Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection

Jesus F. Salazar-Gonzalez,[1] Maria G. Salazar,[1] Brandon F. Keele,[1] Gerald H. Learn,[1] Elena E. Giorgi,[2,3] Hui Li,[1] Julie M. Decker,[1] Shuyi Wang,[1] Joshua Baalwa,[1] Matthias H. Kraus,[1] Nicholas F. Parrish,[1] Katharina S. Shaw,[1] M. Brad Guffey,[1] Katharine J. Bar,[1] Katie L. Davis,[1] Christina Ochsenbauer-Jambor,[1] John C. Kappes,[1] Michael S. Saag,[1] Myron S. Cohen,[4] Joseph Mulenga,[5] Cynthia A. Derdeyn,[6] Susan Allen,[6] Eric Hunter,[6] Martin Markowitz,[7,8] Peter Hraber,[2] Alan S. Perelson,[2] Tanmoy Bhattacharya,[2,9] Barton F. Haynes,[10] Bette T. Korber,[2,9] Beatrice H. Hahn,[1] and George M. Shaw[1]

[1]University of Alabama at Birmingham, Birmingham, AL 35294
[2]Los Alamos National Laboratory, Los Alamos, NM 87545
[3]University of Massachusetts, Amherst, MA 01002
[4]The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
[5]Zambia-Emory HIV Research Project, Lusaka, Zambia
[6]Emory University, Atlanta, GA 30329
[7]Aaron Diamond AIDS Research Center, New York, NY 10016
[8]The Rockefeller University, New York, NY 10065
[9]Santa Fe Institute, Santa Fe, NM 87501
[10]Duke University Medical Center, Durham, NC 27710

Identification of full-length transmitted HIV-1 genomes could be instrumental in HIV-1 pathogenesis, microbicide, and vaccine research by enabling the direct analysis of those viruses actually responsible for productive clinical infection. We show in 12 acutely infected subjects (9 clade B and 3 clade C) that complete HIV-1 genomes of transmitted/founder viruses can be inferred by single genome amplification and sequencing of plasma virion RNA. This allowed for the molecular cloning and biological analysis of transmitted/founder viruses and a comprehensive genome-wide assessment of the genetic imprint left on the evolving virus quasispecies by a composite of host selection pressures. Transmitted viruses encoded intact canonical genes (*gag-pol-vif-vpr-tat-rev-vpu-env-nef*) and replicated efficiently in primary human CD4+ T lymphocytes but much less so in monocyte-derived macrophages. Transmitted viruses were CD4 and CCR5 tropic and demonstrated concealment of coreceptor binding surfaces of the envelope bridging sheet and variable loop 3. 2 mo after infection, transmitted/founder viruses in three subjects were nearly completely replaced by viruses differing at two to five highly selected genomic loci; by 12–20 mo, viruses exhibited concentrated mutations at 17–34 discrete locations. These findings reveal viral properties associated with mucosal HIV-1 transmission and a limited set of rapidly evolving adaptive mutations driven primarily, but not exclusively, by early cytotoxic T cell responses.

CORRESPONDENCE
George M. Shaw:
gshaw@uab.edu

Abbreviations used: C.I., confidence interval; HD, Hamming distance; LTR, long terminal repeat; MRCA, most recent common ancestor; p_n, nonsynonymous divergence; p_s, synonymous divergence; sCD4, soluble CD4; SGA, single genome amplification; SIV, simian immunodeficiency virus.

Transmission of HIV-1 generally results from virus exposure at mucosal surfaces followed by virus replication in submucosal and locoregional lymphoid tissues, and subsequently, by overt systemic infection (1–7). Because of the inaccessibility of early sites of replication, the molecular details of HIV-1 transmission and early virus evolution remain largely unknown.

J.F. Salazar-Gonzalez, M.G. Salazar, and B.F. Keele contributed equally to this paper.

© 2009 Salazar-Gonzalez et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see http://www.jem.org/misc/terms.shtml). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 3.0 Unported license, as described at http://creativecommons.org/licenses/by-nc-sa/3.0/).

The Rockefeller University Press $30.00
J. Exp. Med. Vol. 206 No. 6 1273-1289
www.jem.org/cgi/doi/10.1084/jem.20090378

The Journal of Experimental Medicine

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

Analysis of individual HIV-1 genes from viruses that are responsible for productive clinical infection in humans can be instrumental in elucidating viral properties and biological events underlying the transmission process (8–12). However, identification and characterization of full-length genomes from such viruses could go much further in elucidating, on a genome-wide basis, those properties of transmitted/founder viruses and their progeny that are essential for virus transmission and the establishment of viral persistence. This is true for naive individuals who become infected by HIV-1 and for subjects who are immunized with candidate HIV-1 vaccines but experience breakthrough infection (http://www.hvtn.org/science/1107.html).

Recently, we developed a mathematical model of HIV-1 sequence evolution in acute infection and an experimental strategy based on single genome amplification (SGA) of plasma vRNA/cDNA, followed by direct sequencing of uncloned DNA amplicons, that allowed us to infer the nucleotide sequences of full-length envelope (*env*) genes of transmitted/founder viruses in 98 out of 102 consecutively studied patients (11). The model assumes that a transmitted virus replicates exponentially with a generation time of 2 d (13), a reproductive ratio (R_0) of 6 (14), and an RT error rate of 2.16×10^{-5} (15); that it diversifies under no selection; and that it exhibits a constant mutation rate across positions and lineages and undergoes no back mutations. These assumptions were based on estimated parameters of virus replication and a timing of virus sampling before the development of detectable immune responses (11, 16–20). The model predicts a Poisson distribution of mutations and a star-like phylogeny that coalesces to an inferred consensus sequence at or near the time of virus transmission. We obtained direct experimental evidence in support of this model in Indian rhesus macaques mucosally infected by simian immunodeficiency virus (SIV) strain $SIV_{mac}251$ (21) and in humans infected by HIV-1 from a known sexual partner (unpublished data) (9). In each case, an *env* sequence in the SIV inoculum stock or in the blood of a chronically infected sexual partner was found to be identical to that of the transmitted/founder virus identified in the recipient. We also tested the model by Monte Carlo simulation and against an empirical dataset of 3,449 SGA-derived complete *env* sequences (11). The results supported the model and its assumptions. Importantly, the model and the empirical findings allowed us to infer that in ~70–80% of the cases of sexual transmission of HIV-1, a single virus (or infected cell) is responsible for establishing productive clinical infection, a conclusion now supported by studies in seven additional patient cohorts infected by HIV-1 subtypes A, B, C, or D (unpublished data) (8–10, 12). In the present study, we asked if the experimental strategy for identifying transmitted/founder *env* sequences can be applied successfully to full-length HIV-1 vRNA genomes, which are nearly four times longer than *env* genes (9 vs. 2.6 kb), and whether identification of such genomes can provide new insight into the biology of HIV-1 transmission, and the kinetics and pathways of virus diversification and adaptation leading to viral persistence.

RESULTS

Study subjects

Plasma specimens from 12 adult subjects (10 male and 2 female) with acute HIV-1 infection were analyzed in this study (Table I). Nine subjects were infected by HIV-1 subtype B and three were infected by subtype C. At the initial sampling time point, 10 subjects were plasma $vRNA^+/Ab^-$ (Fiebig stage II; the HIV-1 clinical staging system is discussed in references 11, 17), and two subjects were $vRNA^+/ELISA^+/WB$ indeterminant (Fiebig stage IV). Three subjects were sampled longitudinally through as many as 85 wk of follow-up. Peak plasma viral loads ranged from 394,649 to 26,700,000 vRNA copies per ml. Four subjects admitted to heterosexual exposure as their only HIV-1 risk factor, and eight were men who had sex with men. No subject admitted to injecting drug use.

SGA and sequencing

Between 5 and 18 complete viral genomes (median = 9) were derived by amplification of individual plasma vRNA/cDNA molecules from each subject (108 amplicons in total; Table II). Each of the 108 amplicons was sequenced directly without interim cloning. Sequence chromatograms of 62 amplicons were unambiguous at every position. Sequence chromatograms of 46 amplicons had mixed bases at one to five positions per sequence. Because the proportion of PCR-positive wells at endpoint cDNA dilution was <20%, and because mixed bases generally represented only a subset of polymorphisms in any one sequence, we could infer that most mixed bases on chromatograms resulted from *Taq* polymerase errors in the initial PCR cycles and not from amplification from more than one original vRNA/cDNA template; in such cases, a correct assignment of the ambiguous base could be made. In five instances where one or more mixed bases represented the only polymorphisms in a sequence, this was not possible. Thus, we could make an unambiguous assignment of nucleotides at each position in the nucleotide sequences of 103 HIV-1 genomes and at all but nine positions in five others. From three subjects (CH40, CH58, and CH77), an additional 209 overlapping half genomes and 177 shorter sequence fragments were determined from time points beginning before first antibody detection (Fiebig stage II) and extending to 350–592 d later (Fiebig stage VI).

HIV-1 diversity

In a maximum likelihood phylogenetic tree, viral sequences from the nine US subjects clustered significantly with prototype B clade viruses, whereas sequences from the three Zambian subjects clustered with prototype C clade viruses (Fig. 1). Maximum interstrain diversity among all 108 full-length genomes was >25%, reflecting differences typically observed between different clade B and C viruses. Within individual subjects, maximum virus diversity was far less, ranging from 0.04% in subject SUMA0874 to 2.46% in subject ZM247F (Table II). There was no interspersion of sequences among study subjects. Maximum within-patient viral diversity was distinctly lower in 11 subjects (<0.14% in each) compared

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

with the 12th subject, ZM247F (2.46%). We postulated that the observed differences in maximum viral diversity observed within individuals might reflect the numbers of viruses responsible for establishing productive infection in these subjects, as shown previously for *env* diversity (11). We formally tested this hypothesis by comparing observed viral genome diversities in each subject with estimates, based on model predictions, of the maximum diversity one could expect within 100 d after transmission of a single virus (0.60%; 0.54–0.68% confidence interval [C.I.]) (11). 11 out of the 12 subjects had sequences that fell well below the 0.6% threshold, whereas 1 subject (ZM247F) had sequences that fell far above it (Table II). We also used the model to estimate in each subject the minimum number of days that would be required to explain the observed within-patient HIV-1 genome diversification from a single most recent common ancestor (MRCA) sequence, as we had done previously for *env* diversification (11). In this analysis, we did not adjust for mutations that are selected against and go unobserved because they result in unfit viruses; as a consequence, the timing estimates based on a comparison of the observed data to the model tend to be biased toward a low estimate. 11 subjects with lower viral diversity had minimum estimates for days since a MRCA virus that fell well within model predictions for infection by a single virus (11–33 d; 95% C.I. = 7–38 d) and within a time frame consistent with each subject's Fiebig clinical stage (Table II). Conversely, sequence diversity in subject ZM247F corresponded to a minimum estimate for a MRCA of 493 d, far beyond the range of plausibility for recent infection based on this subject's Fiebig stage II, which has an average duration from virus transmission of 22 d (95% C.I. = 16–39 d) (11, 17). Interestingly, sequences from ZM247F fell into two distinct low diversity phylogenetic lineages that differed from each other by an average of 2.4% (Fig. 1). Sequence diversity within lineage 1 ranged from 0.01–0.09%, and within lineage 2 from 0.02–0.08%. Within each lineage, sequences exhibited a star-like phylogeny and a Poisson distribution of mutations. Model estimates for time from a MRCA for lineage 1 was 21 d (95% C.I. = 14–28 d), and for lineage 2 was 21 d (95% C.I. = 14–28 d). Based on this analysis, we concluded that ZM247F had most likely been infected by two viruses at the same time and from the same sexual partner (Fig. 1 and Table II).

Model testing and analysis of HIV-1 evolution

To explore how full-length HIV-1 genomic sequences sampled near peak viremia conform to model predictions, we obtained for each subject the frequency distribution of all intersequence Hamming distances (HDs; defined as the number of base positions at which two genomes differ) and determined whether it deviated from a Poisson model by using a χ^2 goodness of fit test. Four sequences from three subjects exhibited G-to-A hypermutation. Once these were eliminated from the analysis, sequences from 11 out of 12 subjects conformed to the Poisson model (Table II). Sequences from subject ZM247F did not conform to a Poisson model of variation but did so once the two low diversity lineages evident in the phylogenetic tree (Fig. 1) were analyzed individually. We next investigated whether or not sequences evolved under a star-phylogeny model (i.e., all observed sequences coalesce at the founder) in the expected time frame based on Fiebig stage. Sequences from 6 out of 12 subjects, including each lineage in subject ZM247F, exhibited a star phylogeny (Table II). Among the samples that deviated from a star phylogeny were sequences from two subjects (CH40 and CH77) that exhibited early changes in potential or confirmed CTL epitopes corresponding to the HLA types of these subjects (see Goonetilleke

Table I. Subject demographics, risk group, and baseline laboratory data

Subject	HIV-1 subtype	Geographic location	Gender	Risk group	Sampling date	VL	Clinical specimen	EIA	WB	Fiebig stage[a]
WIT04160	B	Alabama	M	Heterosexual	08.04.00	325,064	Plasma	neg	neg	II
SUMA0874	B	Alabama	M	MSM	05.13.91	939,260	Plasma	neg	neg	II
WEAU0575	B	Alabama	M	MSM	05.30.90	216,415	Plasma	neg	neg	II
TRJ04551	B	Alabama	M	MSM	10.10.01	8,121,951	Plasma	neg	neg	II
04013226-2	B	New York	M	MSM	11.20.02	26,700,000	Plasma	neg	neg	II
04013396-0	B	New York	M	MSM	08.16.05	1,600,000	Plasma	pos	ind	IV
CH40 (700010040)	B	North Carolina	M	MSM	07.11.06	2,197,248	Serum	neg	neg	II
CH40 (700010040)					07.27.06	298,026	Plasma	pos	pos ($p31^-$)	V
CH58 (700010058)	B	North Carolina	M	MSM	08.22.06	92,581	Serum	neg	neg	II
CH58 (700010058)					08.31.06	394,649	Plasma	pos	neg	III
CH77 (700010077)	B	North Carolina	M	MSM	08.25.06	3,565,728	Serum	neg	neg	II
CH77 (700010077)					09.08.06	144,145	Plasma	pos	pos ($p31^-$)	V
ZM246F	C	Zambia	F	Heterosexual	01.14.03	10,013,800	Plasma	neg	neg	II
ZM249M	C	Zambia	M	Heterosexual	08.05.03	>2,000,000	Plasma	pos	ind	IV
ZM247F	C	Zambia	F	Heterosexual	10.28.03	10,823,500	Plasma	neg	neg	II

EIA, enzyme-linked immunosorbant assay for HIV-1; ind, indeterminant; MSM, men who have sex with mena; VL, viral load (RNA molecules/ml); WB, Western blot for HIV-1.
[a]Fiebig clinical serological staging system (17).

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

et al. [22] on p. 1253 of this issue). When putative CTL escape mutations and rare shared polymorphisms arising from early stochastic nucleotide substitutions (11) were excluded from the analysis, sequences from the remaining six subjects conformed to a star phylogeny.

We next analyzed the diversity among *env* genes within full-length genomes compared with a much larger number of *env* genes amplified as *env*-only sequences from the same subjects (Table S1). Because the latter sequences were amplified using different primers than were used for complete genomes, the first question posed was whether both sets of sequences coalesced phylogenetically to the same *env* (or e*nv*s) within each subject, as would be expected for sequences emanating from the same transmitted/founder viruses. This was the case in each of the 12 subjects. The second question posed was, given that there were fewer *env* sequences derived from full-length genomes available for analysis (an average of 11 per subject; median = 8) compared with *env*-only sequences (an average of 36 per subject; median = 34), were the estimated days since a MRCA comparable between the two datasets? Again, the answer was affirmative, with C.I.s for the MRCAs overlapping in every subject except one (Table S1).

For each subject, we examined *gag*, *pol*, *env*, and *nef* genes individually for levels of genetic diversity (Table S1). Within individual subjects, there was considerable heterogeneity among genes in their levels of variation, a finding largely attributable to relatively short gene lengths, small numbers of sequences examined, and the overall low degree of viral diversity present in acute infection samples. Three individuals lacked any variability in *gag* or *nef* sequences. The maximum pairwise HD was greatest in *pol* (6 in subject CH77) and *env* (6 in subjects 04013226-2 and CH77). Estimates of days since the MRCA varied among genes, with *env* and *nef* generally agreeing well with the estimates based on greater numbers of

Table II. Diversity analysis of full-length HIV-1 genomes derived from patients with primary infection

Subject	Fiebig stage	Total number of HIV-1 genomes	Maximum nt length of viral genome analyzed	Minimum nt length of viral genome analyzed	Number of genomes analyzed[a]	Nucleotide sequence diversity (mean %)	Nucleotide sequence diversity (range %)	Max HD	Poisson estimated days since MRCA[b]	Lambda[c]	Goodness of fit p-value	HD fit to Poisson	Star phylogeny	Deviation from star phylogeny	Number of transmitted viruses: env-only analysis[d]	Number of transmitted viruses: Full-length genome analysis
WITO4160	II	18	9,027	7,698	14	0.04	0–0.09	8	17 (12, 22)	3.57	0.923	Yes	Yes		1	1
SUMA0874	II	6	9,033	9,020	6	0.03	0.01–0.04	4	11 (7, 15)	2.33	0.546	Yes	Yes		1	1
WEAU0575	II	9	9,028	9,022	9	0.03	0–0.08	7	14 (8, 21)	3.06	0.49	Yes	No	Early stochastic	1	1
TRJO4551	II	7	9,051	9,046	7	0.06	0.03–0.07	7	20 (14, 26)	4.29	0.836	Yes	Yes		1	1
04013226-2	II	9	9,068	9,038	8	0.08	0.04–0.11	10	33 (28, 38)	7.04	0.269	Yes	No	Early stochastic	1	1
04013396-0	IV	5	9,049	9,048	4	0.03	0–0.06	5	12 (7, 17)	2.5	0.704	Yes	Yes		1	1
CH40 (700010040)	II	26[e]	3,593	2,977	26	0.04	0–0.11	4	16 (12, 21)	1.372	0.251	Yes	Yes		1	1
CH40 (700010040)	V	9	8,990	8,989	9	0.05	0.01–0.1	9	21 (14, 28)	4.44	0.46	Yes	No	CTL selection	1	1
CH58 (700010058)	III	7	9,027	8,717	7	0.06	0.02–0.1	9	24 (14, 33)	5.05	0.941	Yes	No	Early stochastic	1	1
CH77 (700010077)	II	23[e]	2,785	2,785	23	0.05	0–0.14	4	20 (13,26)	1.3	0.449	Yes	Yes		1	1
CH77 (700010077)	V	8	9,048	9,037	8	0.06	0.02–0.13	12	27 (16, 37)	5.75	0.46	Yes	No	CTL selection	1	1
ZM246F	II	10	8,978	8,977	10	0.04	0–0.09	8	17 (10, 24)	3.56	0.88	Yes	No	Early stochastic	1	1
ZM249M	IV	7	9,054	8,735	7	0.03	0.01–0.06	5	15 (9, 20)	3.1	0.671	Yes	Yes		1	1
ZM247F	II	13	9,040	7,957	12	1.19	0.01–2.46	220	493 (314, 673)	106.197	0	No	No	Two strains	2	2
ZM247F lng 1[f]	II	9	9,040	7,957	8	0.05	0.01–0.09	8	21 (14, 28)	4.5	0.967	Yes	Yes		1	1
ZM247F lng 2[f]	II	4	9,034	9,033	4	0.05	0.02–0.08	7	21 (14, 28)	4.5	0.939	Yes	Yes		1	1

[a]Genomes with multiple APOBEC G-to-A hypermutations or large deletions were excluded from model analysis.
[b]Predicted minimum number of days needed to achieve observed within-patient diversity.
[c]Mean of best fitting Poisson distribution. Lambda is the free parameter that defines a Poisson.
[d]Env-only analyses were previously reported (unpublished data) (11, 12).
[e]SGA sequences representing 3′ half subgenomic fragments from the screening sample.
[f]lng 1= lineage 1; lng 2= lineage 2.

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

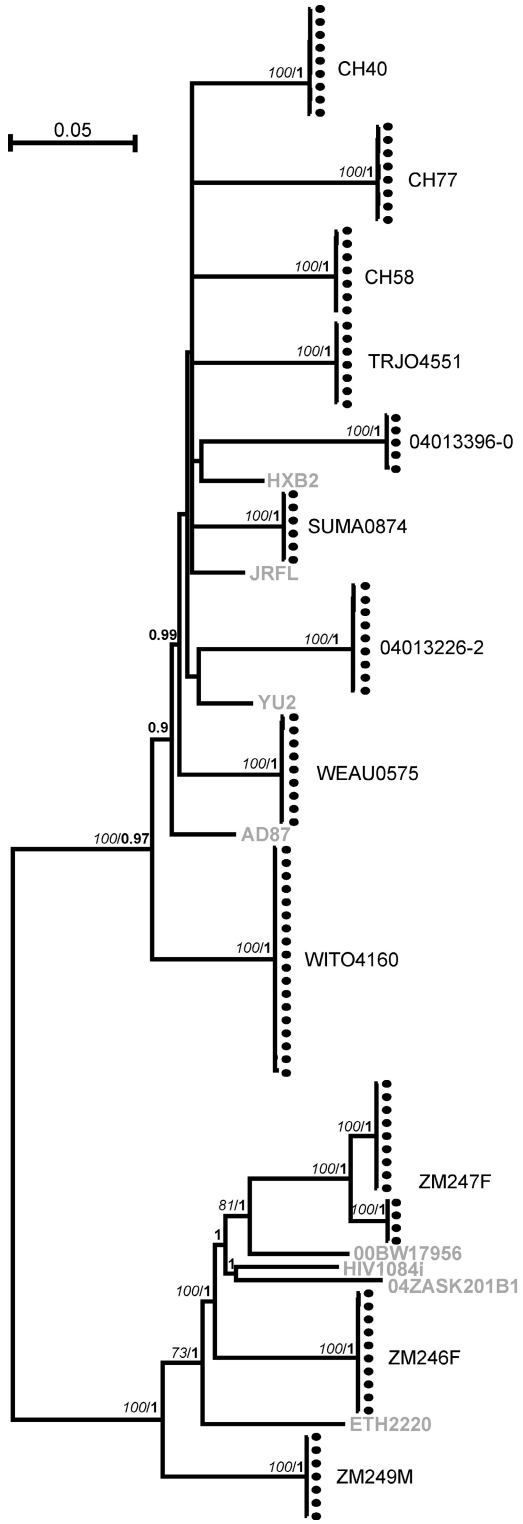

Figure 1. Maximum likelihood phylogenetic tree of full-length HIV-1 genome sequences from 12 subjects. Individual sequences are shown as filled ovals. HIV-1 clade B and C reference sequences from the LANL database (http://www.hiv.lanl.gov/content/index) are shown in gray. Numerals at nodes indicate phylogenetic support for sequences included at that node; maximum likelihood bootstrap support ≥ 70% is shown in italics. Bayesian posterior probability values ≥ 0.9 are shown in bold. The scale bar indicates 5% genetic distance (5 nt differences per 100 sites).

env-only sequences. Contrary to an earlier study (23), there was no evidence of greater variability in *gag* compared with *env*. Table S2 shows the total number of synonymous and nonsynonymous differences among all sequences for each subject and for each gene. There was an excess of nonsynonymous substitutions. However, when taking into account that most random replacements in a coding sequence are nonsynonymous (~79%), the rate of synonymous divergence (p_s) tended to exceed that of the nonsynonymous divergence (p_n) for each gene suggestive of purifying selection. This difference did not reach statistical significance in the present study but did reach significance in a previous one where far greater numbers of sequences were available for analysis (11, 24).

Identification of transmitted or early founder viral genomes

Fig. 2 illustrates the phylogeny of full-length viral sequences from subject WITO4160 together with *Highlighter* plots depicting the positions and identities of nucleotide polymorphisms, insertions, and deletions across the genomes. Among the 18 WITO4160 sequences, no two were identical. Seven sequences contained a total of 15 mixed bases due primarily, if not exclusively, to *Taq* polymerase errors in the initial one or two cycles of PCR amplification. This experimental result (15 *Taq* errors in 18 genomes × 9,000 bp per genome × 3 DNA strands synthesized in the initial two complete cycles of *Taq* polymerization = 3×10^{-5}) is consistent with the error rate reported for *Taq* polymerase of $2.7–8.5 \times 10^{-5}$ (25). Excluding mixed bases (which are denoted by an International Union of Pure and Applied Chemistry designation), single nucleotide insertions (C1, C10, B3, and C9), large deletions (G7 and C4), and G-to-A hypermutation (H3 and C3), each of the sequences differed from the others by 0–11 nt. Nucleotide substitutions exhibited a Poisson distribution and star-like phylogeny (Table II). The consensus of the sequences (WITO_fl.CON) was the same whether or not sequences containing double peaks were included in the analysis; this is an expected result because *Taq* polymerase errors, like HIV-1 RT errors, are essentially randomly distributed across the genomes. Out of the 18 genomes, 9 had intact open reading frames for all essential viral genes, whereas 9 others contained stop codons or insertions or deletions ranging in length from 1 to 1,329 nt. Fig. 3 illustrates in a second subject, ZM247F, many of the same features of early virus diversification, but from two transmitted/founder viruses rather than one. Each transmitted/founder virus in ZM247F was represented by a distinct low diversity lineage evident in both the phylogenetic tree and the *Highlighter* plot. Among all 13 sequences, no two were identical. 8 out of 13 sequences contained mixed bases at one or two positions. Nine sequences contained all essential open reading frames intact. 3 out of 13 sequences contained deletions of between 1 and 1,083 nt; one additional sequence contained a nucleotide substitution resulting

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

in a translational stop codon. Neither APOBEC-related G-to-A hypermutation nor viral recombination between lineages was observed in these very early ZM247F sequences. Sequences comprising each viral lineage (with or without mixed bases included) coalesced to unique, unambiguous transmitted/founder genomes that differed by 2.36%.

Sequences from each of the two viral lineages in ZM247F and from the single lineages in WITO4160, SUMA0874, TRJO4551, 04013396-0, and ZM249M exhibited Poisson distributions of mutations and star-like phylogenies (Figs. 2 and 3; Fig. S1; and Table II), thus allowing for a definitive identification of the transmitted/founder viruses in each subject. Sequences from six other subjects, however, exhibited shared polymorphisms, which can confound the identification of transmitted/founder sequences (11). Examples of shared polymorphisms are illustrated in Fig. 4 for subjects CH40 and CH77, each of whom had just entered Fiebig stage V (vRNA$^+$/ELISA$^+$/WB$^+$[P31$^-$]) at the time of sampling of these sequences. Three shared polymorphisms were evident in sequences of subject CH40 (Fig. 4 A) and five were evident in subject CH77 (Fig. 4 B). Shared polymorphisms can result if two very closely related viruses are acquired during the transmission event, most commonly from a donor who himself/herself is acutely infected (11), or they can arise as a consequence of RT errors in early virus replication cycles of a single transmitted virus and persist alongside the transmitted viral lineage (11). A third possibility is that one or more of the many mutations that occur with sequential replication cycles provides a selective advantage to the virus that results in rapid preferential accumulation of its progeny; this is a particularly likely scenario for CH40 and CH77 at Fiebig stage V, where in a separate study we identified early virus-specific CTL responses (22). In subjects CH40 (Fig. 4 A), CH77 (Fig. 4 B), CH58, ZM246F, 04013226-2, and WEAU0575 (Fig. S2), shared polymorphisms initially precluded a definitive identification of transmitted/founder virus sequences, but this uncertainty could be resolved by analysis of viral sequences from earlier or later time points. This is because a clear directionality in shifting proportions of shared polymorphisms could be established. In subject CH40, for example, where three shared polymorphisms at position 6,705 (G) in *env* and positions 9,360 (A) and 9,371 (A) in *nef* (Figs. 4 A and 5 A; and Fig. S3) were evident in the Fiebig stage V sample, analysis of sequences obtained from a still earlier plasma specimen when the subject was at Fiebig stage II revealed no polymorphisms in *nef*. 4, 12, 24, and 60 wk after enrollment, 63 out of 63 (100%) sequences carried mutations at the two polymorphic sites in *nef*, or within a 30-nt sequence spanning them (Fig. 5 A and Fig. S3). Thus, it was obvious from this analysis that the transmitted/founder sequence at these polymorphic *nef* positions was represented by sequence CH40_fl.CON (Fig. 4 A). We interrogated the shared polymorphism in *env* in a similar manner. In this case, however, 0 out of 26 sequences from the earliest Fiebig stage II sample, 6 out of 23 (26%) sequences

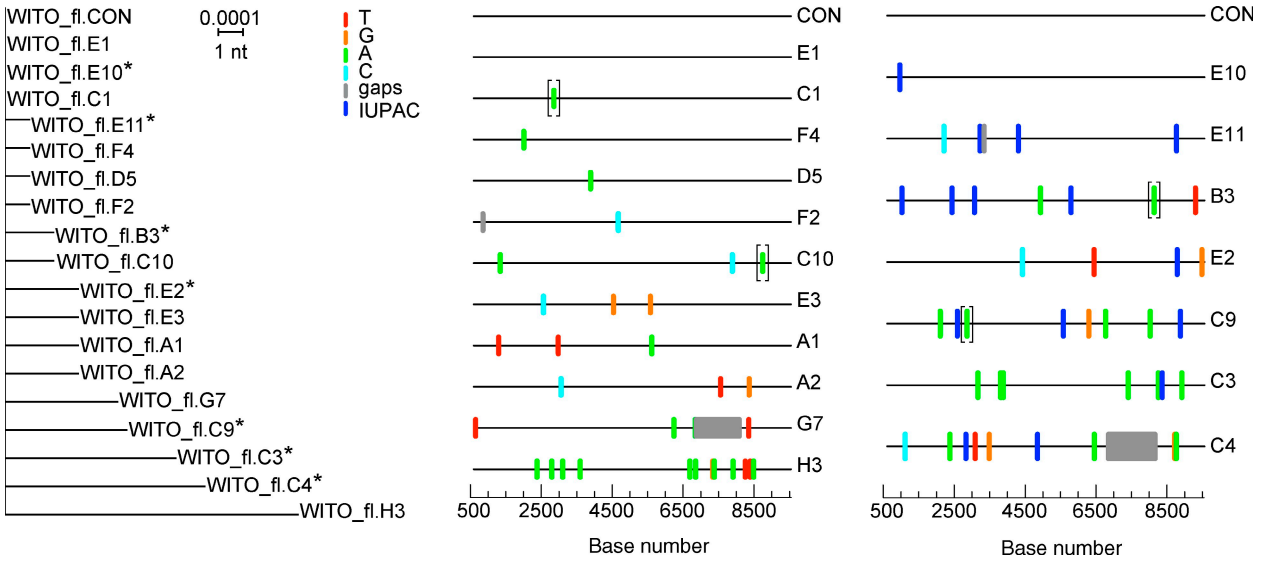

Figure 2. Phylogenetic and *Highlighter* analyses of HIV-1 sequences from subject WITO4160. WITO_fl.CON is the inferred transmitted/founder sequence. In the phylogram (left), asterisks indicate sequences that had International Union of Pure and Applied Chemistry ambiguous state assignments caused by mixed bases. Mixed bases were attributed to *Taq* polymerase errors and, thus, are not reflected in branch lengths of phylograms in this or other figures. The scale bar indicates a genetic distance of 0.0001 (0.01%, or 1 nt difference per genome). In the *Highlighter* plots (middle and right), nucleotide differences from the transmitted/founder consensus sequence are indicated by tic marks. Brackets denote insertions of single bases. Gray tics and boxes indicate a deleted sequence. The consensus (CON) is the same whether or not mixed bases are included. The middle *Highlighter* plot shows sequences without ambiguous (mixed) base assignments. The *Highlighter* plot on the right includes sequences with one or more ambiguous bases (shown as dark blue tic marks). Sequences H3 and C3 contained APOBEC-related G-to-A hypermutation. The horizontal axis indicates nucleotide positions in the HIV-1 genome. Sequences begin at nucleotide position 582 in the 5′ LTR (U5) and extend to position 9,606 in the 3′ LTR (R) based on HXB2 reference sequence numbering (http://www.hiv.lanl.gov/content/sequence/HIV/REVIEWS/HXB2.html).

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

from the Fiebig V sample, and 0 out of 63 sequences from samples 4, 12, 24, or 60 wk later contained the shared polymorphism (Figs. 5 A and S3). Thus, this shared nucleotide polymorphism represented an early stochastic mutation that in all likelihood occurred shortly after virus transmission as predicted in the model (11), was represented as a minor population in the earliest samples, exhibited no fitness advantage, and did not accumulate; instead, it was lost as a minor variant in the expanding quasispecies. Thus, in subject CH40, we could conclude definitively that the transmitted/founder virus genome was represented by the CH40_fl.CON sequence.

Five shared polymorphisms in subject CH77 could be similarly deconvoluted and the transmitted/founder virus identified (Fig. 4 B). The initial sample that we analyzed from this subject was also obtained at early Fiebig stage V, similar to the enrollment sample of subject CH40. At this time point at position 7,285 in *env*, 49 out of 66 (74%) CH77 sequences shared a common nucleotide (A), whereas 17 others shared a different nucleotide (16 sequences with G and 1 sequence with C; Fig. S4). 14 d earlier when the patient was at Fiebig stage II, 23 out of 23 sequences were identical in having a G at this position (Fig. S4). 3–24 wk after enrollment, 29 out of 29 (100%) sequences contained nucleotide mutations at this position or within a 30-nt span that encompassed it. Thus, we could conclude that the transmitted/founder sequence at this position was represented by CH77E_fl.A2 and not by what was the consensus sequence at the enrollment time point (Fig. 4 B). A second set of shared nucleotide polymorphisms was evident in *tat* sequences (position 6,021) where two full-length genomes (C7 and C3) shared a common (T) polymorphism. We could resolve which sequence represented the transmitted/founder virus by determining viral sequences 14 d before and 3–24 wk after this time point. The results revealed that 23 out of 23 (100%) earlier Fiebig II sequences were identical at position 6,021 (containing a T residue) and in a 30-nt region spanning it, and that they were different from 90 out of 92 (98%) sequences from the subsequent time points in this same region (Fig. S4). Thus, at position 6,021, sequences CH77E_fl.C7 and CH77E_fl.C3 corresponded to the transmitted/founder virus. The three other shared polymorphisms in *pol* at positions 2,625, 4,021, and 4,104 represented uncommon sequences in the Fiebig stage V sample that were not prominent in earlier or later samples (Fig. S5 and not depicted). Thus, they represented stochastic mutations that occurred sometime shortly after transmission and persisted at a detectable frequency only transiently. From these analyses, we could infer that none of the sequences shown in Fig. 4 B corresponded to the transmitted/founder genome in subject CH77. Instead, the transmitted/founder full-length genome in subject CH77 corresponded to a sequence identical to the CH77E_fl.A2 sequence except for a single C to T substitution at position 6,021. Full-length genomes from subject CH58 at Fiebig stage III contained two out of seven sequences with a single polymorphism (C) at position 4,408 in *pol* (Fig. S2 B). By examining 12 sequences from Fiebig stage II 9 d earlier and 23 sequences from 45–350 d later (Fig. S6), we could establish definitively the transmitted/founder sequence. The full-length transmitted/founder virus genomes for CH40, CH77, and CH58 represented in Fig. 5 reflect these phylogenetic inferences. For four other subjects whose sequences exhibited rare shared polymorphisms and where very early (Fiebig stage II) or sequential samples were available for analysis, all showed evidence that their consensus sequences indicated in Fig. S2 corresponded to transmitted/founder viruses. Thus, in 12 out of 12 study subjects, we could identify with a high level of confidence the transmitted/founder viral genomes.

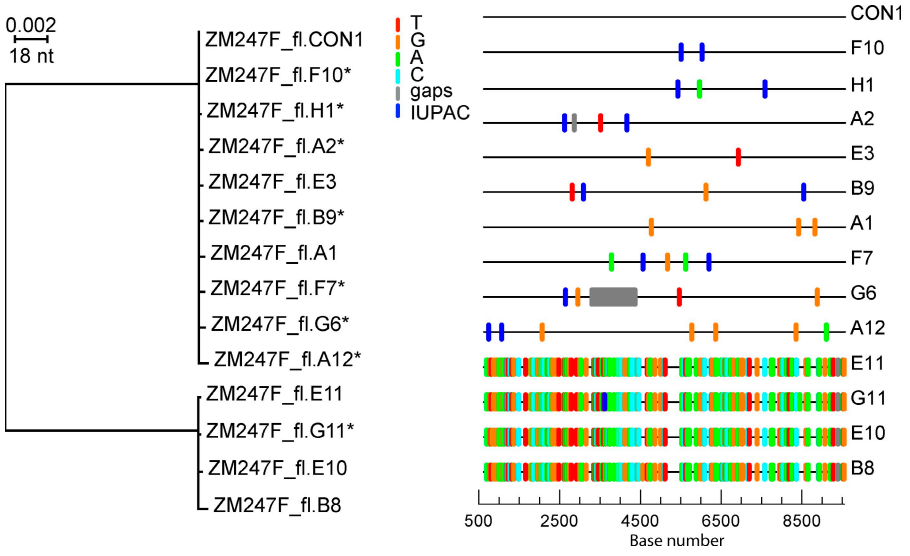

Figure 3. Phylogenetic and *Highlighter* analyses of HIV-1 sequences from subject ZM247F. Subject ZM247F had sequences comprising two distinct lineages that were 2.4% different from each other, as shown by the phylogram (left) and *Highlighter* plot (right). ZM247F_fl.CON1 is the consensus of lineage 1. The second lineage (sequences B8, E10, E11, and G11) differs from ZM247F_fl.CON1 by 213–217 nt differences, as indicated by tic marks.

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

Genetic and biological analysis of transmitted/founder viruses

If early viral sequences coalesce to *actual* transmitted/founder viruses responsible for productive clinical infection, then we would expect three predictions to be borne out experimentally: (a) transmitted/founder *env* sequences inferred from full-length viral genome analysis must be identical to transmitted/founder sequences inferred from *env*-only (subgenomic) SGA analysis; (b) all essential viral gene open reading frames in transmitted/founder full-length genomes must be intact; and (c) inferred transmitted/founder full-length genome sequences must encode replication-competent viruses. We tested and affirmed all three predictions: first, consensus transmitted/founder *env* genes derived by subgenomic amplifications of *env* and by full genome amplifications were identical in each of the 12 subjects (Table II). Second, transmitted/founder complete genomes from each of the 12 subjects contained intact *gag*, *pol* (*rt*, *pro*, and *int*), *vif*, *vpr*, *tat*, *rev*, *vpu*, *env*, and *nef* open reading frames. Third, three complete HIV-1 clade C proviral genomes corresponding to transmitted/founder viruses for subjects ZM246F and ZM247F were constructed either by chemical synthesis or by PCR amplification of viral nucleic acid followed by cloning into plasmid expression vectors (Fig. 6 A). All three genomes (pZM246F-10, pZM247Fv1, and pZM247Fv2) yielded replication-competent virus after transfection and expression in human 293T cells. Each of the three virus strains replicated in activated primary human $CD4^+$ T lymphocytes with kinetics and yields comparable to six control viruses (YU2, SG3, NL4.3, BaL, ADA, and JRCSF) (Fig. 6 B). Surprisingly, each of the three transmitted/founder viruses failed to replicate efficiently when passaged onto human monocyte-derived macrophages obtained from the same normal donor, whereas three prototypic primary macrophage-tropic control viruses (YU2, BaL, and ADA) used as positive controls replicated efficiently in both cell types (Fig. 6 B). This experiment was repeated four times using four different normal uninfected blood donors as the source of lymphocytes and monocyte-derived macrophages and using virus initially generated either in 293T cells

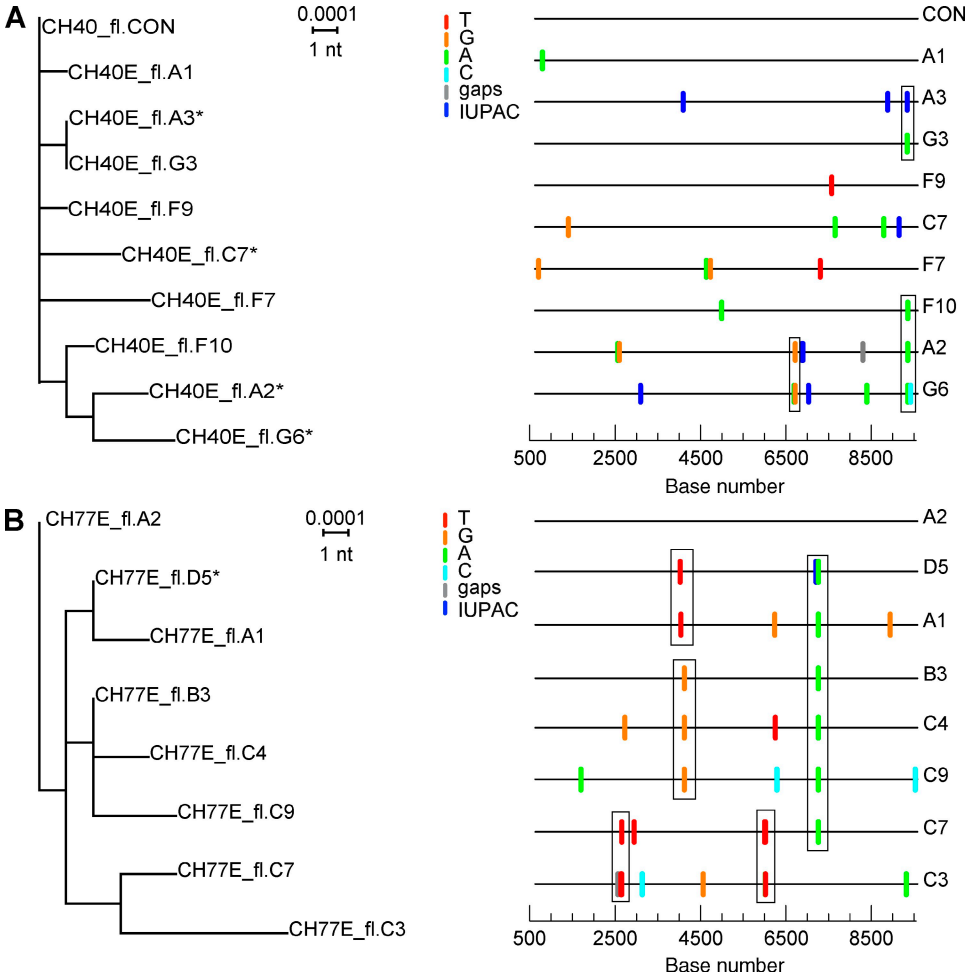

Figure 4. Phylogenetic and *Highlighter* analyses of HIV-1 sequences. (A and B) Analyses are from subjects CH40 (A) and CH77 (B). In *Highlighter* plots, the boxed tic marks indicate nucleotide polymorphisms shared between two or more sequences. Such shared mutations are also reflected by branches emanating from common nodes in the phylograms. In B, sequences are rooted on sequence A2, not on a consensus of the sequences. This is because sequence A2 was most similar to the transmitted/founder sequence, differing from it by only a single nucleotide at position 6,021 (see Results for explanation).

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

or in activated human lymphocytes, each time with similar results. Seven molecular clones of full-length clade B transmitted/founder viruses were also generated as part of a separate study and were tested for replication in human cells (unpublished data). These transmitted/founder viruses replicated efficiently in activated $CD4^+$ T lymphocytes but much less so in monocyte-derived macrophages.

To examine further the biological and antigenic properties of transmitted/founder viruses, we tested HIV-$1_{pZM246F-10}$, HIV-$1_{pZM247Fv1}$, and HIV-$1_{pZM247Fv2}$ for sensitivity to the receptor inhibitor soluble CD4 (sCD4); coreceptor inhibitors TAK-779 and AMD3100; fusion inhibitors T20 and T1249; and mAbs specific for the Env coreceptor binding surface (17b and 21c), CD4 binding site (b12), V3 loop (447-52D, F425-B4e8, and 3074), membrane proximal external region (2F5 and 4E10), and cell-surface CD4 (RPA-T4). We also tested the transmitted/founder viruses for sensitivity to heterologous subtype B and C plasmas to assess their overall neutralization sensitivities. The results are summarized in Table III and show that the transmitted/founder viruses exhibit properties typical of primary virus strains including CD4 and CCR5 dependence, effective concealment of the coreceptor binding bridging sheet and V3 loop structures, and generalized resistance to neutralization by heterologous plasma antibodies. The three clade C transmitted/founder viruses were, however, intermediately sensitive to 4E10 (IC_{50} = 5–48 μg/ml) and in one case (HIV-$1_{pZM246F-10}$) to b12 (IC_{50} = 21 μg/ml).

Molecular pathways and kinetics of virus diversification and adaptation

HIV-1 virions in plasma have an exceedingly short lifespan (<6 h), as do productively infected lymphocytes (~1 d) (13, 26–28). As a result, the genetic composition of plasma virus can change quickly and can provide a sensitive indicator of selection pressures acting on virus and virus-producing cells (16, 18, 20, 27, 28). These properties, combined with the identification of full-length transmitted/founder virus genomes, provided a unique opportunity to evaluate the kinetics and precise molecular pathways of HIV-1 sequence diversification and evolution on a genome-wide basis, because mutations could be mapped to specific viral genomes at or near the moment of transmission. In none of the 12 subjects did we find evidence of positive selection in viral sequences obtained before first antibody detection (i.e., before Fiebig stage III). Instead, we found evidence of selection against amino acid–conferring substitutions during this very early period, with $p_n - p_s$ values tending to be less than predicted for neutral evolution (Table S2). This is an expected finding, because evolutionary theory predicts most nucleotide substitutions to be neutral or deleterious. Early sequence diversification was notable for a high proportion of defective genomes: 27 sequences had frame-shifting insertions or deletions; 3 sequences had large in-frame deletions 318–1,329 nt in length; 5 sequences had gene open reading frames that were truncated because of in-frame stop codons arising from nucleotide point mutations; and 4 sequences contained evidence of APOBEC-mediated G-to-A hypermutation. Collectively, 34 out of 108 sequences contained mutations that rendered virus progeny overtly defective. Even this high proportion of defective genomes is an underestimate of the actual proportion of viruses or proviruses that contain defective genomes, because we excluded amplicons <7 kb in length and we did not examine the integrated proviral DNA compartment directly. In a separate study, we have done the latter and have found that approximately two thirds of HIV-1–infected cells in acutely infected subjects harbor overtly defective genomes (unpublished data).

We next examined sequences from a subset of subjects (CH40, CH58, and CH77) beginning at or near peak viremia (Fiebig stage II) and extending through 12–20 mo of follow-up (Fiebig stage VI). We did this first by identifying the transmitted/founder viral genome in each subject and then (using the same SGA-direct sequencing approach to avoid *Taq*-induced nucleotide substitutions and in vitro–generated recombination events) (12) by determining genomic sequences of plasma virus amplified as overlapping half genomes over the ensuing period (Fig. 5). In contrast to sequences obtained from subjects in Fiebig stage II, which generally contained few if any shared polymorphisms (Figs. 2 and 3; and Figs. S1 and S2), sequences from subjects at Fiebig stages V (CH40 and CH77) and III (CH58) exhibited shared polymorphisms and nonstar-like phylogenies (Figs. 4 and 5). This included nucleotide positions 2,625, 4,021, 4,104, 6,002, 6,021, and 7,285 in subject CH77; nucleotide positions 6,705, 9,360 and 9,371 in subject CH40; and nucleotide position 4,408 in subject CH58. By 32–45 d after the initial screening time point (labeled S in Fig. 5), clear patterns of nonrandom substitutions leading to nearly complete replacement of the transmitted/founder virus were evident in the sequences from each subject (Fig. 5). There was strong statistical evidence for positive selection in these highly focused regions that generally were embedded in epitopes that were early targets of the cellular immune system (22). The most rapid replacement of transmitted/founder virus sequences by mutant virus was observed in subject CH77, where within a period of 2 wk (between screen and day 14) nearly the entire replicating virus population in the body that contributed virus to the plasma compartment was replaced by viruses that shared concentrated mutations in two short regions, each spanning <30 nt in length, in *tat* and *env*. Equally remarkable was the finding that this mutant virus population in CH77 was in turn completely replaced by still another population that contained two additional sets of concentrated mutations in *nef*, with each again spanning <30 nt (compare day 14 and 32 sequences in CH77; Fig. 5). In these four discreet regions of selection in *tat*, *env*, and *nef* of subject CH77, we counted a total of 22 different nucleotide substitutions among the sequences at days 14 and 32. 20 out of 22 (91%) of these nucleotide substitutions were nonsynonymous. Selection for mutations was nearly as rapid in subjects CH40 and CH58 (Fig. 5), where such changes were first evident <9–16 d after antibody detection, which marks the end of Fiebig stage II and the beginning of stage III. Note in subject CH58 that selected mutations at positions 7,980

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

Figure 5. ***Highlighter*** **plots of sequences from longitudinal samples.** (A–C) Samples were from subjects CH40 (A), CH58 (B), and CH77 (C). The dark line represents the transmitted/founder sequence (T/F) for each subject. Sequences from different sample dates are indicated by the grouped horizontal lines, with tic marks indicating differences from the T/F sequence. The screening sample (S) was in each case obtained from a preseroconversion Fiebig stage II sample near peak viremia. The sampling times in days after screening and corresponding viral loads are indicated (left). Light blue tic marks indicate

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

and 7,986 in *env* had begun to accumulate as plasma virus load was still increasing. Based on inspection of sequences (Fig. 5), it was apparent by 32–45 d after screen that selected changes had resulted in nearly complete virus replacement at two genomic loci in CH58 (both in *env*) and five loci in CH40 (in *gag*, *vif*, *vpr*, and *nef*). Again, it was notable that nucleotide substitutions in each of these seven discreet loci were restricted to spans of <30 nt, and 17 out of 18 (94%) distinct nucleotide substitutions were nonsynonymous. The combination of a very high proportion of nonsynonymous nucleotide substitutions (>90%) concentrated in short spans of sequence (generally <30 nt) and affecting a high proportion of viral genomes (>90%) argues strongly for substantial selection pressure on the replicating virus quasispecies, a conclusion supported by formal statistical analysis and phenotypic testing (22). By 12–20 mo after infection, 17–34 discreet loci across the genomes showed evidence of strong mutational selection. Based on these findings, we concluded that the genetic imprint of a composite of host selection pressures influencing virus fitness in the critical period beginning at or near the moment of virus transmission and extending through the establishment of productive clinical infection and setpoint viremia had been captured in the sequences depicted in Fig. 5. In two companion studies, we show this genetic imprint to reflect both cellular and humoral immune responses (unpublished data) (22).

DISCUSSION

Understanding the precise molecular and biological events underlying HIV-1 transmission and subsequent steps leading to productive clinical infection could prove valuable in the development of effective HIV-1 vaccines and microbicides. But elucidating these events in a biologically and physiologically relevant context has been problematic because neither the particular viruses responsible for productive infection nor their initial target cells were known with certainty (2, 3, 5, 6, 29). In this paper, we take a step toward addressing this challenge by showing that complete genomes of transmitted/founder viruses can be identified, cloned, expressed, and analyzed phenotypically, and their evolution mapped precisely.

The current study builds on recent studies that describe a model of HIV-1 evolution in acute infection (11, 21) and empirical analyses of early *env* gene or *pro-pol* gene diversification (8–12, 21). We sought to extend this work to an analysis of full-length HIV-1 genomes and to address five questions specifically. First, can methods for SGA-direct sequencing of plasma vRNA/cDNA *env* sequences be applied successfully to full-length HIV-1 genomes? The answer is yes, but because of the fourfold increased length of the amplified genome, shared nucleotide polymorphisms and mixed bases are detected approximately four times more frequently. Such sequence ambiguities or polymorphisms are predicted by the model (see the mathematical model in Materials and methods) (11) and can be problematic for the correct identification of transmitted/founder viruses. However, we showed in this paper that they could be reconciled by the analysis of sequential sequences. A second and third question posed was if early diversification of full-length HIV-1 genomes conformed to a model of random evolution and if the identification of transmitted/founder viruses and estimates of time to a MRCA based on full-genome analyses corresponded to findings based on *env*-only gene analyses? The answer to both queries was yes. We have corroborated these findings by additional studies where we identified the transmitted/founder virus from PBMC proviral DNA and plasma vRNA from the same individuals using primer sets that amplified complete viral genomes, 5′ and 3′ half genomes, and *env*-only genomes (unpublished data) (9).

We also asked whether transmitted/founder full-length sequences have intact principal gene open reading frames and if they encode replication-competent viruses, as would be expected of transmitted/founder viruses responsible for spawning productive clinical infection? Indeed, each of the 13 transmitted/founder genomes that we identified had intact *gag*, *pol* (*rt*, *pro*, *int*), *vif*, *vpr*, *tat*, *rev*, *vpu*, *env*, and *nef* genes. To determine if transmitted/founder genomes encoded replication-competent viruses, we selected three full-length clade C sequences for analysis. Obtaining full-length recombinant clones of proviral genomes whose sequences match exactly those of transmitted/founder viruses can be challenging because of errors introduced by Superscript III MuLV RT or by *Taq* DNA polymerase in individual HIV-1 genomic sequences. Further complicating the cloning of replication-competent virus is an inherent pathogenicity of some *env* genes when replicated in bacteria. We circumvented these problems by pursuing two different strategies. First, we chemically synthesized and subcloned the complete transmitted/founder proviral genome of subject ZM246F in three overlapping subgenomic fragments, none of which included an intact *env* gene. The three fragments were cloned into a low copy number plasmid vector (Fig. 6 A, top), which was grown at reduced temperatures and for shorter durations than normal to prevent spontaneous deletions or other inactivating mutations. As a second approach, we used the high fidelity Phusion DNA polymerase to amplify in two overlapping proviral half genomes (neither containing an intact *env* gene) from preseroconversion high molecular weight lymphocyte DNA each of the two transmitted/founder viruses of subject ZM247F (Fig. 6 A, bottom).

differences in noncoding regions of the genomes. Gray tic marks indicate deletions. Green tic marks indicate differences that are synonymous in the reading frames indicated at the bottom of the figure. Red tic marks indicate differences that are nonsynonymous in a coding gene. Short horizontal lines (labeled A, B, and C) indicate short (~1-kb) sequences from a screening sample used to verify the presence or absence of a shared polymorphism (Figs. S3–S6). The horizontal axis indicates nucleotide positions in the alignment beginning at position 582 in the 5′ LTR (U5) and extending to position 9,606 in the 3′ LTR (R) based on HXB2 reference sequence numbering (http://www.hiv.lanl.gov/content/sequence/HIV/REVIEWS/HXB2.html).

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

This was followed by ligation and cloning of the complete proviruses into a low copy plasmid vector. The nucleotide sequence of each of the three proviruses was confirmed to be identical to the respective transmitted/founder viral sequences. Progeny viruses derived from 293T cells transfected with these three proviral clones replicated efficiently in activated primary human CD4+ T lymphocytes (Fig. 6 B, top). Surprisingly, none of the three transmitted/founder viruses replicated efficiently in monocyte-derived macrophages from the same donors (Fig. 6 B, bottom). This result was reproducible

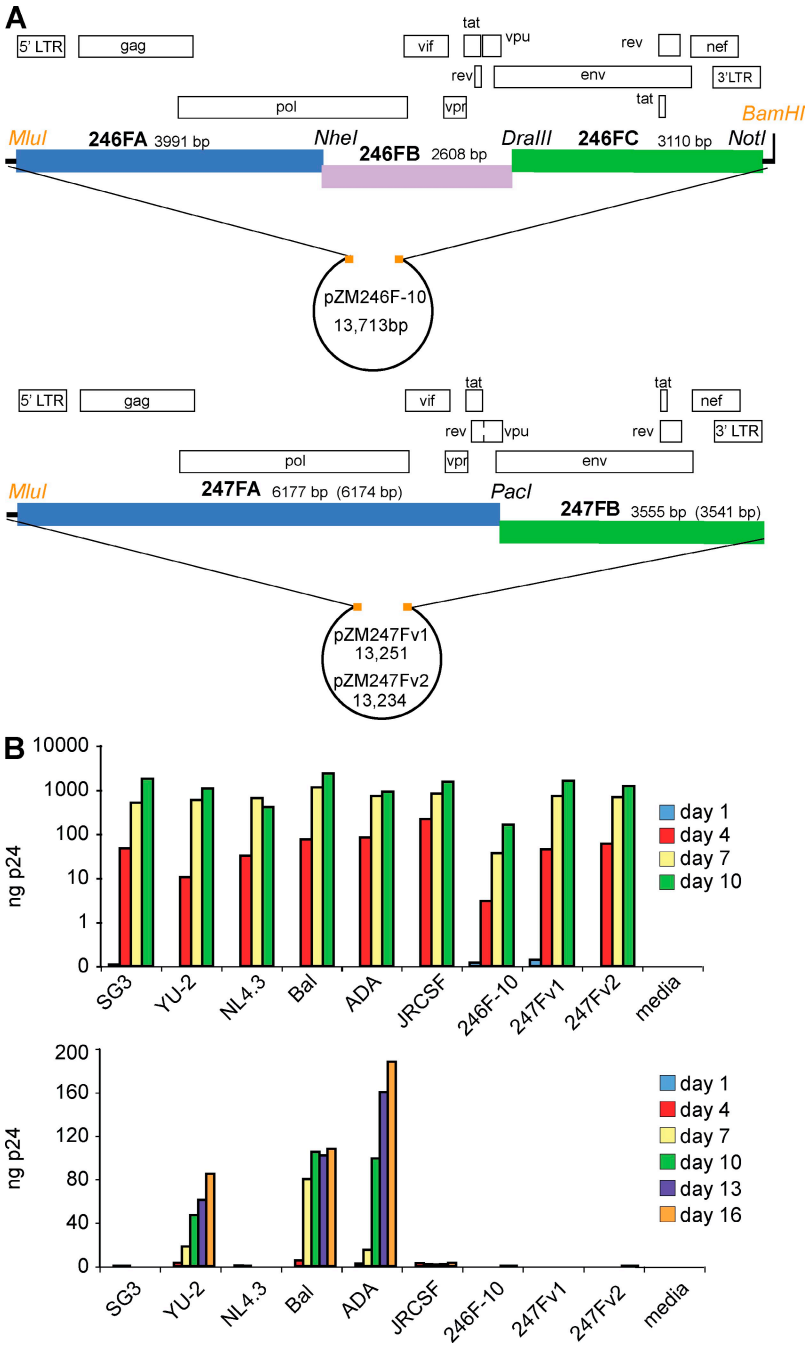

Figure 6. Cloning strategy and functional analysis of three subtype C full-length molecular clones from subjects ZM246F and ZM247F. (A) The full-length transmitted/founder proviral sequence for ZM246F was chemically synthesized in three overlapping fragments. The two transmitted/founder proviruses in subject ZM247F were PCR amplified from PBMC DNA in two overlapping halves. Proviral fragments were cloned into the indicated plasmid vectors. (B) Plasmid DNAs were transfected into 293T cells, and progeny viruses were examined for their ability to replicate in activated primary human CD4+ lymphocytes (top) and monocyte-derived macrophages (bottom) from the same donor. Culture supernatants were analyzed at days 1–16 for p24 production. The results shown were replicated four times in lymphocytes and macrophages from four different normal blood donors.

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

in lymphocytes and monocyte-derived macrophages from four different normal uninfected donors and was subsequently corroborated using viruses derived from full-length transmitted/founder proviral DNA clones from seven clade B subjects (unpublished data). We note that other investigators have previously observed that only a subset of primary R5 tropic viruses replicate efficiently in human monocyte-derived macrophages (30–33). These findings suggest that prototypic macrophage-tropic HIV-1 strains such as YU2, ADA, and BaL, which have frequently been used to model HIV-1 transmission, may not be representative of a substantial proportion of transmitted/founder viruses with respect to cell tropism. Our findings further suggest that during the initial stages of infection between transmission and peak viremia, replication of the transmitted/founder virus and its progeny in macrophages is unlikely to contribute substantially to overall virus production compared with lymphocyte-tropic viruses. This conclusion is consistent with other studies that have found that ostensibly "resting" memory CD4+ T cells (i.e., those lacking conventional markers of cell activation) and activated memory $CD4^+$ T cells, and not macrophages, are the principal early targets of HIV-1 and SIV infection in humans and primates (1, 2, 7, 29, 34–36).

Other biological properties of the three transmitted/founder viruses that we studied included their sensitivity to a large number of Env-specific ligands (Table III). Viruses HIV-$1_{pZM246F-10}$, HIV-$1_{pZM247Fv1}$, and HIV-$1_{pZM247Fv2}$ exhibited properties similar to viruses pseudotyped by other primary subtype B (11) or subtype C (37) HIV-1 Envs with respect to their CD4 and CCR5 dependence and sensitivity to the gp41 peptide fusion inhibitors T20 and T1249 and to the MPER mAb 4E10. Only HIV-$1_{pZM246F-10}$ was sensitive to B12, and none of the three were sensitive to 2G12 or 2F5, likely because of epitope variation on C clade viruses. The coreceptor binding surface of diverse HIV-1 strains is antigenically conserved (38), as is the V3 loop (39, 40), but HIV-$1_{pZM246F-10}$, HIV-$1_{pZM247Fv1}$, and HIV-$1_{pZM247Fv2}$ were resistant to CD4i (17b and 21C) and V3 (447-52D, F425-B4e8, and 3074) mAbs at concentrations as high as 25 μg/ml. When CD4i and V3 epitopes are exposed by sCD4 binding (38, 39) or in the context of HIV-2/HIV-1 Env chimeras (39, 40), the five CD4i and V3 mAbs neutralize viruses at IC_{50} concentrations between 0.001 and 5 μg/ml, indicating that the cognate epitopes are present but concealed in the native Env trimer (38–40). These findings suggest that CD4i bridging sheet and V3 epitopes on native clade C Env trimers are effectively shielded, as they are in native clade B Env trimers (11, 38–40). Finally, we found that 11 clade B and 22 clade C plasmas (all heterologous and containing high titers of CD4i and V3 cross-reactive polyclonal antibodies) (39, 40) exhibited low Nab titers against the three clade C viruses (IC_{50} < 1:40). This again suggested that the functional Env trimer on these transmitted/founder clade C viruses is effectively masked from antibody recognition.

The last question we asked was if the identification of transmitted/founder full-length genomes and their progeny could provide new insight into the kinetics and molecular pathways of HIV-1 diversification and adaptation leading to virus persistence? Between transmission and peak viremia (Fiebig stage II), we found that the virus diversified in an essentially random fashion, leaving no or little evidence of host-related selective pressures imprinted on its genome. 9–16 d later, evidence of striking selection on the virus quasispecies emerged in all three subjects studied, and by 32–45 d nearly the entire replicating virus population in each subject was replaced by viruses harboring mutations at two to five distinct loci (Fig. 5). In an accompanying paper, we model the kinetics and fitness costs associated with these rapidly evolving mutations and estimate the contributions of early CTL-mediated cell killing to virus containment (22). It is notable that for HIV-1 to establish early productive clinical infection (<3 mo from transmission), as few as two to five mutations were required across the entire transmitted/founder virus proteome. But for chronic persistent replication at ~6 mo after infection, some 4–14 loci of mutational selection were evident, and by 12–20 mo

Table III. Phenotype of transmitted HIV-1 clade C viruses compared with prototype clade B viruses

Virus[a]	sCD4	HIVIG	Clade B plasma	Clade C plasma	17b	21c	b12	2G12	2F5	4E10	Z13e1	447-52D	447-52D+ sCD4	F425-B4e8	F425-B4e8+ sCD4	3074	3074+ sCD4	T-20	T-1249	Anti-CD4 mab	TAK-779	Co-receptor usage
	nM	*μg/ml*	*(IC_{50})*[b]	*(IC_{50})*[c]	*μg/ml*	*μg/ml*	*μg/ml*	*μg/ml*	*μg/ml*	*μg/ml*	*μg/ml*	*μg/ml*	*μg/ml*	*μg/ml*	*μg/ml*	*μg/ml*	*μg/ml*	*μg/ml*	*μg/ml*	*μg/ml*	*μM*	
YU-2	13	619	20	20	>10	>10	7.16	>50	>50	>50	>50	13.36	1.48	>25	1.85	>25	>25	0.49	0.094	0.18	0.047	R5
NL4.3	1	60	92	31	0.21	0.59	0.01	1.08	0.24	1.05	>50	0.03	0.01	>25	>25	>25	>25	0.47	0.003	0.63	>10	X4
ADA	105	1,000	19	17	>10	>10	1.31	>50	33.2	>50	>50	>25	0.03	>25	0.11	>25	6.81	0.15	0.094	0.31	0.307	R5
NL.Bal. ecto	8	113	50	32	>10	>10	0.11	1.39	12.3	15.1	>50	0.14	0.04	0.54	0.05	>25	1.47	0.09	0.015	0.44	1.04	R5
ZM246F	125	748	13	31	>10	>10	21.35	>50	>50	5.50	>50	>25	>25	>25	>25	>25	3.35	0.07	0.012	0.17	0.003	R5
ZM247Fv1	155	>1,000	13	18	>10	>10	>50	>50	>50	11.33	>50	>25	>25	>25	>25	>25	>25	0.05	0.011	0.11	0.032	R5
ZM247Fv2	125	>1,000	12	20	>10	>10	>50	>50	>50	47.11	>50	>25	>25	>25	>25	>25	>25	0.08	0.027	0.16	0.034	R5

[a]Viruses derived by transfection of 293T cells with replication-competent molecularly cloned proviral DNA.
[b]Reciprocal geometric mean neutralization titers (IC_{50}) of heterologous plasma specimens from 11 HIV-1 clade B chronically infected subjects.
[c]Reciprocal geometric mean neutralization titers (IC_{50}) of heterologous plasma specimens from 22 HIV-1 clade C chronically infected subjects.

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

this number increased to 17–34 selected sites. Elucidation of adaptive and innate immune pressures that act to shape the evolving virus quasispecies, ultimately affecting viral load set-point and disease progression, is an important area for future investigation (22). Mapping mutations on a genome-wide basis against transmitted/founder virus sequences in recently infected individuals, including vaccine breakthrough infections, represents a unique strategy to explore the constellation of immune pressures acting on HIV-1 (8–12, 22), SIV (21), and other viral pathogens (41–43).

MATERIALS AND METHODS

Study subjects. Peripheral blood samples were obtained from 12 adult subjects who gave informed consent under clinical protocols approved by the participating institutions' human use review boards, including those at the University of Alabama at Birmingham, the University of North Carolina at Chapel Hill, Emory University, the Rockefeller University, and Duke University. Blood was generally collected in acid citrate dextrose and plasma separated and stored at −70°C.

Laboratory staging. Plasma samples were tested for HIV-1 RNA, p24 antigen, and viral-specific antibodies as previously described (11, 17). The duration of the eclipse phase (the period between virus transmission and the detection of viral RNA in plasma) was estimated to be 10 d (range = 7–21 d) (11, 44–49).

Viral RNA extraction and cDNA synthesis. Approximately 20,000 viral RNA copies were extracted using the QIAamp Viral RNA Mini Kit (QIAGEN) or the BioRobot EZ1 Workstation with EZ1 Virus Mini Kit (version 2.0; QIAGEN). Samples with low vRNA loads (200–10,000 copies/ml) were concentrated by centrifugation at 23,600 *g* for 1 h at 4°C before extraction. RNA was recovered from spin columns in a final elution volume of 60 µl. RNA was either frozen at −80°C or immediately used to synthesize cDNA. 5,000 (or fewer) vRNA molecules were reverse transcribed using SuperScript III according to the manufacturer's instructions (Invitrogen) and modified as follows. RT reactions were performed in 40-µl reactions. First, 15 µl RNA was mixed with 11 µl of a master mix A containing 0.5 mM of each deoxynucleoside triphosphate (dNTP) and 0.25 µM of reverse primer 1.R3.B3R (5′-ACTACTTGAAGCACTCAAGGCAAGCTTTATTG-3′; nt 9,611–9,642; HXB2 numbering) and incubated for 5 min at 65°C to denature secondary RNA structure. cDNA generated by reverse primer 1.R3.B3R served as a template for PCR amplification of a 9-kb or a 5-kb fragment corresponding to the 3′ half viral genome. A separate RT reaction was performed using reverse primer B5R1 (5′-CTTGCCACACAATCATCACCTGCCAT-3′; nt 5,052–5,077) to generate cDNA that served as a template to amplify a 5-kb 5′ half viral genome fragment. After denaturation, the tube was incubated on ice for 1 min. Then, 26 µl of the reaction mixture was combined with a master mix B–containing RT buffer (1×), 5 mM dithiothreitol, 2 U/µl of the RNase inhibitor RNaseOUT, and 5 U/µl SuperScript III. Tubes were then incubated at 50°C for 1.5 h. An additional 1 µl SuperScript III was added and tubes were incubated for another 1.5 h at 55°C. After the completion of the RT step, the reaction mixture was inactivated by heat (70°C for 15 min), followed by the addition of 1 µl RNase H, and incubated at 37°C for 20 min. The resulting cDNA was aliquoted and frozen at −80°C until further analysis or used immediately for PCR amplification.

SGA. cDNA was serially diluted in replicates of eight PCR wells and subjected to nested PCR amplification with HIV-specific primers that yield a 9-kb fragment beginning at the first nucleotide of the U5 region of the 5′ long terminal repeat (LTR) and extending to the last nucleotide of the R region of the 3′ LTR. cDNA dilutions that yielded 40% or fewer PCR-positive wells were retested in 96-well plates to identify a dilution where <20% of wells were positive for amplification products. First-round PCR was performed in 1× Expand Long Template buffer 1 (1.75 mM $MgCl_2$ final concentration), 0.35 mM of each dNTP, 0.3 µM of forward primer 1.U5.B1F (5′-CCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGT-3′; nt 538–571), reverse primer 1.R3.B3R, and 3.75 U/µl Expand Long Template Mix (Roche) in a 50-µl reaction mixture. The PCR mixtures were set up in optical 96-well reaction plates (MicroAmp; Applied Biosystems) and sealed with MicroAmp adhesive film. PCR conditions were 94°C for 2 min, followed by 10 cycles of 94°C for 15 s, 55°C for 30 s, and 68°C for 8 min, followed by 25 cycles of 94°C for 15 s, 55°C for 30 s, and 68°C for 8 min, with cumulative increments of 20 s at 68°C with each successive cycle and a final extension period of 10 min at 68°C. Second-round PCR was performed by transferring 1 µl of the first-round product into a final volume of 50 µl of a reaction mixture containing 1× Expand Long Template buffer, 0.35 mM of each dNTP, and 0.3 µM of forward primer 2.U5.B4F (5′-AGTAGTGTGTGCCCGTCTGTTGTGTGACTC-3′; nt 552–581) and reverse primer 2.R3.B6R (5′-TGAAGCACTCAAGGCAAGCTTTATTGAGGC-3′; nt 9,607–9,636). PCR conditions were the same as for the first-round PCR. The amplicons were sized on precast 1% agarose E-gel 48 (Invitrogen). All products derived from cDNA dilutions yielding < 20% PCR-positive wells and >7-kb in length by analytical gel electrophoresis were sequenced. Because genomic length viral RNA is flanked by the R region of the LTR, only partial LTR sequences and near-complete viral genomes can be obtained in a single PCR amplification step. Specifically, the internal nested 5′ primer (2.U5.B4F) corresponds to the first 30 nt of U5 (HXB2 nt 552–581), and the internal nested 3′ primer (2.R3.B6R) corresponds to the last 30 nt of the 3′ R region (HXB2 nt 9,607–9,636), thus rendering these two short stretches of 30 nt primer derived rather than virus derived. To determine complete transmitted/founder viral RNA/cDNA sequences that include viral sequences occupied by 5′ U5 and 3′ R primers, a separate PCR reaction was performed using primers located in the 5′ R region and in *gag*. For this amplification, nested PCR primers corresponded to the 5′ R (HXB2 nt 452–474 and 456–482) and *gag* (HXB2 nt 1,401–1,422 and 1,432–1,452), and a 1-kb amplicon was generated and sequenced. PCR reaction conditions were the same as for the full-length amplification except for a reduction in extension time to 1 min. Collectively, the resulting amplicons and sequences encompassed the complete HIV-1 viral genome. To amplify the complete HIV-1 genome in two partially overlapping 5-kb fragments, we modified the SGA procedure previously described for the HIV-1 *env* gene (11). PCR amplifications were performed in the presence of 1× High Fidelity Platinum PCR buffer, 2 mM $MgSO_4$, 0.2 mM of each dNTP, 0.2 µM of each primer, and 0.025 U/µl Platinum Taq High Fidelity polymerase (Invitrogen) in a 20-µl reaction. First-round PCR primers for amplifying the 5′ half genome were 1.U5.B1F and B5R1. PCR reactions were performed in MicroAmp 96-well reaction plates as follows: 1 cycle of 94°C for 2 min, and 35 cycles of a denaturing step of 94°C for 15 s, an annealing step of 55°C for 30 s, and an extension step of 68°C for 5 min, followed by a final extension of 10 min at 68°C. 2 µl from the first-round PCR product was added to a second-round PCR that included the 5′ half genome forward primer 2.U5.B4F and reverse primer B5R2 (5′-CAATCATCACCTGCCATCTGTTTTCCATA-3′; nt 5,040–5,068). The second-round PCR reaction was performed under the same conditions indicated for the first-round PCR but with a total of 45 cycles. To amplify the 3′ half genome, first-round PCR primers included B3F1 (5′-ACAGCAGTACAAATGGCAGTATT-3′; nt 4,749–4,771) and 1.R3.B3. The first-round PCR reaction conditions were the same as for the 5′ half genome. Next, 2 µl from the first-round PCR product was added to a second-round PCR that included primers B3F3 (5′-TGGAAAGGTGAAGGGGCAGTAGTAATAC-3′; nt 4,956-4,983) and 2.R3.B6R. PCR conditions were the same as for the second-round amplification of the 5′ half genome. All products derived from cDNA dilutions yielding <30% PCR-positive wells and >4-kb amplicons were sequenced. All PCR procedures were performed under PCR clean-room conditions (11).

DNA sequencing. Amplicons were directly sequenced, aligned, and analyzed as previously described (11, 50, 51). APOBEC-associated G-to-A hypermutated sequences were identified and excluded from model analyses. The full

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

sequence alignment is available as a supplemental data file (www.hiv.lanl.gov/content/sequence/hiv/user_alignments/Salazar), and sequences are available from GenBank/EMBL/DDBJ under accession nos. FJ495804–FJ496214, EU578998–EU579004, EU579006–EU579015, EU579018–EU579019, EU576471–EU576521, and FJ919955–FJ919967.

Mathematical model. The model used in the present study has been fully described (11). Under this model, with no selection pressure and fast expansion, one can expect small samples from homogeneous virus populations to have evolved from a founder strain in a star-like phylogeny, with all sequences coalescing at the founder (11, 21). Occasional deviations from a star phylogeny are, however, expected. The sampling of 10 sequences, for example, from a later generation of an exponentially growing population with sixfold growth per generation ($R_0 = 6$) has ~4% chance of including a pair sharing the first four generations, 19% chance of including sequences that share three, and 75% chance of sharing two. With a predicted rate of mutations of approximately one per five generations for the full-length 9-kb HIV-1 genome, there is ~75% chance of finding among 10 sequences 2 that share one mutation, ~20% chance of finding 2 sequences that share a pair of mutations, and <2% chance of sharing more than that. These probabilities are slightly enhanced by early stochastic events that can lead to the virus producing less than six descendants in some generations but are diminished by the chances that mutations cause a fitness disadvantage that results in early purifying selection, as previously observed (11, 24).

Proviral DNA cloning. To obtain an infectious molecular clone of the transmitted/founder virus from subject ZM246F, the entire 9.7-kb sequence was first inferred from SGA analysis of plasma viral RNA and then chemically synthesized in three subgenomic fragments of 4, 2.6, and 3.1 kb (Blue Heron Biotechnology). These fragments overlapped at unique *Nhe*I (HXB2 position 3,998) or *Dra*III (HXB2 position 6,600) sites, respectively, and were propagated in plasmid vectors. A 5′ terminal *Mlu*I cloning site that was attached during chemical synthesis and a *Bam*HI polylinker site at the 3′ end of the genome were used for subsequent manipulation. To facilitate assembly of a full-length proviral clone, a unique *Mlu*I site was engineered by oligonucleotide adapter ligation between *Eco*RI and *Hind*III of pBR322. The 4-kb 5′ fragment was then cloned in *Mlu*I-NheI of the modified pBR322 vector. After restriction enzyme digestion of this recombinant plasmid with *Nhe*I and *Bam*HI, the full-length proviral clone pZM246F-10 was generated by simultaneous ligation with the 2.6-kb middle *Nhe*I-*Dra*III and the 3.1-kb 3′ *Dra*III-*Bam*HI fragments. The integrity of the entire proviral consensus sequence was confirmed by restriction endonuclease mapping, and finally, by nucleotide sequence analysis.

To obtain infectious molecular clones of the two transmitted/founder viruses from subject ZM247F, we used a high fidelity DNA polymerase and proviral DNA as a template to obtain both 5′ and 3′ half genomes, each with a complete LTR element. Primers were designed to correspond exactly to the transmitted/founder sequence as determined by SGA-direct amplicon sequencing. For the 5′ half genome, the entire 5′ U3-R-U5, *gag*, *pol*, *vif*, *vpr tat1*, *rev1*, and *vpu*, and part of *env* were amplified in a single-round PCR. For the 3′ half genome, the entire 3′ U5-R-U3, *nef*, *tat2*, and *rev2*, and the remainder of *env* were amplified. High molecular weight genomic DNA was isolated from PBMCs obtained on the same day that plasma virus was used to identify the two unique transmitted/founder virus lineages. The complete proviral genomes corresponding to both virus lineages were amplified and cloned in two overlapping halves encompassing a unique restriction site, *Pac*I. Single-round PCR amplification was performed in the presence of 1× Phusion Hot Start HF Buffer, 0.2 mM of each dNTP, 0.5 μM of each primer, 3% final concentration of DMSO, and 0.02 U/μl Phusion Hot Start High Fidelity DNA polymerase (New England Biolabs, Inc.) in a 40-μl reaction. PCR primers for the variant 1 are V15′F (5′-TGGAAGGGTTAATTTACTCCAAGAAAAGG-3′; nt 1–29; HXB2 numbering) and V15′R (5′-CACTGTCTTCTGCCCTTTCTCTAATTCTTT-3′; nt 6,192–6,221), and V13′F (5′-TAGGAAATTGGTAAGACAAAGAAAAATAGACTGG-3′; nt 6,151–6,184) and V13′R (5′-TGCTAGAGATTTTCCACACTACCAAAATGG-3′; nt 9,689-del-9,719). PCR primers for variant 2 are V25′F (5′-TGGAAGGGTTAGTTTACTCCAAGAAAAGG-3′; nt 1–29) and V25′R (5′-CATGGTGTTGGTATTATTGCTGGTAGCAG-3; nt 6,643-del-6,686), and V23′F (5′-GGCTCCATAGCTTAGGGCAACATATCTATA-3′; nt 5,671–5,700) and V23′R (5′-TGCTAGAGATTTTCCACACTACCAAAATGG-3′; nt 9,689-del-9,719). PCR was performed under the following conditions: 1 cycle of 99°C for 1 min, and 35 cycles of 99°C for 8 s, 65°C for 30 s, and a 72°C extension for 4 min, followed by a final extension of 72°C for 10 min. Correctly sized fragments (~6 kb for 5′ and ~4 kb for 3′) were identified by gel electrophoresis. An adenine overhang was added to each purified fragment using *Taq* polymerase with 1× buffer (Promega) and 0.2 mM of each dNTP incubated at 94°C for 2 min followed by a single extension of 72°C for 10 min. Each half genome was then independently T/A cloned into the TOPO-XL vector (Invitrogen) and transformed into XL2-Blue MRF competent bacteria (Agilent Technologies). Bacteria were plated on lysogeny broth (LB) agar plates supplemented with 50 μg/ml kanamycin and cultured overnight at 30°C. Single colonies were selected and grown overnight in liquid LB at 30°C with 225-rpm shaking followed by plasmid isolation. Each molecular clone was sequence confirmed to be identical to the transmitted/founder viral sequence for each lineage. The 5′ clone for each lineage was ligated to the cognate 3′-XL vector by using *Pac*I and *Mlu*I (variant 1) or *Not*I (variant 2) restriction digestion and ligation (New England Biolabs Inc.), thereby generating the infectious clones for lineage 1 (pZM247Fv1) and lineage 2 (pZM247Fv25).

Virus phenotypic analysis. The ability of cloned viral genomes to express replication-competent virus was assessed using 293T cells for DNA transfection and primary human $CD4^+$ lymphocytes, monocyte-derived macrophages, and JC53BL-13 cells (TZM-bl; National Institutes of Health AIDS Research and Reference Reagent Program) as target cells, as previously described (20, 52–54). Virus neutralization by Env-specific antibodies and ligands was similarly tested as previously described (11, 20, 38, 52). All samples were tested in duplicate and all experiments were repeated at least three times to ensure reproducibility.

Virus replication was assessed in activated primary human $CD4^+$ T cells and in monocyte-derived macrophages obtained from the same normal human donors. Peripheral blood was collected into ACD-A anticoagulant by venipuncture and mononuclear cells (PBMCs) purified by standard Ficoll-Hypaque density gradient methods. For $CD4^+$ T cell isolation, PBMCs were incubated with human anti-CD4–coated magnetic beads (Militenyi Biotec) and positively selected using a cell separator (autoMACS; Militenyi Biotec). They were then incubated in 10-cm polystyrene tissue culture plates for 2 h at 37°C in HBSS with 10 mM Ca^{2+} and Mg^{2+} to remove adherent monocytes. Nonadherent cells were collected, washed in RPMI 1640 plus 15% FBS, and resuspended at 10^6 cells/ml in RPMI 1640 plus 15% FBS with 3 μg/ml staphyloccocal enterotoxin B (Sigma-Aldrich) for 48 h at 37°C to activate the lymphocytes. 5×10^5 cells were incubated with 50,000 IU of virus overnight at 37°C in 300 μl RPMI 1640 with 15% FBS. Cells were washed three times and plated in 24-well polystyrene tissue culture plates in a volume of 2 ml RPMI 1640 with 15% FBS and 30 U IL-2/ml. 60 μl of media was removed for day 1 p24 and RT baseline analysis. Every 3 d, 60 μl of media was removed and frozen for p24/RT analysis and half of the media was removed from each well and replaced with fresh media. For monocyte-derived macrophage isolation, 3×10^6 PBMCs per well were plated in 24-well plates in HBSS plus 10 mM Ca^{2+} and Mg^{2+} plus 10% human AB serum (Sigma-Aldrich) and incubated at 37°C for 2 h. Nonadherent cells were removed and DMEM with 10% giant cell tumor (GCT)–conditioned media (Irvine Scientific) plus 10% human AB serum was added to wells with 5 U/ml rhMCSF (R&D Systems). After 3 d of incubation, wells were washed vigorously with PBS three times, and media containing DMEM with 10% GCT, 10% FBS, and 5 U rhMCSF was added to wells. After an additional 3 d of incubation, media was removed and macrophages were incubated with 100,000 IU of virus in 300 μl per well. After a 2-h incubation, 500 μl of media was added per well. After overnight incubation, each well was washed three times, 1.5 ml of media was added, and 60 μl of media was removed for day 1 p24/RT

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

baseline analysis. Every 3 d, 60 µl of culture supernatant was removed and frozen for p24/RT analysis, whereas the remainder of the media was removed from each well and replaced with 1.5 ml of fresh media. Cultures were continued for 16 d.

Online supplemental material. Fig. S1 shows phylogenetic and *Highlighter* analyses of HIV-1 sequences from subjects SUMA0847, TRJO4551, 04013396-0, and ZM249M. Fig. S2 shows phylogenetic and *Highlighter* analyses of HIV-1 sequences from subjects ZM246F, CH58, 04013226-2, and WEAU0575. Fig. S3 shows nucleotide sequence alignments of polymorphic regions in *env* and *nef* in early samples from subject CH40. Fig. S4 shows nucleotide sequence alignments of polymorphic regions in *tat* and *env* in early samples from subject CH77. Fig. S5 shows nucleotide sequence alignments of polymorphic regions in *pol* in early samples from subject CH77. Fig. S6 shows nucleotide sequence alignments of polymorphic regions in *pol* in early samples from subject CH58. Table S1 shows nucleotide gene diversity in four protein-coding regions in early progeny of transmitted/early founder viruses. Table S2 provides an analysis of p_n and p_s for *gag*, *pol*, *env*, and *nef*. Online supplemental material is available at http://www.jem.org/cgi/content/full/jem.20090378/DC1.

We thank D. McPherson, Y. Chen, and B. Cochran for technical assistance; the clinical cores of the Center for HIV/AIDS Vaccine Immunology and the University of Alabama at Birmingham Center for AIDS Research; and J. White for manuscript preparation.

This work was supported by the Center for HIV/AIDS Vaccine Immunology and by grants from the National Institutes of Health (AI067854, AI061734, and AI27767) and the Bill and Melinda Gates Foundation (37874).

The authors have no conflicting financial interests.

Submitted: 17 February 2009
Accepted: 1 May 2009

REFERENCES

1. Brenchley, J.M., T.W. Schacker, L.E. Ruff, D.A. Price, J.H. Taylor, G.J. Beilman, P.L. Nguyen, A. Khoruts, M. Larson, A.T. Haase, and D.C. Douek. 2004. CD4$^+$ T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *J. Exp. Med.* 200:749–759.
2. Haase, A.T. 2005. Perils at mucosal front lines for HIV and SIV and their hosts. *Nat. Rev. Immunol.* 5:783–792.
3. Margolis, L., and R. Shattock. 2006. Selective transmission of CCR5-utilizing HIV-1: the 'gatekeeper' problem resolved? *Nat. Rev. Microbiol.* 4:312–317.
4. Mehandru, S., M.A. Poles, K. Tenner-Racz, A. Horowitz, A. Hurley, C. Hogan, D. Boden, P. Racz, and M. Markowitz. 2004. Primary HIV-1 infection is associated with preferential depletion of CD4$^+$ T lymphocytes from effector sites in the gastrointestinal tract. *J. Exp. Med.* 200:761–770.
5. Pope, M., and A.T. Haase. 2003. Transmission, acute HIV-1 infection and the quest for strategies to prevent infection. *Nat. Med.* 9:847–852.
6. Shattock, R.J., and J.P. Moore. 2003. Inhibiting sexual transmission of HIV-1 infection. *Nat. Rev. Microbiol.* 1:25–34.
7. Veazey, R.S., I.C. Tham, K.G. Mansfield, M. DeMaria, A.E. Forand, D.E. Shvetz, L.V. Chalifoux, P.K. Sehgal, and A.A. Lackner. 2000. Identifying the target cell in primary simian immunodeficiency virus (SIV) infection: highly activated memory CD4(+) T cells are rapidly eliminated in early SIV infection in vivo. *J. Virol.* 74:57–64.
8. Abrahams, M.R., J.A. Anderson, E.E. Giorgi, C. Seoighe, K. Mlisana, L.-H. Ping, G.S. Athreya, F.K. Treurnichat, B.F. Keele, N. Wood, et al. 2009. Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J. Virol.* 83:3556–3567.
9. Haaland, R.E., P.A. Hawkins, J. Salazar-Gonzalez, A. Johnson, A. Tichacek, E. Karita, O. Manigart, J. Mulenga, B.F. Keele, G.M. Shaw, et al. 2009. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS Pathog.* 5:e1000274.
10. Kearney, M., F. Maldarelli, W. Shao, J.B. Margolick, E.S. Daar, J.W. Mellors, V. Rao, J.M. Coffin, and S. Palmer. 2009. Human immunodeficiency virus type 1 population genetics and adaptation in newly infected individuals. *J. Virol.* 83:2715–2727.
11. 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, et al. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. USA.* 105:7552–7557.
12. 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, et al. 2008. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J. Virol.* 82:3952–3970.
13. Markowitz, M., M. Louie, A. Hurley, E. Sun, M. Di Mascio, 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.
14. 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.
15. 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.
16. Borrow, P., H. Lewicki, X. Wei, M.S. Horwitz, N. Peffer, H. Meyers, J.A. Nelson, J.E. Gairin, B.H. Hahn, M.B. Oldstone, and G.M. Shaw. 1997. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat. Med.* 3:205–211.
17. 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.
18. Jones, N.A., X. Wei, D.R. Flower, M. Wong, F. Michor, M.S. Saag, B.H. Hahn, M.A. Nowak, G.M. Shaw, and P. Borrow. 2004. Determinants of human immunodeficiency virus type 1 escape from the primary CD8$^+$ cytotoxic T lymphocyte response. *J. Exp. Med.* 200:1243–1256.
19. Richman, D.D., T. Wrin, S.J. Little, and C.J. Petropoulos. 2003. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl. Acad. Sci. USA.* 100:4144–4149.
20. Wei, X., J.M. Decker, S. Wang, H. Hui, J.C. Kappes, X. Wu, J.F. Salazar-Gonzalez, M.G. Salazar, J.M. Kilby, M.S. Saag, et al. 2003. Antibody neutralization and escape by HIV-1. *Nature.* 422:307–312.
21. Keele, B.F., H. Li, G.H. Learn, P. Hraber, E.E. Giorgi, T. Grayson, C. Sun, Y. Chen, W. Yeh, N.L. Letvin, et al. 2009. Low-dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1. *J. Exp. Med.* 206:1117–1134.
22. Goonetilleke, N., M.K.P. Liu, J.F. Salazar-Gonzalez, G. Ferrari, E. Giorgi, V.V. Ganusov, B.F. Keele, G.H. Learn, E.L. Turnbull, M.G. Salazar, et al. 2009. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J. Exp. Med.* 206:1253–1272.
23. Zhang, L.Q., P. MacKenzie, A. Cleland, E.C. Holmes, A.J. Brown, and P. Simmonds. 1993. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *J. Virol.* 67:3345–3356.
24. Wood, N., T. Bhattacharya, B.F. Keele, E. Giorgi, M. Liu, B. Gaschen, M. Daniels, G. Ferrari, B.F. Haynes, A. McMichael, et al. 2009. HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. *PLoS Pathog.* 5:e1000414.
25. Bracho, M.A., A. Moya, and E. Barrio. 1998. Contribution of Taq polymerase-induced errors to the estimation of RNA virus diversity. *J. Gen. Virol.* 79:2921–2928.
26. Ramratnam, B., S. Bonhoeffer, J. Binley, A. Hurley, L. Zhang, J.E. Mittler, M. Markowitz, J.P. Moore, A.S. Perelson, and D.D. Ho. 1999. Rapid production and clearance of HIV-1 and hepatitis C virus assessed by large volume plasma apheresis. *Lancet.* 354:1782–1785.

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026

27. Wei, X., S.K. Ghosh, M.E. Taylor, V.A. Johnson, E.A. Emini, P. Deutsch, J.D. Lifson, S. Bonhoeffer, M.A. Nowak, B.H. Hahn, et al. 1995. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature*. 373:117–122.
28. Ho, D.D., A.U. Neumann, A.S. Perelson, W. Chen, J.M. Leonard, and M. Markowitz. 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*. 373:123–126.
29. Hladik, F., P. Sakchalathorn, L. Ballweber, G. Lentz, M. Fialkow, D. Eschenbach, and M.J. McElrath. 2007. Initial events in establishing vaginal entry and infection by human immunodeficiency virus type-1. *Immunity*. 26:257–270.
30. Duenas-Decamp, M.J., P. Peters, D.R. Burton, and P.R. Clapham. 2009. Determinants flanking the CD4 binding loop modulate macrophage-tropism of HIV-1 R5 envelopes. *J. Virol*. 83:2575–2583.
31. Dunfee, R.L., E.R. Thomas, P.R. Gorry, J. Wang, J. Taylor, K. Kunstman, S.M. Wolinsky, and D. Gabuzda. 2006. The HIV Env variant N283 enhances macrophage tropism and is associated with brain infection and dementia. *Proc. Natl. Acad. Sci. USA*. 103:15160–15165.
32. Peters, P.J., W.M. Sullivan, M.J. Duenas-Decamp, J. Bhattacharya, C. Ankghuambom, R. Brown, K. Luzuriaga, J. Bell, P. Simmonds, J. Ball, and P.R. Clapham. 2006. Non-macrophage-tropic human immunodeficiency virus type 1 R5 envelopes predominate in blood, lymph nodes, and semen: implications for transmission and pathogenesis. *J. Virol*. 80:6324–6332.
33. Xu, Y., H. Zhu, C.K. Wilcox, A. van't Wout, T. Andrus, N. Llewellyn, L. Stamatatos, J.I. Mullins, L. Corey, and T. Zhu. 2008. Blood monocytes harbor HIV type 1 strains with diversified phenotypes including macrophage-specific CCR5 virus. *J. Infect. Dis*. 197:309–318.
34. Mattapallil, J.J., D.C. Douek, B. Hill, Y. Nishimura, M. Martin, and M. Roederer. 2005. Massive infection and loss of memory CD4+ T cells in multiple tissues during acute SIV infection. *Nature*. 434:1093–1097.
35. Schacker, T., S. Little, E. Connick, K. Gebhard, Z.Q. Zhang, J. Krieger, J. Pryor, D. Havlir, J.K. Wong, R.T. Schooley, et al. 2001. Productive infection of T cells in lymphoid tissues during primary and early human immunodeficiency virus infection. *J. Infect. Dis*. 183:555–562.
36. Veazey, R.S., M. DeMaria, L.V. Chalifoux, D.E. Shvetz, D.R. Pauley, H.L. Knight, M. Rosenzweig, R.P. Johnson, R.C. Desrosiers, and A.A. Lackner. 1998. Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection. *Science*. 280:427–431.
37. Li, M., J.F. Salazar-Gonzalez, C.A. Derdeyn, L. Morris, C. Williamson, J.E. Robinson, J.M. Decker, Y. Li, M.G. Salazar, V.R. Polonis, et al. 2006. Genetic and neutralization properties of subtype C human immunodeficiency virus type 1 molecular env clones from acute and early heterosexually acquired infections in Southern Africa. *J. Virol*. 80:11776–11790.
38. Decker, J.M., F. Bibollet-Ruche, X. Wei, S. Wang, D.N. Levy, W. Wang, E. Delaporte, M. Peeters, C.A. Derdeyn, S. Allen, et al. 2005. Antigenic conservation and immunogenicity of the HIV coreceptor binding site. *J. Exp. Med*. 201:1407–1419.
39. Davis, K.L., F. Bibollet-Ruche, H. Li, J.M. Decker, O. Kutsch, L. Morris, A. Salomon, A. Pinter, J.A. Hoxie, B.H. Hahn, et al. 2009. Human immunodeficiency virus type 2 (HIV-2)/HIV-1 envelope chimeras detect high titers of broadly reactive HIV-1 V3-specific antibodies in human plasma. *J. Virol*. 83:1240–1259.
40. Davis, K.L., E.S. Gray, P.L. Moore, J.M. Decker, A. Salomon, D.C. Montefiori, B.S. Graham, M.C. Keefer, A. Pinter, L. Morris, et al. 2009. High titer HIV-1 V3-specific antibodies with broad reactivity but low neutralizing potency in acute infection and following vaccination. *Virology*. 387:414–426.
41. Kuntzen, T., J. Timm, A. Berical, L.L. Lewis-Ximenez, A. Jones, B. Nolan, J. Schulze zur Wiesch, B. Li, A. Schneidewind, A.Y. Kim, R.T. Chung, G.M. Lauer, and T.M. Allen. 2007. Viral sequence evolution in acute hepatitis C virus infection. *J. Virol*. 81:11658–11668.
42. Morimoto, K., D.C. Hooper, H. Carbaugh, Z.F. Fu, H. Koprowski, and B. Dietzschold. 1998. Rabies virus quasispecies: implications for pathogenesis. *Proc. Natl. Acad. Sci. USA*. 95:3152–3156.
43. Towner, J.S., M.L. Khristova, T.K. Sealy, M.J. Vincent, B.R. Erickson, D.A. Bawiec, A.L. Hartman, J.A. Comer, S.R. Zaki, U. Stroher, et al. 2006. Marburgvirus genomics and association with a large hemorrhagic fever outbreak in Angola. *J. Virol*. 80:6497–6516.
44. Clark, S.J., M.S. Saag, W.D. Decker, S. Campbell-Hill, J.L. Roberson, P.J. Veldkamp, J.C. Kappes, B.H. Hahn, and G.M. Shaw. 1991. High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection. *N. Engl. J. Med*. 324:954–960.
45. Gaines, H., M. von Sydow, P.O. Pehrson, and P. Lundbegh. 1988. Clinical picture of primary HIV infection presenting as a glandular-fever-like illness. *BMJ*. 297:1363–1368.
46. Lindbäck, S., A.C. Karlsson, J. Mittler, A. Blaxhult, M. Carlsson, G. Briheim, A. Sonnerborg, and H. Gaines. 2000a. Viral dynamics in primary HIV-1 infection. Karolinska Institutet Primary HIV Infection Study Group. *AIDS*. 14:2283–2291.
47. Lindbäck, S., R. Thorstensson, A.C. Karlsson, M. von Sydow, L. Flamholc, A. Blaxhult, A. Sonnerborg, G. Biberfeld, and H. Gaines. 2000b. Diagnosis of primary HIV-1 infection and duration of follow-up after HIV exposure. Karolinska Institute Primary HIV Infection Study Group. *AIDS*. 14:2333–2339.
48. Little, S.J., A.R. McLean, C.A. Spina, D.D. Richman, and D.V. Havlir. 1999. Viral dynamics of acute HIV-1 infection. *J. Exp. Med*. 190:841–850.
49. Schacker, T., A.C. Collier, J. Hughes, T. Shea, and L. Corey. 1996. Clinical and epidemiologic features of primary HIV infection. *Ann. Intern. Med*. 125:257–264.
50. Maddison, W.P., and D.R. Maddison. 2001. MacClade 4: Analysis of Phylogency and Character Evolution. Sinauer Associates, Inc., Sunderland, MA. 503 pp.
51. Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22:4673–4680.
52. Wei, X., J.M. Decker, H. Liu, Z. Zhang, R.B. Arani, J.M. Kilby, M.S. Saag, X. Wu, G.M. Shaw, and J.C. Kappes. 2002. Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob. Agents Chemother*. 46:1896–1905.
53. Mascola, J.R., P. D'Souza, P. Gilbert, B.H. Hahn, N.L. Haigwood, L. Morris, C.J. Petropoulos, V.R. Polonis, M. Sarzotti, and D.C. Montefiori. 2005. Recommendations for the design and use of standard virus panels to assess neutralizing antibody responses elicited by candidate human immunodeficiency virus type 1 vaccines. *J. Virol*. 79:10103–10107.
54. Li, M., F. Gao, J.R. Mascola, L. Stamatatos, V.R. Polonis, M. Koutsoukos, G. Voss, P. Goepfert, P. Gilbert, K.M. Greene, et al. 2005. Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J. Virol*. 79:10108–10125.

Downloaded from http://rupress.org/jem/article-pdf/206/6/1273/1910479/jem_20090378.pdf by guest on 09 February 2026