1128/JVI.80.4.1637-1644.2006 (2006).
50. Sede, M. M., Moretti, F. A., Laufer, N. L., Jones, L. R. & Quarleri, J. F. HIV-1 tropism dynamics and phylogenetic analysis from longitudinal ultra-deep sequencing data of CCR5- and CXCR4-using variants. *PLoS One* **9**, e102857, doi: 10.1371/journal.pone.0102857 (2014).
51. Strathdee, S. A. *et al.* Needle exchange is not enough: lessons from the Vancouver injecting drug use study. *AIDS* **11**, F59–65 (1997).
52. Woods, C. K. *et al.* Automating HIV drug resistance genotyping with RECall, a freely accessible sequence analysis tool. *J Clin Microbiol* **50**, 1936–1942, doi: 10.1128/JCM.06689-11 (2012).
53. Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679, doi: 10.1093/bioinformatics/bti079 (2005).
54. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307–321, doi: 10.1093/sysbio/syq010 (2010).
55. Fourment, M. & Gibbs, M. J. PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evol Biol* **6**, 1, doi: 10.1186/1471-2148-6-1 (2006).
56. Brumme, Z. L. *et al.* CCR5Delta32 and promoter polymorphisms are not correlated with initial virological or immunological treatment response. *AIDS* **15**, 2259–2266 (2001).
57. Cotton, L. A. *et al.* HLA class I sequence-based typing using DNA recovered from frozen plasma. *J Immunol Methods* **382**, 40–47, doi: 10.1016/j.jim.2012.05.003 (2012).
58. Knapp, D. J. *et al.* “Deep” sequencing accuracy and reproducibility using Roche/454 technology for inferring co-receptor usage in HIV-1. *PLoS One* **9**, e99508, doi: 10.1371/journal.pone.0099508 (2014).
59. Swenson, L. C., Dong, W., Mo, T., Woods, C., Zhong, X., Thielen, A., Jensen, M., Biswas, P., Ellery, S., Lewis, M., James, I., Chapman, D., Valdez, H. & Harrigan, R. In *European AIDS Conference*. (European AIDS Clinical Society).
60. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214, doi: 10.1186/1471-2148-7-214 (2007).
61. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797, doi: 10.1093/nar/gkh340 (2004).
62. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**, 512–526 (1993).
63. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R News*, 7–11 (2006).
64. O’Brien, S. J., Gao, X. & Carrington, M. HLA and AIDS: a cautionary tale. *Trends Mol Med* **7**, 379–381, doi: S1471-4914(01)02131-1 [pii] (2001).
65. Wilen, C. B. *et al.* Phenotypic and immunologic comparison of clade B transmitted/founder and chronic HIV-1 envelope glycoproteins. *J Virol* **85**, 8514–8527, doi: 10.1128/JVI.00736-11 (2011).
66. Herbeck, J. T. *et al.* Demographic processes affect HIV-1 evolution in primary infection before the onset of selective processes. *J Virol*, doi: 10.1128/JVI.02697-10 (2011).
67. Zolla-Pazner, S. *et al.* The cross-clade neutralizing activity of a human monoclonal antibody is determined by the GPGR V3 motif of HIV type 1. *AIDS Res Hum Retroviruses* **20**, 1254–1258, doi: 10.1089/0889222042545054 (2004).
68. Draenert, R. *et al.* Constraints on HIV-1 evolution and immunodominance revealed in monozygotic adult twins infected with the same virus. *J Exp Med* **203**, 529–539, doi: 10.1084/jem.20052116 (2006).
69. Bunnik, E. M., Quakkelaar, E. D., van Nuenen, A. C., Boeser-Nunnink, B. & Schuitemaker, H. Increased neutralization sensitivity of recently emerged CXCR4-using human immunodeficiency virus type 1 strains compared to coexisting CCR5-using variants from the same patient. *J Virol* **81**, 525–531, doi: 10.1128/JVI.01983-06 (2007).
70. Bleul, C. C., Wu, L., Hoxie, J. A., Springer, T. A. & Mackay, C. R. The HIV coreceptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes. *Proc Natl Acad Sci USA* **94**, 1925–1930 (1997).
71. Berkowitz, R. D., Beckerman, K. P., Schall, T. J. & McCune, J. M. CXCR4 and CCR5 expression delineates targets for HIV-1 disruption of T cell differentiation. *J Immunol* **161**, 3702–3710 (1998).
72. Baroncelli, S. *et al.* HIV-1 coreceptor switch during 2 years of structured treatment interruptions. *Eur J Clin Microbiol Infect Dis* **32**, 1565–1570, doi: 10.1007/s10096-013-1911-z (2013).
73. Frange, P. *et al.* Sexually-transmitted/founder HIV-1 cannot be directly predicted from plasma or PBMC-derived viral quaspecies in the transmitting partner. *PLoS One* **8**, e69144, doi: 10.1371/journal.pone.0069144 (2013).
74. Sagar, M. *et al.* Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. *J Infect Dis* **199**, 580–589, doi: 10.1086/596557 (2009).
75. Jabara, C. B., Jones, C. D., Roach, J., Anderson, J. A. & Swanstrom, R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc Natl Acad Sci USA* **108**, 20166–20171, doi: 10.1073/pnas.1110064108 (2011).

Acknowledgements

We thank Chanson Brumme for database assistance, Rachel McGovern, Manal Rahman, and Denis Chopera for laboratory assistance and Xiaomei (Tallie) Kuang for critical review of the manuscript and figures. We thank Evan Wood and P. Richard Harrigan for specimen and data access. We also thank Dr.

Harrigan for access to NGS sequencing and for helpful discussions. This study was supported by operating grants from the Canadian Institutes for Health Research (CIHR) MOP-93536 to ZLB, HOP-111406 to AFYP, HOP-115700 to ZLB/AFYP, and by The Canadian HIV Cure Enterprise Team Grant HIG-133050 (to ZLB/AFYP) from the CIHR in partnership with CANFAR and IAS. The VIDUS and ACCESS studies were supported by the United States National Institutes of Health (VIDUS: U01DA038886; ACCESS: R01DA021525). This research was undertaken, in part, thanks to funding from the Canada Research Chairs program through a Tier 1 Canada Research Chair in Inner City Medicine that supports Dr. Evan Wood. AQL was funded by the Canadian Institutes of Health Research (CIHR) Frederick Banting and Charles Best Masters Award. RM is supported by a scholarship from the CIHR Strategic Training Program in Bioinformatics at the University of British Columbia. RD was supported by the Merck-Canada Training of Aboriginal Youth in Biomedical Labs (TAYBL) program at Simon Fraser University. KH is supported by a CIHR Fellowship. KH is supported by a CIHR New Investigator Award (MSH-141971). M-JM is supported in part by the United States National Institutes of Health (NIDA R01-DA021525). AFYP is the recipient of a Scholar Award from the MSFHR/St. Paul's Hospital Foundation - Providence Health Care Research Institute Career Investigator program, and a CIHR New Investigator Award. ZLB was supported by a CIHR New Investigator Award and currently holds a Scholar Award from the Michael Smith Foundation for Health Research.

Author Contributions

A.Q.L., A.F.Y.P. and Z.L.B. conceived the study, experiments and analyses. A.Q.L., J.T., W.D. and R.D. collected the data. A.Q.L., R.M., C.W. and A.F.Y.P. analyzed and interpreted the data. K.H. and M.-J.M. provided cohort access and expertise. A.Q.L. and Z.L.B. drafted the manuscript, which was critically reviewed by all authors.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Le, A. Q. *et al.* Differential evolution of a CXCR4-using HIV-1 strain in CCR5wt/wt and CCR5 Δ 32/ Δ 32 hosts revealed by longitudinal deep sequencing and phylogenetic reconstruction. *Sci. Rep.* 5, 17607; doi: 10.1038/srep17607 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>