JOURNAL OF VIROLOGY, Apr. 2009, p. 3556–3567 Vol. 83, No. 8
0022-538X/09/$08.00+0 doi:10.1128/JVI.02132-08
Copyright © 2009, American Society for Microbiology. All Rights Reserved.

Quantitating the Multiplicity of Infection with Human Immunodeficiency Virus Type 1 Subtype C Reveals a Non-Poisson Distribution of Transmitted Variants▽†

M.-R. Abrahams,[1] J. A. Anderson,[2] E. E. Giorgi,[3,4] C. Seoighe,[1] K. Mlisana,[5] L.-H. Ping,[2] G. S. Athreya,[3] F. K. Treurnicht,[1] B. F. Keele,[6] N. Wood,[1] J. F. Salazar-Gonzalez,[6] T. Bhattacharya,[3,7] H. Chu,[2] I. Hoffman,[2] S. Galvin,[2] C. Mapanje,[8] P. Kazembe,[8] R. Thebus,[1] S. Fiscus,[2] W. Hide,[9] M. S. Cohen,[2] S. Abdool Karim,[5] B. F. Haynes,[10] G. M. Shaw,[6] B. H. Hahn,[6] B. T. Korber,[3,7] R. Swanstrom,[2] and C. Williamson[1]* for the CAPRISA Acute Infection Study Team and the Center for HIV-AIDS Vaccine Immunology Consortium

Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa[1]; University of North Carolina at Chapel Hill, Chapel Hill, North Carolina[2]; Los Alamos National Laboratory, Los Alamos, New Mexico[3]; University of Massachusetts, Amherst, Massachusetts[4]; Centre for the AIDS Programme of Research in South Africa, University of Kwa-Zulu Natal, Durban, South Africa[5]; University of Alabama at Birmingham, Birmingham, Alabama[6]; Santa Fe Institute, Santa Fe, New Mexico[7]; Kamuzu Central Hospital, Lilongwe, Malawi[8]; South African Bioinformatics Institute, University of Western Cape, Cape Town, South Africa[9]; and Duke University Medical Center, Durham, North Carolina[10]

Received 9 October 2008/Accepted 24 December 2008

Identifying the specific genetic characteristics of successfully transmitted variants may prove central to the development of effective vaccine and microbicide interventions. Although human immunodeficiency virus transmission is associated with a population bottleneck, the extent to which different factors influence the diversity of transmitted viruses is unclear. We estimate here the number of transmitted variants in 69 heterosexual men and women with primary subtype C infections. From 1,505 env sequences obtained using a single genome amplification approach we show that 78% of infections involved single variant transmission and 22% involved multiple variant transmissions (median of 3). We found evidence for mutations selected for cytotoxic-T-lymphocyte or antibody escape and a high prevalence of recombination in individuals infected with multiple variants representing another potential escape pathway in these individuals. In a combined analysis of 171 subtype B and C transmission events, we found that infection with more than one variant does not follow a Poisson distribution, indicating that transmission of individual virions cannot be seen as independent events, each occurring with low probability. While most transmissions resulted from a single infectious unit, multiple variant transmissions represent a significant fraction of transmission events, suggesting that there may be important mechanistic differences between these groups that are not yet understood.

The development of a human immunodeficiency virus (HIV) vaccine remains a global priority. Since vaccines need to target viruses at, or close to, the time of infection, there is great interest in understanding both the genetic characteristics of viruses that are successfully transmitted and the genetic diversification that ensues during the earliest stages of infection. There is evidence from cross-sectional studies of individuals with acute and early HIV infections that transmission is associated with a population bottleneck (4, 9, 16, 35, 36, 39, 46–48). The best evidence of transmission bottlenecks is obtained from studies of HIV discordant couples, where infection is associated with transmission of one or a few genetic variants despite high degrees of viral diversity in the transmitting donors (13). A recent study was able to quantify this bottleneck by using methodologies that could identify precisely the transmitted, founder virus population (16). That study found that as many as 76% of transmission events involved productive clinical infection by only a single genetic variant in HIV-1 subtype B transmission by either heterosexual or homosexual routes. These findings are encouraging since in the majority of infections a vaccine-mediated response may have to protect against only a low-multiplicity infection event that is associated with low viral diversity in the initial days and weeks of infection.

However, it is also apparent that the transmission bottleneck can be overcome, as evidenced by the transmission of multiple viral variants (9, 10, 12, 16, 35, 36). The clinical implications of multiple variant transmission are potentially serious because coinfection with genetically divergent viral lineages has been associated with more severe disease progression (10, 12, 37). Current estimates of the frequency of multiple variant transmissions vary widely with between 0 and 50% of successful sexual transmissions estimated to involve the transfer of multiple genetic variants (9, 16, 35).

It is still an open question as to whether there are distinctive features of viruses that enhance their transmissibility and whether different risk behaviors are associated with different

* Corresponding author. Mailing address: Institute of Infectious Diseases and Molecular Medicine, Division of Medical Virology, Faculty of Health Sciences, University of Cape Town, Observatory, Cape Town 7925, South Africa. Phone: 27-21-406 6683. Fax: 27-21-406 6682. E-mail: carolyn.williamson@uct.ac.za.

† Supplemental material for this article may be found at http://jvi.asm.org/.

▽ Published ahead of print on 4 February 2009.

numbers of transmitted viruses. For example, intravenous infection has been associated with more heterogeneous infections than intravaginal infection in a rhesus macaque model (11). Besides the potential identification of previously unrecognized transmission risk factors, attempts to discover the variables influencing multivariant transmission event frequencies should also provide insights into natural barriers to HIV infection. Unfortunately, because of methodological variations between studies it has not been possible to directly compare results and to assess the impacts of biological factors such as gender, routes of transmission, the presence of sexually transmitted diseases and viral subtype on multiple variant transmission frequencies (39). Complex cofactors in a given risk category may impact rates of transmission of multiple variants just as they impact overall transmission rates (32).

In the present study, we investigated the genetic characteristics of subtype C variants transmitted via a heterosexual route in order to understand mechanisms involved in HIV transmission at genital sites. We used here the approach described by Keele et al. (16), which allowed us to directly compare multiple transmission frequencies between subtype B and C viruses. Our results indicate that similar proportions of subtype B and subtype C infections involve the transmission of multiple variants. Analysis of a total of 171 transmission events in both the subtype B and C studies showed that infection with multiple variants does not follow a Poisson distribution, suggesting that transmissions of multiple variants are not independent events in a setting of a low probability of infection. The question of the biological basis of the mucosal viral transmission bottleneck and factors associated with its breach, however, remains unanswered.

MATERIALS AND METHODS

Cohort description. Plasma was obtained from 69 individuals experiencing early HIV-1 subtype C infection. Twenty-six were identified through prospective monthly monitoring of HIV-negative women using rapid HIV tests and RNA PCR (44) (CAPRISA 002 cohort, Durban, South Africa); 22 were identified (i) with negative rapid HIV or enzyme immunoassay assay (EIA) antibody test results but positive HIV RNA and/or p24 antigen test results or (ii) who were antibody positive but antibody negative or Western blot indeterminate within the previous 45 days (CHAVI 001 cohorts, South Africa and Kamazu Central Hospital Sexually Transmitted Disease Clinic, Malawi). An additional 21 individuals enrolled from the Sexually Transmitted Disease Clinic in Malawi were identified as HIV PCR positive with a rapid HIV or EIA antibody-negative test or who had indeterminate HIV Western blot results (30).

This study was approved by institutional review boards, and all participants provided written informed consent.

Staging of HIV infection. The durations of HIV-1 infection were categorized into six stages based on evolving HIV-1 RNA or antibody profiles developed by Fiebig et al. (8). Plasma was tested for HIV RNA by using Roche Amplicor vRNA assays (Rotkreuz, Switzerland) and for antibodies by EIA (BEP 2000 [Dade Behring, Marburg, Germany] or Determine AntiHIV-1/2 3rd Generation EIA [Abbott, Illinois] and Uni-Gold Recombigen [Trinity Biotech, Ireland]) and a GS HIV-1 Western blot analysis kit (Bio-Rad, Washington). Individuals classified as being in stage I were viral RNA positive and p24 antigen and EIA antibody negative, those in stage II were RNA and p24 antigen positive but EIA antibody negative, those in stage III were EIA antibody positive but Western blot negative, those in stage IV had an indeterminate Western blot result, those in stage V were Western blot positive but without reactivity to the p31 integrase band, and those in stage VI were Western blot positive with a p31 band present.

PCR amplification and sequencing. HIV-1 RNA was extracted from 140 to 200 μl of ACD or EDTA plasma and eluted in a 50-μl final volume. The full volume of RNA was reverse transcribed to cDNA (100-μl reaction volume) by using a Superscript III reverse transcriptase (RT) system (Invitrogen, California) with an OFM-19 primer as described previously (16) or with oligo(dT). The cDNA was then serially diluted to obtain no more than 30% positive amplification reactions so that each amplicon would theoretically be amplified from a single template more than 80% of the time (39). This limiting dilution approach of *env* gene amplification was initially described by Simmonds et al. (41) and Edmonson and Mullins (7) and was thereafter modified by Palmer et al. (29) by sequencing of the amplicon directly and finally modified by Salazar-Gonzalez et al. (39), who showed that single genome amplification of the *env* gene with direct sequencing precludes recombination and Taq-induced error, and provides proportional representation of each viral sequence. PCR products were directly sequenced by using an ABI 3000 genetic analyzer (Applied Biosystems, Foster City, CA) and BigDye terminator reagents. To ensure that sequences reflected single templates from the viral populations in vivo, amplicons with sequence chromatograms with "double peaks," indicative of coamplification of more than one template, were excluded. Sequences with deletions larger than 100 nucleotides compared to the intraparticipant consensus were excluded.

Sequence analysis. Differences in sequences were visualized by using neighbor-joining trees (MEGA 3.1) (43) and Highlighter nucleotide transition and transversion plots (www.hiv.lanl.gov). Pairwise DNA distances were computed by using MEGA 3.1.

Conformance of intraparticipant sequence diversity to a mathematical model of random evolution was evaluated as described by Keele et al. (16) and Lee et al. (H. Y. Lee, E. E. Giorgi, B. F. Keele, B. Gaschen, G. S. Athreya, J. F. Salazar-Gonzalez, K. T. Pham, P. A. Geopfert, J. M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, B. H. Hahn, G. M. Shaw, B. T. Korber, T. Bhattacharya, and A. S. Perelson, submitted for publication) whereby exponential viral replication from a single lineage is assumed to fix mutations at a constant rate, in the absence of positive selection, using the following parameters: an HIV-1 generation time of 2 days (24), a reproductive ratio of 6 (i.e., assuming that the virus replicates exponentially, for each currently infected cell six new cells will be infected in the next generation) (42), and a replication error rate of 2.16×10^{-5} substitutions per site per replication cycle (23).

Time of divergence from the most recent common ancestor (MRCA) was estimated by using BEAST (i.e., Bayesian Evolutionary Analysis Sampling Trees, v1.4.7) (5, 6) with a relaxed (uncorrelated exponential) molecular clock and general time-reversible substitution model, with relative substitution rate parameters estimated by using HyPhy (31), as described previously (16). We used a gamma distribution with four categories and a proportion of invariant sites to model rate heterogeneity across sites. Substitution rates were unlinked across codon positions with a mean fixed at 2.16×10^{-5} substitutions per site per generation (23).

Sequences were analyzed for evidence of APOBEC3G-induced hypermutation by using the Hypermut 2.0 tool (www.hiv.lanl.gov). Sequences with a P value of ≤ 0.1 were considered enriched for mutations consistent with APOBEC3G signatures. In sequence sets showing evidence of enrichment for APOBEC3G-driven G-to-A transitions but with no single significantly hypermutated sequence, hypermutation was tested for after superimposition of all mutations within that sequence set onto a single representative sequence.

We used randomization to test whether there was evidence of clustering of mutations within 10-amino-acid stretches putatively associated with cytotoxic-T-lymphocyte (CTL) immune responses (14). Sites were classified as mutated in a given patient if there was at least one sequence from the patient differing from the patient consensus at that site. For each mutated site, we calculated its nearest-neighbor distance as the distance between the site and the closest mutated site to the left or right of it on the intrapatient sequence alignment. We then compared the number of mutations with a nearest neighbor within 10 amino acids to the number expected by chance by randomizing the locations of the mutated sites 1,000 times. A P value was calculated as the fraction of randomized datasets for which the proportion of mutated sites within 10 amino acids of another mutated site was equal to or greater than the observed proportion. The null hypothesis here is random distribution of mutated sites, which is consistent with the model of neutral drift of the infecting virus in acute infection, proposed by Keele et al. (16). Rejection of the null hypothesis suggests clustering of the mutations on a scale consistent with escape from CTL responses.

Multivariant transmission was considered if (i) within-patient *env* diversity was heterogeneous, with multimodal distribution of pairwise Hamming distances (HDs; that is, the number of differing sites between sequences) and structure within the phylogenetic trees, and (ii) if these deviations from the model of random evolution from a single founder virus could not be accounted for by APOBEC3G mutations, immune pressure, or stochastic events. For individuals infected with more than one variant, our expectation was that when multiple variants were transmitted, the sequences in the recipient should coalesce at a time predating the estimated time of infection. The number of infecting variants was enumerated, after accounting for recombination, by identification of distinct

TABLE 1. Demographic, clinical, and virological characteristics of individuals infected with a single variant compared to infection with multiple HIV-1 variants

Parameter	Individuals infected with:	
	A single variant	More than one variant
Cohort		
Total no. of subjects	54	15
No. of female subjects	32	7
No. of male subjects	22	8
No. of subjects at laboratory stage of infection:		
I/II	17	5
III/IV	14	6
V/VI	23	4
Mean age in yr (female) ± SD	29 ± 10	25 ± 7
Infecting virus		
Maximum % DNA distance range	0.08–0.64	0.32–8.49
Median no. of infecting variants (range)	1	3 (2–5)
Viral load at 12 mo postinfection[a]	39,300 ± 106,072 (400–376,000)	70,100 ± 532,413 (400–1,230,000)
CD4 counts at 12 mo postinfection[b]	468 ± 193 (303–1,030)	274 ± 251 (188–695)

[a] Expressed as median copies/ml ± the standard deviation (n = 24).
[b] Expressed as median cells/μl ± the standard deviation (n = 24).

lineages on phylogenetic trees, together with examination of Highlighter transition and transversion plots (www.hiv.lanl.gov).

N-linked glycosylation sites were identified by using the N-glycosite program (www.hiv.lanl.gov). HXB2 *env* protein amino acid locations for putative epitopes were obtained by using the HIV sequence locator tool (www.hiv.lanl.gov). Motif Scan (www.hiv.lanl.gov) was used to detect HLA anchor residue motifs within putative epitopes and to search for matching potential epitopes.

Recombination analysis. The GARD (www.datamonkey.org/GARD/) (17, 18), RAP Beta version (www.hiv.lanl.gov), and RDP version 3.27 (http://darwin.uvigo.es/rdp/rdp.html) (25) tools were used to detect recombination in intraparticipant sequence sets.

Modeling the probability of multivariant infection. We have assumed that established infection with a single genotypic variant signifies that a single virus particle was involved in the transmission event and that, in the setting of low probability of transmission, this represents the minimal infectious dose. The Poisson distribution was used to model the frequency of transmission of one, two, or more variants under the assumption that the transmission of each variant occurs with the same probability, i.e., as independent events. A left truncated Poisson model was used (since the zero events, i.e., no transmission, are not observed), and the model fitted to all of the data with a maximum-likelihood method; this then allowed the frequency of zero events to be estimated given the distribution of one, two, etc., variants being detected. A corresponding transmission probability was estimated as the sum of all probabilities for observing one or more variants.

Statistical testing. Categorical variables were compared by using Fisher exact tests (two-tailed). P values of <0.05 were considered significant.

GenBank accession numbers. The GenBank accession numbers for the 1,505 *env* sequences are FJ443128 to FJ444362.

RESULTS

Cohort description and timing of infection. Blood plasma samples were obtained from 69 study participants (30 men and 39 women) from South Africa and Malawi who had recently acquired HIV-1 infection through heterosexual contact (Table 1). A total of 42 of the participants were in the very early stages of infection: 22 were in stage I/II (viral RNA positive, p24 antigen and EIA antibody negative, or RNA and p24 antigen positive but EIA antibody negative), 6 were in stage III (EIA antibody positive but negative by Western blotting), and 14 were in stage IV (EIA antibody positive with an indeterminate Western blot result). Individuals in these stages had high viral loads (mean of >6 million copies/ml) typical of acute infection (Fig. 1). Twenty-seven participants were in the later stages of primary infection, eighteen in stage V (Western blot positive but without reactivity to the p31 integrase band) and eight in stage VI (Western blot positive with a p31 band present), with lower viral loads (medians of 84,572 and 21,176 copies/ml, respectively); one individual was classified as being in stage V or VI, with reactivity to p31 integrase not yet determined.

Characterization of *env* gene diversity. A total of 1,505 *env* gene sequences encoding gp160 (with an average of 22 se-

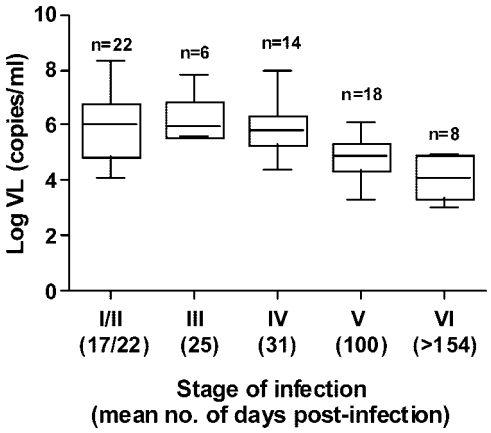


FIG. 1. Log viral loads of 68 participants categorized into stage I/II (viral RNA positive, p24 antigen and EIA antibody negative, or RNA and p24 antigen positive but EIA antibody negative), stage III (EIA antibody positive but negative by Western blot), stage IV (EIA antibody positive with an indeterminate Western blot result), stage V (Western blot positive but without reactivity to the p31 integrase band), and stage VI (Western blot positive with a p31 band present) (8). Boxes represent the 70th percentile; horizontal bars represent the median, and whisker bars correspond to minimum and maximum values. p24 antigen tests, which differentiate stages I and II of infection, were not carried out for all participants; therefore, HIV RNA-positive, enzyme-linked immunosorbent assay-negative individuals were classified as stage I/II. One participant was classified as being in stage V or VI of infection and is thus not included in this figure. The mean number of days postinfection was determined by Fiebig et al. (8) and modified by Keele et al. (16).

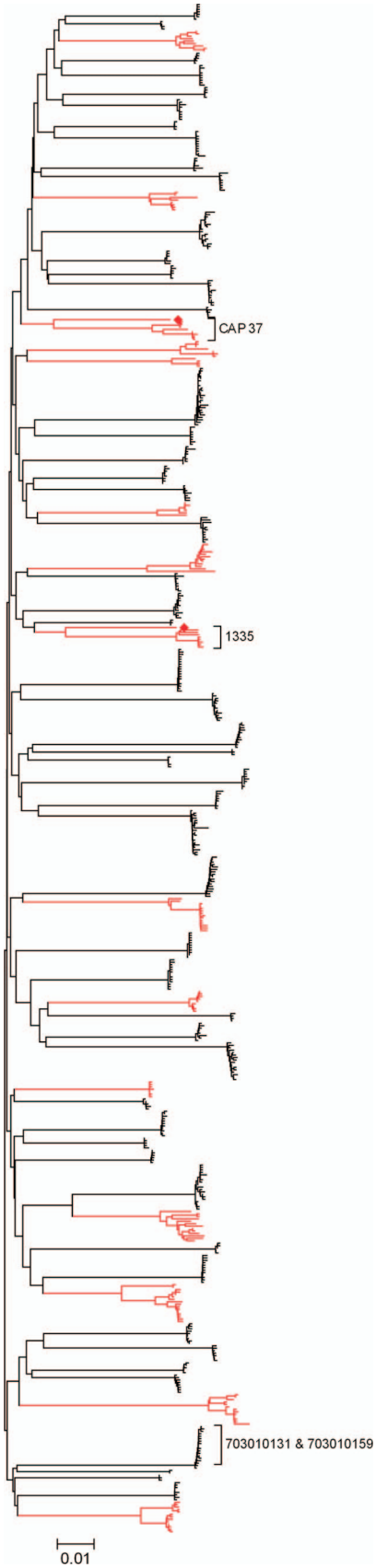


FIG. 2. Neighbor-joining tree of *env* sequences from each of the 69 study participants from South Africa and Malawi. To limit the size of the tree, only 626 sequences of the 1,505 sequences generated were analyzed: where individuals harbored viruses with multiple identical *env* sequences, only the unique sequences were included. Red branches represent sequences from participants infected with more than one variant. Diamonds represent sequences with >6% divergence from the participant sequence population. The sequences for participants 703010131 and 703010159, a donor-recipient pair, cluster together, as indicated.

quences per participant; range, 15 to 42) were generated from the 69 participants using single genome amplification and direct sequencing of amplicons. All sequences were classified as HIV-1 subtype C (Fig. 2). Fourteen participants clearly harbored highly heterogeneous viral populations (maximum intrapatient DNA distances of 0.73 to 8.49%) forming more than one phylogenetic lineage. These individuals were postulated to be infected with multiple variants and were further characterized to quantify the number of infecting variants. The remaining 55 participants harbored virus populations with lower diversity (maximum intrapatient DNA distances ranging from 0.08 to 0.64%) with sequences clustering as single lineages with no, or very little, structure within phylogenetic trees (see Tables S1A to S1C in the supplemental material). On further analysis, one of the 55 individuals with a low diversity *env* population, participant 1176, was subsequently found to be infected with three closely related viruses.

All sequences were initially analyzed for evidence of APOBEC3G-induced G-to-A hypermutation and, after removal of these mutations from the relevant sequence sets, all sequences were compared to a model of random evolution in which infection with a single founder virus is followed by neutral expansion characterized by exponential population growth (16). This is expected to result in a star-like phylogeny and a Poisson distribution for the intrapatient HD. Sequence sets that did not conform to the model after the removal of the G-to-A hypermutation were analyzed for evidence of immune pressure, selection, or evidence of scattered APOBEC3G-related hypermutation, although not detected as significant ($P > 0.1$), to determine possible reasons for the deviation from the model. Finally, the time of infection estimated from laboratory staging was compared to the estimated time to the MRCA of the sequences.

Two study participants were identified as being a donor-recipient pair based on their sexual history and the close phylogenetic linkage of their viruses (Fig. 2). Participant 703010131 was assumed to be the donor since he was classified as being in stage III infection, while the assumed recipient, participant 703010159, was found to be in stage I/II.

Low *env* diversity after transmission. Of the 55 individuals harboring viruses with low diversity, sequences from 30 individuals displayed Poisson distributed HDs and a star-like phylogeny consistent with infection with a single founder virus and with the subsequent incorporation of randomly distributed mutations (Fig. 3A and see Table S1A in the supplemental material). Sequences from a further five individuals conformed to the model following the removal of G-to-A substitutions embedded in APOBEC3G signature patterns after showing sig-

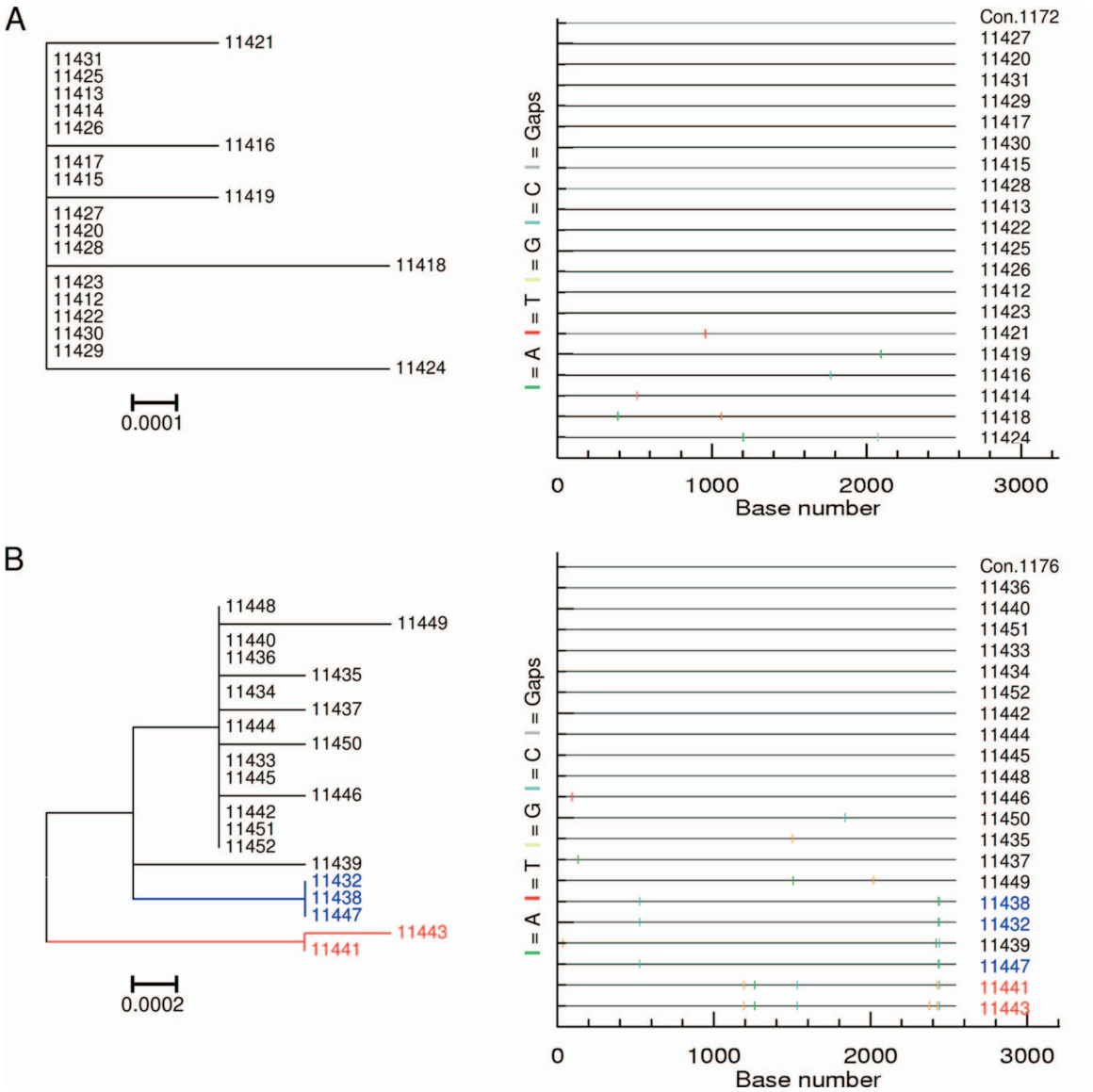


FIG. 3. *env* sequence diversity was visually determined by the structure of the phylogenetic tree (left), and the pattern of nucleotide base mutations within sequences was observed on a Highlighter plot (right). The Highlighter plots compare sequences from each participant's sequence set to an intraparticipant consensus (uppermost sequence) and illustrate the positions of nucleotide base transitions and transversions using short, color-coded bars. (A) Participant 1172 with a highly homogeneous *env* sequence population displaying limited structure on a tree and a few nucleotide changes from the intraparticipant consensus. (B) Participant 1176 infected with three closely related *env* populations (indicated in black, blue, and red, respectively) based on the clustering of sequences into individual clades on a tree and the shared patterns of mutations observed on a Highlighter plot. Both participants were viral RNA positive but ELISA negative (stage I/II of infection).

nificant enrichment of these substitutions within a single sequence and/or within the overall participant sequence set with no single overtly hypermutated sequence ($P < 0.1$) (see Table S1B in the supplemental material). In a sixth individual with significant APOBEC3G-driven hypermutation ($P = 0.0067$), participant CAP85, sequences did not conform to a star-like phylogeny even after removal of APOBEC-driven substitutions (data not shown). This was not investigated further. Sequence sets for an additional two individuals (CAP225 and 704810053) conformed to the model following removal of scattered G-to-A mutations, despite the fact that this hypermutation was not significant ($P > 0.1$) (see Table S1B in the supplemental material).

Thus, in seven of eight individuals, the model violation prior to the removal of G-to-A substitutions would likely be due to an increased relative mutation rate in these sites. This result is similar to that reported for a small set of individuals identified in the acute subtype B-infected cohort of Keele et al. (16), where an overall enrichment of G-to-A substitutions in APOBEC motifs was scattered throughout the available sequences in 7 of 81 homogeneous infections, with an additional six subjects carrying one or more overtly hypermutated sequence. The

balance between HIV-1 Vif and APOBEC3G (22) may be altered in such cases. In addition, a recent study suggests that this type of pattern may be associated with improved clinical outcome (19).

The remaining 17 of the 55 individuals with low-diversity *env* populations had no evidence of enrichment for G-to-A hypermutation but harbored virus populations that did not conform to the model due to their having either non-Poisson-distributed HDs and/or non-star-like phylogenies. Six of these individuals were classified as infected with a single founder virus based on divergence within a single lineage ($n = 2$) and/or on patterns of shared mutations between sequences ($n = 4$) (i.e., internal branching in the intrapatient tree inconsistent with the expected star-like phylogeny). These shared mutations were possibly due to stochastic accumulation of neutral mutant alleles, either as neutral mutations that occurred very early in the infection and could thus be established and retained in a high enough frequency to be sampled or due to very early CTL escape mutations.

A further 7 of these 17 individuals had evidence of early antibody pressure with sequences showing loss or gain of N-linked glycosylation sites or changes in envelope loop length; all 7 were recruited in stages IV to VI of infection (Fig. 4). In four of the seven participants, these changes occurred in the V1/V2 loop region, which may indicate that this area is an early target for immune pressure during primary infection.

Three of the seventeen individuals had evidence of cellular immune pressure. In one individual there was a mutation associated with a known epitope (CAP8, HLA-B*0801). In the other two individuals there was clustering of mutations within 10-mers (participants 70310054 and 705010015) (Fig. 4), a phenomenon previously found to be associated with CTL escape (14). The clustered mutations within these two participants' 10-mers fell within sequences found to contain HLA anchor residue motifs for a range of HLA genotypes and which were also matched to potential epitopes by Motif Scan (www.hiv.lanl.gov). However, since the HLA types for these two individuals are not known, the inference of putative CTL escape remains speculative, and the relevance of these mutations unknown.

Finally, the last individual harboring low diversity virus, subject 1176, had sequences that exhibited shared mutations between sequence subsets and an estimated time to MRCA of 88 days, which exceeded the estimated time of infection since this individual was in stage I/II infection; thus, we infer that this subject was infected with three closely related viruses (average DNA distance, 0.1%) (Fig. 3B).

In summary, of the 55 individuals with low diversity after transmission, 54 were classified as likely to be infected with a single variant, and 1 was classified as likely to be infected with three closely related variants. For participants identified as infected with a single infectious unit, the estimated time to the MRCA was consistent with the expected time of infection with one exception, participant CAP217, whose infecting virus displayed lower genetic diversity (mean, 0.02%) than expected, given this participant was in Fiebig stage IV of infection (see Table S1A in the supplemental material).

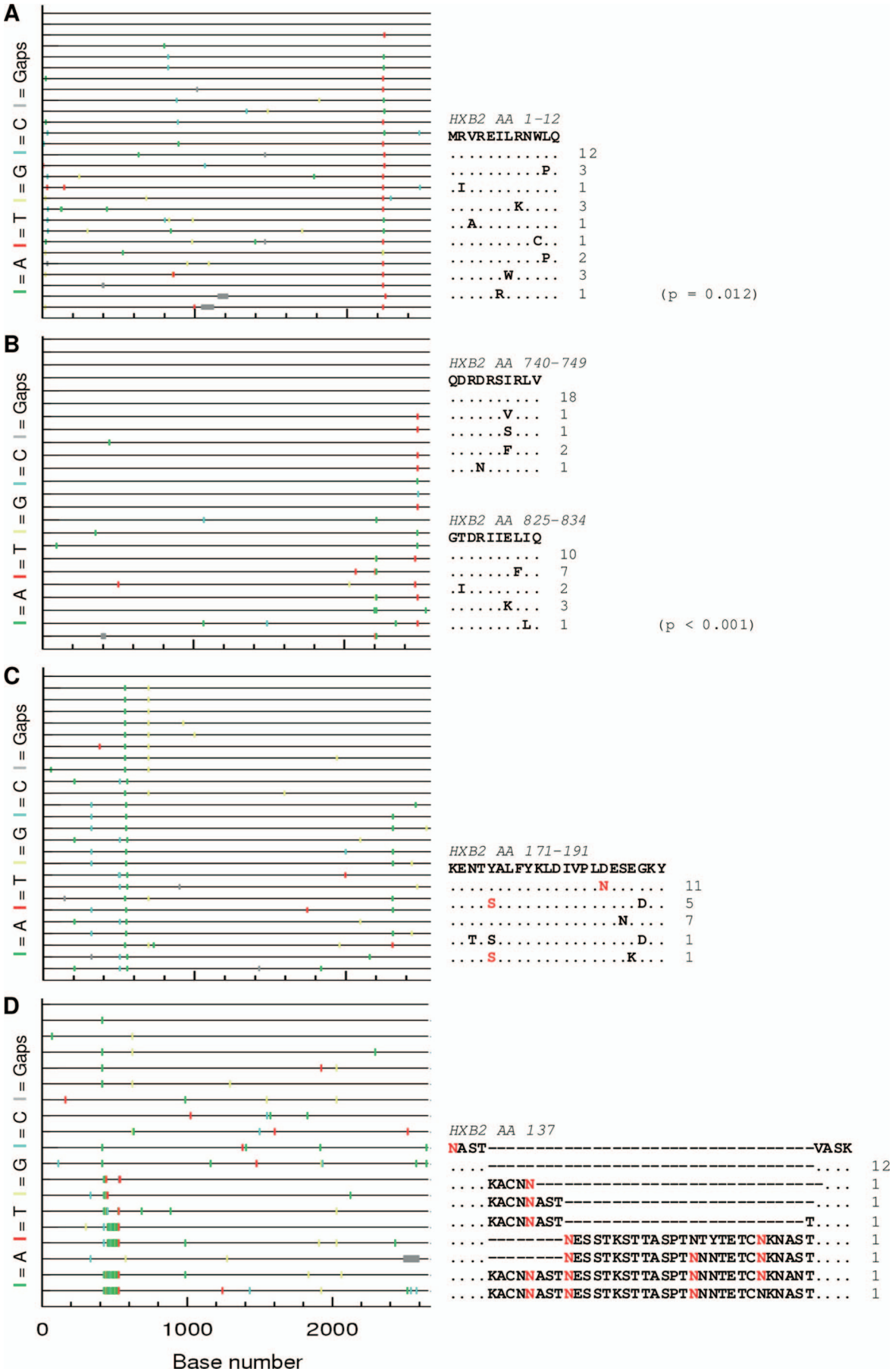
Infection with multiple unique infectious units. After correcting for recombination and hypermutation, phylogenetic analyses of the high diversity sequence sets indicated that all 14 were infected with more than one viral variant (see Table S1D in the supplemental material). Including participant 1176 infected with three closely related viruses brings the total number of individuals with multivariant infections to 15. Interlineage recombination between transmitted variants was observed in 10 of the 14 individuals, using the Recombination Analysis Program (www.hiv.lanl.gov) (Fig. 5). In 9 of these 10 cases, recombination was also detected by GARD or RDP3.27 (18, 19, 26) or both. However, since donor samples were not available for these individuals, the transmission of recombinants cannot conclusively be ruled out. In all 14 sequence sets, the estimated number of days since the MRCA significantly exceeded the period for which the associated individual could realistically have been infected (MRCA range, 605 to 5,998 days).

Therefore, including the individual infected with three closely related variants, a total of 15 individuals were inferred to be infected with multiple viruses. The median number of readily distinguishable infectious units per individual was three (range, two to five), with each viral lineage having a unique pattern of nucleotide variation (Fig. 5). This represents the minimum number of transmitted viruses in these individuals, given the limitations in sampling. While we cannot exclude the possibility that multiple variants were transmitted and a single variant grew out at the time of detection, there is no bias for the detection of multiple variants at earlier Fiebig stages (Table 1).

Two of the fourteen participants, CAP37 and 1335, were both infected with viruses with *env* sequences differing from one another by more than 6% nucleotide sequence identity (Fig. 2). Although the complete *env* sequences from these individuals clustered as an outlier to the participant sequences, the region encoding gp41 separated into distinct phylogenetic branches separated by epidemiologically unlinked sequences. This suggests that these two subjects had dual infections, although it is not possible to determine whether these individuals were infected by two independent transmission events from different donors or if variants were cotransmitted from a donor who had a dual infection.

Recombinant genomes result from the dual infection of cells, followed by recombination in a subsequent round of infection and thereafter outgrowth, allowing detection. We sought to determine whether the detection of recombination was related to the Fiebig stage. To have a large enough data set, we pooled data from Fiebig stage I to III to compare them to Fiebig stage IV to VI and also included the data reported by Keele et al. (16). We detected the presence of recombinant viruses at a significantly higher rate in the later Fiebig samples ($P = 0.0015$). In most individuals, we detected each recombinant genome once. While we can conclude from these results that the detection of recombinants becomes more likely with later stages of infection, without corresponding donor samples the possibility that detected recombinants were transmitted cannot be ruled out. In particular, this could be the case with participant CAP69 (stage I/II of infection) in whom we saw the outgrowth of some recombinants (Fig. 5B).

Clustering of mutations. We screened all sequences for evidence of the clustering of mutations. We identified tight clusters of mutations by visual inspection, as well as by means of a test based on randomizing the locations of mutated sites within each patient in order to determine whether these mutations


A
= A = T = G = C = Gaps
HXB2 AA 1-12
MRVREILRNWLQ
..... 12
.....P. 3
.I..... 1
.....K.... 3
..A..... 1
.....C.. 1
.....P. 2
.....W..... 3
.....R..... 1 (p = 0.012)
B
= A = T = G = C = Gaps
HXB2 AA 740-749
QDRDRSIRLV
..... 18
.....V... 1
.....S... 1
.....F... 2
...N..... 1
HXB2 AA 825-834
GTDRIIELIQ
..... 10
.....F.. 7
.I..... 2
.....K... 3
.....L. 1 (p < 0.001)
C
= A = T = G = C = Gaps
HXB2 AA 171-191
KENTYALFYKLDIVPLDESEGKY
.....N..... 11
....S.....D.. 5
.....N.... 7
..T.S.....D.. 1
....S.....K... 1
D
= A = T = G = C = Gaps
HXB2 AA 137
NAST----------VASK
....----------.... 12
....KACNN----------... 1
....KACNNAST----------.... 1
....KACNNAST----------T.... 1
....--------NESSTKSTTASPTNTYTETCNKNAST.... 1
....--------NESSTKSTTASPTNNNTETCNKNAST.... 1
....KACNNASTNESSTKSTTASPTNNNTETCNKNANT.... 1
....KACNNASTNESSTKSTTASPTNNNTETCNKNAST.... 1
0
1000
2000
Base number

were significantly more clustered than would be expected by chance. We focused specifically on clustering on a length scale of 10 amino acids (roughly the size of a CTL epitope), since the presence of clusters of mutations within a region of approximately this size is consistent with selective pressure resulting in evasion of early cellular immune responses such as CTL responses (14). We determined the number of mutations with a nearest neighbor within 10 amino acids and compared this to the number expected by chance, estimated by randomizing the locations of the mutated sites 1,000 times. Under a model of neutral divergence (16), we expect mutated sites to be distributed randomly throughout the sequence. In addition, tight clustering of mutations is unlikely even if purifying selection were to affect a proportion of mutations. This would thus suggest that mutations within these clusters would be favored by selection.

Significant clustering was identified in 16 individuals (Fig. 4). Nine of these individuals with clustered mutations had single variant infections; six of these subjects (CAP129, CAP217, 0626, 703010193, 703010217, and 706010164) harbored sequences consistent with the model of neutral evolution from a single founder virus (see Table S1A in the supplemental material), suggesting that the virus populations from these individuals were under early selection despite the failure to reject the model of neutral evolution. Clustered mutations were detected in seven individuals with multivariant transmission.

As expected, most participants harboring sequences with evidence of clustered mutations were in later stages of primary infection (stages IV to VI) although, interestingly, three individuals with very early infection (stages I to III) harbored sequences with clustered mutations, providing evidence for very early selective pressure. In the absence of sequence data from donor individuals, however, it is unknown whether any of these mutations were transmitted.

Multiple variants are not transmitted as independent events with low probability. Transmission of HIV-1 is infrequent (32), and the homogeneity of sequences in acute infection suggests that most transmission events represent the minimal infectious dose. For both the subtype B (16) and the subtype C datasets we found that ca. 80% of transmissions result in infection by a single genotypic variant or a single virus particle. To determine whether transmission of multiple variants represents independent but concurrent infectious events where transmission of the first variant had no effect on the probability of transmission of the second (or third) variant, we modeled the number of transmitted variants using the Poisson distribution and estimated a putative transmission probability as the likelihood that one or more variants are transmitted (equal to 1 − the probability that no variants are transmitted). In Fig. 6 we show the expected number of infections with one, two, or more than two variants for when the probability of transmission in a single exposure is 0.1, 0.25, and 0.4. Transmission probabilities in this range are needed to observe the transmission of multiple variants at significant levels. In contrast, a more realistic transmission rate of 0.01, for example, would result in the transmission of two variants once in 10,000 transmission events.

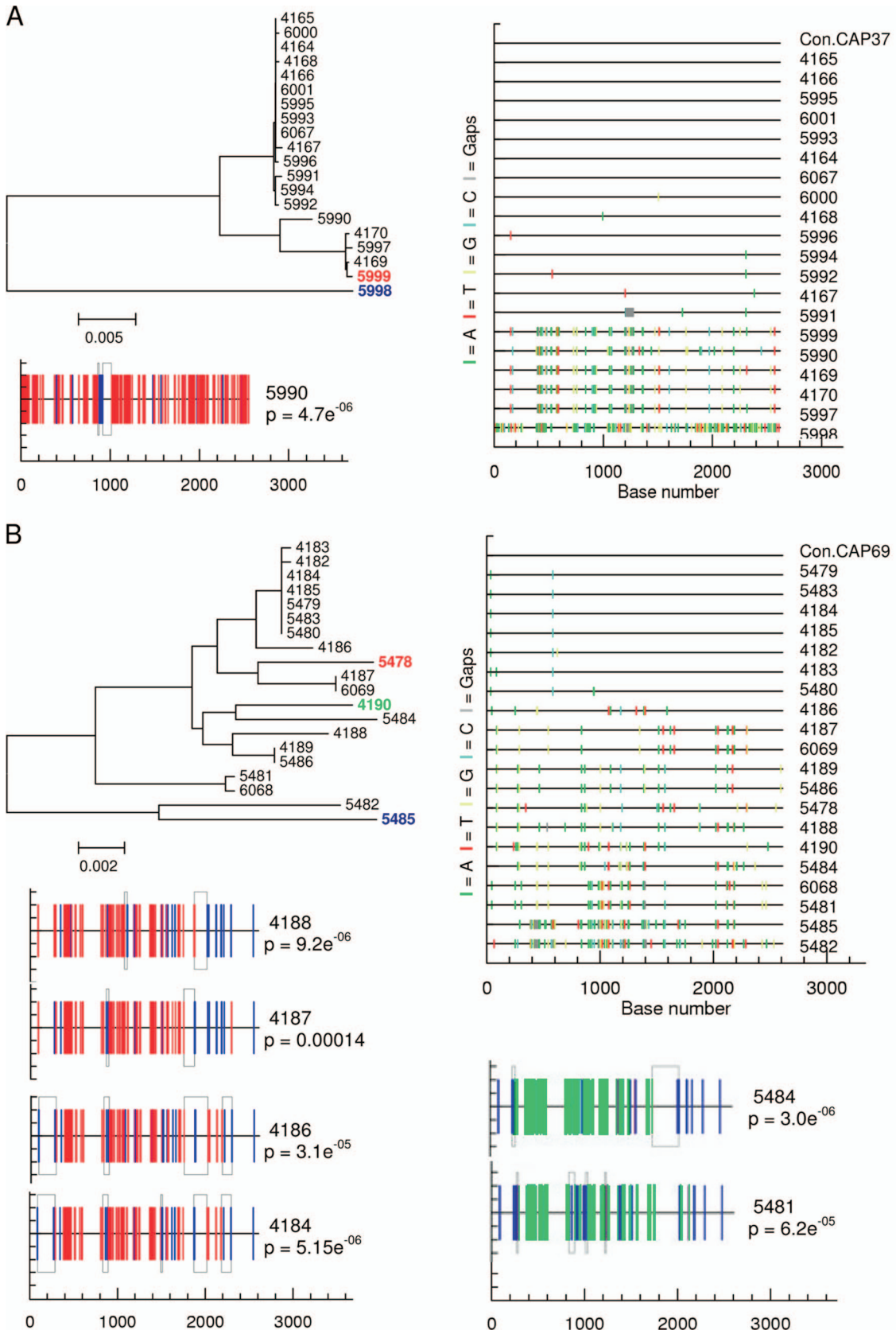
In Fig. 6 we also show the observed number of individuals infected with one, two, or more than two variants in the subtype B and subtype C primary infection cohorts. When these data are fitted to the Poisson model we estimate a transmission rate of 0.48 for the subtype C data set (Poisson mean of 0.65 with a standard error of 0.12) and a transmission rate of 0.55 for the subtype B data set (Poisson mean of 0.80 with a standard error of 0.05). These rates of transmission are unreasonably high compared to reported sexual transmission rates, which have a lower bound of ~0.001 (reviewed in Powers et al. [32]). We therefore conclude that the transmission of multiple variants does not represent low probability independent events but rather results from either transiently high transmission rates or linked transmission of multiple virions.

Multiple variant infection and disease progression. Of the 24 individuals monitored for 1 year postinfection, we found no significant difference between viral RNA load set point or $CD4^+$ T-cell counts in participants who were infected with a single variant (19 of 24) compared to those infected with multiple variants (5 of 24) ($P = 0.3198$ and $P = 0.2232$, respectively) (Table 1). However, 4 of 6 (67%) individuals with multiple variant infections had $CD4^+$ T-cell counts consistently below 350 cells/μl and were classified as rapid progressors, whereas only 4 of 20 (20%) individuals that were infected with single variants fell into this category. This association between rapid disease progression and multiple variant transmission ($P = 0.051$) supports previous studies that have shown that high diversity after infection is associated with increased rates of disease progression (10, 12, 37).

DISCUSSION

This study is the first to provide a comprehensive assessment of the multiplicity of HIV-1 subtype C infection in the context of heterosexual transmission in men and women. Our findings mirror observations of subtype B transmissions via heterosexual or homosexual routes using the same methodology (16). Together with the subtype B study, these data demonstrate that in 171 individuals a single virus is responsible for infection of 77% of individuals, with 23% of individuals infected with

FIG. 4. (A and B) Clustering of mutations within stretches of 10 amino acid residues associated with putative CTL pressure illustrated in participant 703010054 (stage V of infection) (A) and participant 705010015 (stage V of infection) (B). (C and D) Changes in the V1/V2 loop in *env* associated with putative antibody pressure illustrated in participant 704010017 (stage VI of infection) displaying clustered mutations within the V2 loop resulting in the gain of three N-linked glycosylation sites (C) and participant CAP269 displaying multiple amino acid insertions within the V1 loop resulting in the gain of one to four N-linked glycosylation sites (D). Highlighter plots (left) compare sequences from each participant's sequence set to an intraparticipant consensus (uppermost sequence) and illustrate the positions of nucleotide base transitions and transversions using short, color-coded bars. Amino acid identity alignments (right) illustrate regions of clustering of mutations in sequences aligned with an intraparticipant subtype C consensus sequence with corresponding HXB2 *env* protein locations indicated above. The number of sequences harboring a particular mutation is displayed alongside each sequence. N-linked glycosylation sites are indicated as red amino acid residues. Where significant mutational clustering was detected by a randomization test, *P* values are provided in parentheses.


A
4165
6000
4164
4168
4166
6001
5995
5993
6067
4167
5996
5991
5994
5992
5990
4170
5997
4169
5999
5998
0.005
5990
p = 4.7e-06
0
1000
2000
3000
Con.CAP37
4165
4166
5995
6001
5993
4164
6067
6000
4168
5996
5994
5992
4167
5991
5999
5990
4169
4170
5997
5998
I = A I = T = G I = C I = Gaps
Base number
B
4183
4182
4184
4185
5479
5483
5480
4186
5478
4187
6069
4190
5484
4188
4189
5486
5481
6068
5482
5485
0.002
4188
p = 9.2e-06
4187
p = 0.00014
4186
p = 3.1e-05
4184
p = 5.15e-06
Con.CAP69
5479
5483
4184
4185
4182
4183
5480
4186
4187
6069
4189
5486
5478
4188
4190
5484
6068
5481
5485
5482
5484
p = 3.0e-06
5481
p = 6.2e-05

multiple variants. Based on the frequency of the transmission of multiple variants, we find that infection with more than one variant does not occur as independent events at low probability. This implies that transmission of the second variant is linked to transmission of the first variant. Understanding the frequency and cause of multivariant transmission is relevant since individuals infected with multiple variants would require a vaccine that protects against greater initial viral diversity instead of a single homogeneous virus population. In addition, it is clinically important since high diversity following transmission has been associated with faster disease progression (10, 12, 37). In keeping with this observation, the present study found an association between multivariant transmission and disease progression.

Discrepant results due to differences in methodological approaches have hindered a clear understanding of multivariant transmission. A key advantage of our study is that we used the same methodological and analytical approaches to define the founder virus population that was used recently to study subtype B acute infection (16), thus enabling us to clearly enumerate the infecting viruses and also directly compare results. Despite different infecting subtypes and routes of transmission, the frequencies of multivariant transmission were strikingly similar: we report 22% in subtype C heterosexually infected men and women compared to the 24% of participants infected with subtype B via homosexual and heterosexual transmission reported by Keele et al. (16). Phylogenetic analysis indicated that the multiple variants came from a single donor in 87% of the cases (13 of 15 subjects), and the time to the MRCA demonstrates that the variants diverged at times significantly before the transmission event.

This estimated frequency of multivariant transmissions should, however, be considered a minimum. Many infections in highly epidemic regions have been attributed to transmissions during the acute stage of infection (30). Since this stage is generally associated with a highly homogeneous viral population, multiple variant transmissions in these instances could be missed. In addition, we may miss variants present at a low frequency (with a sample size of 20 sequenced amplicons, there is 95% confidence of detecting sequences present at frequencies greater than 15%) (16).

Although we used a model which assumes neutral evolution (16), deleterious mutations will be lost through purifying selection, and early innate and adaptive host responses are likely to impact the apparent mutation rate, especially in participants sampled after peak viremia. We did in fact identify putative immune pressure in acute infection, with a third of the sequence sets containing evidence of putative CTL pressure (based on clustered mutations) or antibody pressure (based on changes in N-glycosylation sites or variable loop length). The rates of mutation were also influenced by APOBEC3G-mediated hypermutation observed in eight individuals with single variant infections. In addition, sequences were under purifying selection with a higher rate of synonymous (dS) compared to nonsynonymous (dN) substitutions (mean dN/dS ratio of 0.79; variance, 0.44). A mean dN/dS ratio of <1 suggests that the rate of diversification of the sequences could be slightly less than the rate estimated under a strict assumption of neutrality. However, the impact of a relatively small departure from neutrality on the estimated times to the last common ancestors of intrapatient sequence sets is likely to be minor.

The rate of HIV transmission is in the range of one transmission event per 1,000 exposures (34; reviewed in Powers et al. [32]), although two studies reported rates of 31 per 1,000 exposures (26), (40) and 97 per 1,000 exposures(3). However, even rates as high as 0.03 to 0.1 cannot account for a frequency of multivariant transmission of 22 to 24%, if multiple variants are transmitted independently. This suggests that transmission of each variant is not an independent event in the context of a low transmission probability. One explanation of the frequency of multiple variant transmissions is that different cofactors transiently change the rate of multivariant transmission. The distribution of frequency of one, two, and more than two variants can be explained if two rates are incorporated: one rate would account for 70 to 75% of transmission events and have a low probability of transmission with only rare occurrences of the transmission of multiple variants; the second rate would account for 25 to 30% of transmission events. However, a probability of transmission of $\sim$0.8 would be required to result in equal numbers of transmission events of two variants and more than two variants, which would approximate the observed data. It is likely that increased transmission occurs as a result of sexually transmitted infections (38) or traumatic breaks in the epithelium. However, it is also possible that the transmission of multiple variants represents a linked event, i.e., infection by one particle (or genome) is in some way linked to an increased probability of infection with a second particle (or genome). If the infectious unit were an infected $CD4^+$ T cell, which can be infected with multiple viruses (15), this could account for at least some of the multivariant transmission events. A recent report has shown the potential for infected cells to penetrate a disturbed epithelium (45), and the apparent need for infection via infected cells in the case of HTLV-1 provides further support for a cell-mediated mechanism of transmission (33). Alternatively, virus particles could be aggregated by biological molecules such as SEVI (for semen-derived enhancer of virus infection) (27) or tetherin during budding (28), potentiating infection with two particles in a single, rare transmission event.

A previous study from Kenya showed women were generally

FIG. 5. Highly heterogeneous, multiple variant *env* sequence populations are visually represented by phylogenetic trees with extensive branch structure and discernible clades (left) and Highlighter plots with diverse patterns of nucleotide base mutations compared to the intraparticipant consensus (right). *env* variants resulting from recombination between clades are displayed below with RAP (for recombination analysis program) plots. Parental strains for each recombinant are color coded on the trees, with regions within each recombinant likewise color coded to correspond to respective parental strains. Recombination breakpoints are illustrated by empty boxes on RAP plots. (A) Participant CAP37 was infected with three distinct variants. Sequence 5998 differs by up to 6% from the remaining sequences from this participant and is a suspected dual infection. (B) Participant CAP69 was infected with at least five distinct viruses with extensive recombination between clades. Six recombinant strains are illustrated here.

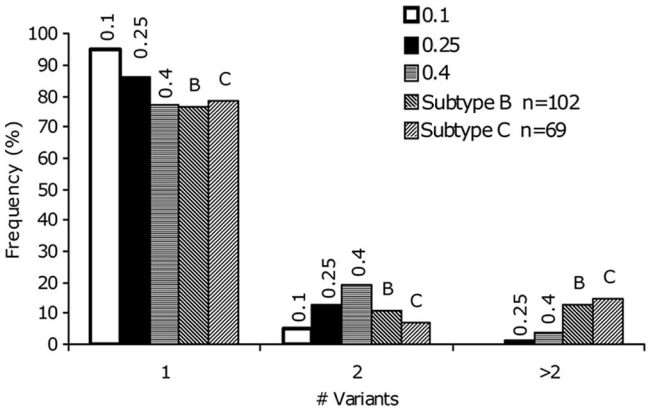


FIG. 6. Model of the rates of transmission of multiple variants using a Poisson distribution. Transmission rates, shown above each bar, are modeled to account for differences in the frequency of transmissions of one, two, or more than two variants using transmission probabilities of 0.1 (Poisson mean, 0.1), 0.25 (Poisson mean, 0.5), and 0.4 (Poisson mean, 0.5). Transmission probability in this setting is defined as the sum of the probabilities of all nonzero events in the Poisson distribution. The frequency of transmission of one, two, or more than two variants is shown for the subtype C cohort described in the present study and from the subtype B cohort described in Keele et al. (16) and indicated by a "C" or "B" above each bar. In modeling the Poisson distribution to fit the cohort data all data were used as counts, not percentages, and values greater than two variants were not pooled but rather modeled in total.

infected with more heterogeneous virus populations compared to men (21). Although in our study, we did not find that females were infected with higher-diversity viral populations compared to men (data not shown), there were differences in frequency of sexually transmitted infections, with genital ulcerative disease being much more common in men from our study than in women (82% versus 13%). Thus, since genital ulcerative disease has an impact on transmission, this could confound our analysis. In addition, uncircumcised men are more susceptible to infection (1, 2); thus, this difference in results between studies may be due to the fact that most of the men in our study were from Malawi where there is a very low frequency of circumcision, whereas the Kenyan study recruited from a cohort of individuals where 87% were circumcised (34).

In conclusion, infection with a single virus in the majority of individuals demonstrates the severity of the genetic bottleneck at transmission. These data in conjunction with the subtype B analysis suggests a universal observation that mucosal HIV-1 infection most frequently originates from a single infectious unit. Less frequently, multiple viral variants are transmitted, which not only increases the genetic diversity, but this increased diversity also provides the virus with greater opportunity to escape early selective pressure through recombination. Although the biological basis for the transmission of multiple variants remains unknown, possible explanations include transiently high rates of transmission due to cofactors, transmission via a multiply infected cell, or transmission of viral aggregates. Since one in five individuals will become infected with multiple infectious variants, it is important to translate how this information impacts on the breadth and targeting needed for protective vaccination.

ACKNOWLEDGMENTS

This study was funded by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, and the U.S. Department of Health and Human Services (AI51794, CAPRISA; DK49381 [M.S.C.], CHAVI), as well as by the National Research Foundation (no. 67385) (South Africa), the South African AIDS Vaccine Initiative, and amFAR grant 106997-43.

We thank the clinical staff and participants from the CAPRISA, CHAVI, and Malawi STI cohorts; Darren Marten for critical comments; and Leslie Arney for assistance with the graphics. We also thank the clinical staff from the CHAVI Lilongwe cohorts, including Francis Martinson, Gift Kamanga, Happiness Kanyamula, and Deborah Kamwendo, for their support.

REFERENCES

1. **Auvert, B., D. Taljaard, E. Lagarde, J. Sobngwi-Tambekou, R. Sitta, and A. Puren.** 2005. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial. PLoS Med. **2:**e298.
2. **Bailey, R. C., S. Moses, C. B. Parker, K. Agot, I. Maclean, J. N. Krieger, C. F. Williams, R. T. Campbell, and J. O. Ndinya-Achola.** 2007. Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. Lancet **369:**643–656.
3. **Cameron, D. W., J. N. Simonsen, L. J. D'Costa, A. R. Ronald, G. M. Maitha, M. N. Gakinya, M. Cheang, J. O. Ndinya-Achola, P. Piot, and R. C. Brunham.** 1989. Female to male transmission of human immunodeficiency virus type 1: risk factors for seroconversion in men. Lancet **ii:**403–407.
4. **Derdeyn, C. A., J. M. Decker, F. Bibollet-Ruche, J. L. Mokili, M. Muldoon, S. A. Denham, M. L. Heil, F. Kasolo, R. Musonda, B. H. Hahn, G. M. Shaw, B. T. Korber, S. Allen, and E. Hunter.** 2004. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. Science **303:**2019–2022.
5. **Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut.** 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. **4:**e88.
6. **Drummond, A. J., and A. Rambaut.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. **7:**214.
7. **Edmonson, P. F., and J. I. Mullins.** 1992. Efficient amplification of HIV half-genomes from tissue DNA. Nucleic Acids Res. **20:**4933.
8. **Fiebig, E. W., D. J. Wright, B. D. Rawal, P. E. Garrett, R. T. Schumacher, L. Peddada, C. Heldebrant, R. Smith, A. Conrad, S. H. Kleinman, and M. P. Busch.** 2003. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. AIDS **17:**1871–1879.
9. **Gottlieb, G. S., L. Heath, D. C. Nickle, K. G. Wong, S. E. Leach, B. Jacobs, S. Gezahegne, A. B. van't Wout, L. P. Jacobson, J. B. Margolick, and J. I. Mullins.** 2008. HIV-1 variation before seroconversion in men who have sex with men: analysis of acute/early HIV infection in the multicenter AIDS cohort study. J. Infect. Dis. **197:**1011–1015.
10. **Gottlieb, G. S., D. C. Nickle, M. A. Jensen, K. G. Wong, J. Grobler, F. Li, S. L. Liu, C. Rademeyer, G. H. Learn, S. S. Karim, C. Williamson, L. Corey, J. B. Margolick, and J. I. Mullins.** 2004. Dual HIV-1 infection associated with rapid disease progression. Lancet **363:**619–622.
11. **Greenier, J. L., C. J. Miller, D. Lu, P. J. Dailey, F. X. Lu, K. J. Kunstman, S. M. Wolinsky, and M. L. Marthas.** 2001. Route of simian immunodeficiency virus inoculation determines the complexity but not the identity of viral variant populations that infect rhesus macaques. J. Virol. **75:**3753–3765.
12. **Grobler, J., C. M. Gray, C. Rademeyer, C. Seoighe, G. Ramjee, S. A. Karim, L. Morris, and C. Williamson.** 2004. Incidence of HIV-1 dual infection and its association with increased viral load set point in a cohort of HIV-1 subtype C-infected female sex workers. J. Infect. Dis. **190:**1355–1359.
13. **Haaland, R. E., P. A. Harkins, J. F. Salazar-Gonzalez, A. Johnson, A. Tichacek, E. Karita, O. Manigart, J. Mulenga, B. F. Keele, G. M. Shaw, B. H. Hahn, S. Allen, C. A. Derdeyn, and E. Hunter.** 2009. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. PloS Pathog. **5:**e1000274.
14. **Jones, N. A., X. Wei, DR. Flower, M. Wong, F. Michor, M. S. Saag, B. H. Hahn, M. A. Nowak, G. M. Shaw, and P. Borrow.** 2004. Determinations of human immunodeficiency virus type 1 escape from the primary $CD8^+$ cytotoxic T lymphocyte response. J. Exp. Med. **200:**1243–1256.
15. **Jung, A., R. Maier, J. P. Vartanian, G. Bocharov, V. Jung, U. Fischer, E. Meese, S. Wain-Hobson, and A. Meyerhans.** 2002. Multiply infected spleen cells in HIV patients. Nature **418:**144.
16. **Keele, B. F., E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham, M. G. Salazar, C. Sun, T. Grayson, S. Wang, H. Li, X. Wei, C. Jiang, J. L. Kirchherr, F. Gao, J. A. Anderson, L. H. Ping, R. Swanstrom, G. D. Tomaras, W. A. Blattner, P. A. Goepfert, J. M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, M. S. Cohen, D. C. Montefiori, B. F. Haynes, B. Gaschen, G. S. Athreya, H. Y. Lee, N. Wood, C. Seoighe, A. S. Perelson, T. Bhattacharya, B. T. Korber, B. H. Hahn, and G. M. Shaw.** 2008. Identification and char-

acterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc. Natl. Acad. Sci. USA **105:**7552–7557.

17. **Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. Frost.** 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. **23:**1891–1901.
18. **Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. Frost.** 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics **22:**3096–3098.
19. **Land, A. M., T. B. Ball, M. Luo, R. Pilon, P. Sandstrom, J. E. Embree, C. Wachihi, J. Kimani, and F. A. Plummer.** 2008. Human immunodeficiency virus (HIV) type 1 proviral hypermutation correlates with CD4 count in HIV-infected women from Kenya. J. Virol. **82:**8172–8182.
20. Reference deleted.
21. **Long, E. M., H. L. Martin, Jr., J. K. Kreiss, S. M. Rainwater, L. Lavreys, D. J. Jackson, J. Rakwar, K. Mandaliya, and J. Overbaugh.** 2000. Gender differences in HIV-1 diversity at time of infection. Nat. Med. **6:**71–75.
22. **Malim, M. H.** 2006. Natural resistance to HIV infection: the Vif-APOBEC interaction. C. R. Biol. **329:**871–875.
23. **Mansky, L. M., and H. M. Temin.** 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J. Virol. **69:**5087–5094.
24. **Markowitz, M., M. Louie, A. Hurley, E. Sun, M. M. Di, A. S. Perelson, and D. D. Ho.** 2003. A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and T-cell decay in vivo. J. Virol. **77:**5037–5038.
25. **Martin, D. P., C. Williamson, and D. Posada.** 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics **21:**260–262.
26. **Mastro, T. D., G. A. Satten, T. Nopkesorn, S. Sangkharomya, and I. M. Longini, Jr.** 1994. Probability of female-to-male transmission of HIV-1 in Thailand. Lancet **343:**204–207.
27. **Munch, J., E. Rucker, L. Standker, K. Adermann, C. Goffinet, M. Schindler, S. Wildum, R. Chinnadurai, D. Rajan, A. Specht, G. Gimenez-Gallego, P. C. Sanchez, D. M. Fowler, A. Koulov, J. W. Kelly, W. Mothes, J. C. Grivel, L. Margolis, O. T. Keppler, W. G. Forssmann, and F. Kirchhoff.** 2007. Semen-derived amyloid fibrils drastically enhance HIV infection. Cell **131:**1059–1071.
28. **Neil, S. J., T. Zang, and P. D. Bieniasz.** 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature **451:**425–430.
29. **Palmer, S., M. Kearney, F. Maldarelli, E. K. Halvas, C. J. Bixby, H. Bazmi, D. Rock, J. Falloon, R. T. Davey, Jr., R. L. Dewar, J. A. Metcalf, S. Hammer, J. W. Mellors, and J. M. Coffin.** 2005. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. J. Clin. Microbiol. **43:**406–413.
30. **Pilcher, C. D., H. C. Tien, J. J. Eron, Jr., P. L. Vernazza, S. Y. Leu, P. W. Stewart, L. E. Goh, and M. S. Cohen.** 2004. Brief but efficient: acute HIV infection and the sexual transmission of HIV. J. Infect. Dis. **189:**1785–1792.
31. **Pond, S. L., S. D. Frost, and S. V. Muse.** 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics **21:**676–679.
32. **Powers, K. A., C. Poole, A. E. Pettifor, and M. S. Cohen.** 2008. Rethinking the heterosexual infectivity of HIV-1: a systematic review and meta-analysis. Lancet Infect. Dis. **8:**553–563.
33. **Proietti, F. A., A. B. Carneiro-Proietti, B. C. Catalan-Soares, and E. L. Murphy.** 2005. Global epidemiology of HTLV-1 infection and associated diseases. Oncogene **24:**6058–6068.
34. **Rakwar, J., L. Lavreys, M. L. Thompson, D. Jackson, J. Bwayo, S. Hassanali, K. Mandaliya, J. Ndinya-Achola, and J. Kreiss.** 1999. Cofactors for the acquisition of HIV-1 among heterosexual men: prospective cohort study of trucking company workers in Kenya. AIDS **13:**607–614.
35. **Ritola, K., C. D. Pilcher, S. A. Fiscus, N. G. Hoffman, J. A. Nelson, K. M. Kitrinos, C. B. Hicks, J. J. Eron, Jr., and R. Swanstrom.** 2004. Multiple V1/V2 env variants are frequently present during primary infection with human immunodeficiency virus type 1. J. Virol. **78:**11208–11218.
36. **Sagar, M., E. Kirkegaard, E. M. Long, C. Celum, S. Buchbinder, E. S. Daar, and J. Overbaugh.** 2004. Human immunodeficiency virus type 1 (HIV-1) diversity at time of infection is not restricted to certain risk groups or specific HIV-1 subtypes. J. Virol. **78:**7279–7283.
37. **Sagar, M., L. Lavreys, J. M. Baeten, B. A. Richardson, K. Mandaliya, B. H. Chohan, J. K. Kreiss, and J. Overbaugh.** 2003. Infection with multiple human immunodeficiency virus type 1 variants is associated with faster disease progression. J. Virol. **77:**12921–12926.
38. **Sagar, M., L. Lavreys, J. M. Baeten, B. A. Richardson, K. Mandaliya, J. O. Ndinya-Achola, J. K. Kreiss, and J. Overbaugh.** 2004. Identification of modifiable factors that affect the genetic diversity of the transmitted HIV-1 population. AIDS **18:**615–619.
39. **Salazar-Gonzalez, J. F., E. Bailes, K. T. Pham, M. G. Salazar, M. B. Guffey, B. F. Keele, C. A. Derdeyn, P. Farmer, E. Hunter, S. Allen, O. Manigart, J. Mulenga, J. A. Anderson, R. Swanstrom, B. F. Haynes, G. S. Athreya, B. T. Korber, P. M. Sharp, G. M. Shaw, and B. H. Hahn.** 2008. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J. Virol. **82:**3952–3970.
40. **Satten, G. A., T. D. Mastro, and I. M. Longini, Jr.** 1994. Modelling the female-to-male per-act HIV transmission probability in an emerging epidemic in Asia. Stat. Med. **13:**2097–2106.
41. **Simmonds, P., P. Balfe, C. A. Ludlam, J. O. Bishop, and A. J. Brown.** 1990. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. J. Virol. **64:**5840–5850.
42. **Stafford, M. A., L. Corey, Y. Cao, E. S. Daar, D. D. Ho, and A. S. Perelson.** 2000. Modeling plasma virus concentration during primary HIV infection. J. Theor. Biol. **203:**285–301.
43. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596–1599.
44. **van Loggernberg, F., K. Mlisana, C. Williamson, S. C. Auld, L. Morris, C. M. Gray, K. Q. Abdool, A. Grobler, N. Barnabas, I. Iriogbe, and S. S. Abdool Karim.** 2008. Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. PLoS ONE **3:**e1954.
45. **Weiler, A. M., Q. Li, L. Duan, M. Kaizu, K. L. Weisgrau, T. C. Friedrich, M. R. Reynolds, A. T. Haase, and E. G. Rakasz.** 2008. Genital ulcers facilitate rapid viral entry and dissemination following intravaginal inoculation with cell-associated simian immunodeficiency virus SIVmac239. J. Virol. **82:**4154–4158.
46. **Wolfs, T. F., G. Zwart, M. Bakker, and J. Goudsmit.** 1992. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. Virology **189:**103–110.
47. **Wolinsky, S. M., C. M. Wike, B. T. Korber, C. Hutto, W. P. Parks, L. L. Rosenblum, K. J. Kunstman, M. R. Furtado, and J. L. Munoz.** 1992. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. Science **255:**1134–1137.
48. **Zhu, T., H. Mo, N. Wang, D. S. Nam, Y. Cao, R. A. Koup, and D. D. Ho.** 1993. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. Science **261:**1179–1181.