Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection

Brandon F. Keele[a], Elena E. Giorgi[b,c], Jesus F. Salazar-Gonzalez[a], Julie M. Decker[a], Kimmy T. Pham[a], Maria G. Salazar[a], Chuanxi Sun[a], Truman Grayson[a], Shuyi Wang[a], Hui Li[a], Xiping Wei[a], Chunlai Jiang[d], Jennifer L. Kirchherr[d], Feng Gao[d], Jeffery A. Anderson[e], Li-Hua Ping[f], Ronald Swanstrom[f], Georgia D. Tomaras[g], William A. Blattner[h], Paul A. Goepfert[a], J. Michael Kilby[a], Michael S. Saag[a], Eric L. Delwart[i], Michael P. Busch[i], Myron S. Cohen[e], David C. Montefiori[g] Barton F. Haynes[d], Brian Gaschen[b], Gayathri S. Athreya[b], Ha Y. Lee[j], Natasha Wood[k], Cathal Seoighe[k], Alan S. Perelson[b], Tanmoy Bhattacharya[b,l], Bette T. Korber[b,l], Beatrice H. Hahn[a,m], and George M. Shaw[a,m,n]

Departments of [a]Medicine and [m]Microbiology, University of Alabama at Birmingham, Birmingham, AL 35223; [b]Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM 87545; [c]Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01002; Departments of [d]Medicine and [g]Surgery, Duke University Medical Center, Durham, NC 27710; Departments of [e]Medicine and [f]Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC 27599; [h]Department of Epidemiology, University of Maryland, College Park, MD 20742; [i]Blood Systems Research Institute and Department of Laboratory Medicine, University of California, San Francisco, CA 94118; [j]Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642; [k]Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Rondebosch 7701, South Africa; and [l]Santa Fe Institute, Santa Fe, NM 87501

Communicated by Robert C. Gallo, University of Maryland, Baltimore, MD, March 5, 2008 (received for review December 20, 2007)

The precise identification of the HIV-1 envelope glycoprotein (Env) responsible for productive clinical infection could be instrumental in elucidating the molecular basis of HIV-1 transmission and in designing effective vaccines. Here, we developed a mathematical model of random viral evolution and, together with phylogenetic tree construction, used it to analyze 3,449 complete *env* sequences derived by single genome amplification from 102 subjects with acute HIV-1 (clade B) infection. Viral *env* genes evolving from individual transmitted or founder viruses generally exhibited a Poisson distribution of mutations and star-like phylogeny, which coalesced to an inferred consensus sequence at or near the estimated time of virus transmission. Overall, 78 of 102 subjects had evidence of productive clinical infection by a single virus, and 24 others had evidence of productive clinical infection by a minimum of two to five viruses. Phenotypic analysis of transmitted or early founder Envs revealed a consistent pattern of CCR5 dependence, masking of coreceptor binding regions, and equivalent or modestly enhanced resistance to the fusion inhibitor T1249 and broadly neutralizing antibodies compared with Envs from chronically infected subjects. Low multiplicity infection and limited viral evolution preceding peak viremia suggest a finite window of potential vulnerability of HIV-1 to vaccine-elicited immune responses, although phenotypic properties of transmitted Envs pose a formidable defense.

HIV-1 vaccines | transmitted HIV-1 envelope | viral evolution | virus transmission

HIV-1 transmission in humans results most commonly from virus exposure at mucosal surfaces (1). For practical reasons, it has been impossible to identify and characterize by direct analytical methods HIV-1 at or near the moment of transmission, yet it is this virus that antibody or cell-based vaccines must interdict. An important step in achieving a molecular understanding of HIV-1 transmission and potentially in the development of an effective HIV/AIDS vaccine is an accurate and precise description of the transmitted or early founder virus (or viruses) and sequences evolving from them during the critical period leading to productive clinical infection.

Previous reports examined the molecular basis of HIV-1 transmission by analyzing the genetic composition of viruses sampled between 1 and 6 months after infection (2–14). A common methodological approach in these studies was to derive viral sequences from the plasma or peripheral blood mononuclear cells of patients by bulk or near-limiting dilution PCR amplification of viral nucleic acid [proviral DNA or viral (v)RNA], followed by cloning, sequencing, and phylogenetic analysis (2–12). Alternatively, bulk amplified viral nucleic acids were analyzed by heteroduplex tracking assay (HTA) where only a small fraction of the gene of interest is interrogated by selective annealing of a short oligonucleotide probe followed by differential migration of the heteroduplex (5–7, 10). Although these approaches provide an approximation of the complexity of virus populations in acute and early infection, they have significant limitations. HTA, for example, does not provide sequence information for phylogenetic analysis and, as a consequence, allows for only qualitative inferences regarding the genetic identity and complexity of virus populations. Bulk or near-endpoint PCR, followed by cloning and sequencing, is compromised by the introduction of *Taq* polymerase errors (15–17), *Taq* polymerase-mediated template switching (recombination) (17–19), and nonproportional representation of target sequences attributable to template resampling or unequal template amplification and cloning (15–17, 20). Based largely on these approaches, previous studies generally have described the virus quasispecies in acute and early infection either as "homogeneous," reflecting transmission of one or few viruses, or "heterogeneous," reflecting a higher multiplicity of infection (2–11). It even has been suggested that HIV-1 infection commonly results from transmission and early replication of multiple virus variants that subsequently undergo a process of homogenization or purifying selection, giving rise to the appearance of a more homogeneous infection (7).

The objective of the present study was to develop and implement an experimental strategy that would enable us to identify unambiguously the transmitted or early founder *env* genes of viruses

Author contributions: B.F.K. and E.E.G. contributed equally to this work; B.F.K., J.F.S.-G., J.M.D., K.T.P., M.G.S., C. Sun, T.G., S.W., H.L., X.W., C.J., J.L.K., F.G., J.A.A., L.-H.P., R.S., G.D.T., W.A.B., P.A.G., J.M.K., M.S.S., E.L.D., M.P.B., M.S.C., D.C.M., B.F.H., B.H.H., and G.M.S. performed research; E.E.G., B.G., G.S.A., H.Y.L., N.W., C. Seoighe, A.S.P., T.B., and B.T.K. contributed new reagents/analytic tools; and B.F.K., E.E.G., R.S., B.F.H., C. Seoighe, A.S.P., T.B., B.T.K., B.H.H., and G.M.S. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. EU574937–EU579293).

[n]To whom correspondence should be addressed. E-mail: gshaw@uab.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0802203105/DCSupplemental.

responsible for establishing productive HIV-1 infection, to track their evolution in the critical period between transmission, peak viremia and seroconversion, and to evaluate their phenotypic properties. Essential to this strategy were two findings. First was the demonstration by Leigh Brown and coworkers (15), Mullins and coworkers (19), Coffin and coworkers (16), and Hahn and coworkers (17) that single genome amplification (SGA) of HIV-1 plasma vRNA followed by direct sequencing of uncloned amplicon DNA precludes *Taq*-induced nucleotide misincorporation and recombination in finished sequences. Second were the observations by Ho and coworkers (21, 22) and Shaw and coworkers (23, 24) that, because of the extremely short *in vivo* lifespan of plasma virus and of productively infected cells ($t_{1/2} < 1$ day), analysis of plasma vRNA could provide a uniquely informative view of HIV-1 replication dynamics and evolution. Thus, we hypothesized that an SGA-based analysis of plasma vRNA obtained from acutely infected individuals in the earliest stages of infection, and evaluated within the context of a model of random viral evolution, would allow us to infer the nucleotide sequences of *env* genes of viruses responsible for establishing productive clinical infection weeks earlier.

Results

Mathematical Model. We first constructed a mathematical model of HIV-1 replication and diversification by using previously estimated parameters of HIV-1 generation time (2 days) (22) reproductive ratio (R_0, 6) (25), and reverse transcriptase (RT) error rate (2.16×10^{-5}) (26) and by assuming that the initial virus replicates exponentially infecting R_0 new cells at each generation and diversifying under a model of evolution that assumes no selection [see supporting information (SI) *Text*]. Under this model, viruses would be expected to exhibit a Poisson distribution of mutations and a star-like phylogeny (27). This model led us to pose two hypotheses concerning the transmitted or early founder virus: (*i*) the progeny of individual virus(es) that establish productive infection can be identified in early stages of HIV-1 infection as distinct genetic lineages with low sequence diversity; and (*ii*) the consensus sequence of each *env* lineage sampled before the onset of immune selection corresponds to the actual *env* sequence of transmitted or founder virus (or viruses) responsible for establishing productive clinical infection.

HIV-1 Envelope Sequence Diversity. We tested these hypotheses by sequencing and analyzing 3,449 full-length *env* genes from plasma vRNA from 102 HIV-1 clade B-infected subjects whom we staged according to the Fiebig classification (28) (Fig. 1*A* and Dataset S1, Dataset S2, Dataset S3, and Dataset S4). Fifty-four of the subjects were regular donors of source plasma for whom serial specimens were available for analysis. As part of routine blood-banking practice, these individuals were regularly questioned (and deferred) for homosexual encounters, sex for money, or i.v. drug use, and they were monitored for acquisition of blood-borne infectious agents (including HIV and hepatitis viruses) that could indicate such risk behaviors. Forty-three other subjects admitted to high risk heterosexual ($n = 23$) or homosexual ($n = 20$) encounters, and four had unknown risks. Only one subject admitted to i.v. exposure as a risk. Fifty-one subjects were viremic without detectable HIV-1 serum antibodies at the time of study (Fiebig stages I or II), 26 others had HIV-1 antibodies detectable by ELISA but negative (Fiebig stage III) or indeterminate (Fiebig stage IV) by Western blot (WB), and 23 had positive WB lacking p31 bands, indicating recent infection (Fiebig stage V). Two subjects were fully WB-positive (Fiebig stage VI). To ensure proportional representation of plasma vRNA and avoid *in vitro*-generated recombination events and *Taq* polymerase errors, we used SGA of plasma vRNA followed by direct sequence analysis of uncloned *env* amplicons (16, 17, 19). Sequences were excluded if the chromatogram revealed "double peaks," indicative of amplification from more than one template or early *Taq* polymerase error. The median number of *env* sequences analyzed per

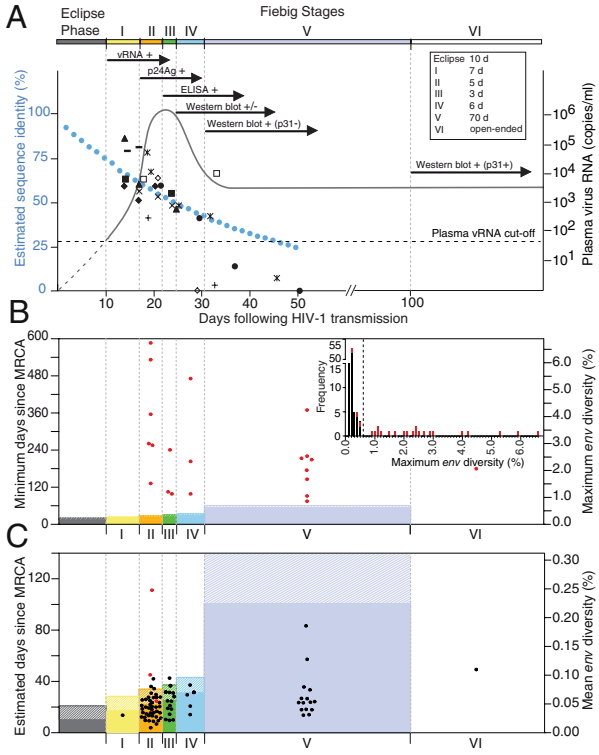

Fig. 1. HIV-1 *env* diversity in relation to Fiebig stage. (*A*) Fiebig stages (28) are defined by HIV-1 clinical laboratory test results. Blue symbols depict a Monte Carlo simulation of *env* sequence identity (%) after transmission (*SI Text*). Black symbols depict percentage identities in *env* sequences measured sequentially in 10 subjects, each represented by a different symbol. (*B*) Frequency distribution of within-patient maximum *env* diversities in all 102 subjects (*Inset*) and among the 21 subjects with more heterogeneous sequences. The vertical dashed line in the *Inset* represents the model prediction of maximum achievable *env* diversity (0.60%; 95% C.I. = 0.54–0.68%) by 100 days after transmission. Colored areas above each Fiebig stage represent model estimates of maximum *env* diversity (hatched area = 95% C.I.) and corresponding minimum days since a MRCA sequence. (*C*) Frequency distribution of within-patient mean *env* diversity and estimated days since MRCA in relation to Fiebig stage for 81 subjects with more homogeneous *env* sequences. Red dots in *B* and *C* and red bars in the *Inset* in *B* indicate subjects shown by phylogenetic analysis of *env* sequences to have been infected by two or more viruses.

subject was 25 (range, 10–203). The maximum within-patient *env* diversity ranged from 0.08% to 6.63%, with very low diversity found in 81 subjects (0.08 – 0.47%) and distinctly higher diversity found in 21 others (range 0.86–6.63%) (Fig. 1*B*, inset). We postulated that the observed differences in maximum *env* diversity may have been explained by differences in the numbers of viruses that infected these individuals (5, 10), and we formally tested this hypothesis by comparing model estimates for each subject (analyzed individually) of the minimum number of days that would be required to explain the observed within-patient *env* diversification from a single most recent common ancestor (MRCA) sequence. In this model, we do 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 would tend to be biased toward a low estimate. Each of 21 Fiebig stage II–VI subjects with more diverse viral sequences had minimum estimates for days since a MRCA virus that exceeded plausible values given their Fiebig stage (Fig. 1*B*). Conversely, all but one of 81 subjects with more homogeneous sequences had *env* diversities and estimated days since a MRCA that fell well within or near model predictions (Fig. 1*C*), suggesting that these individ-

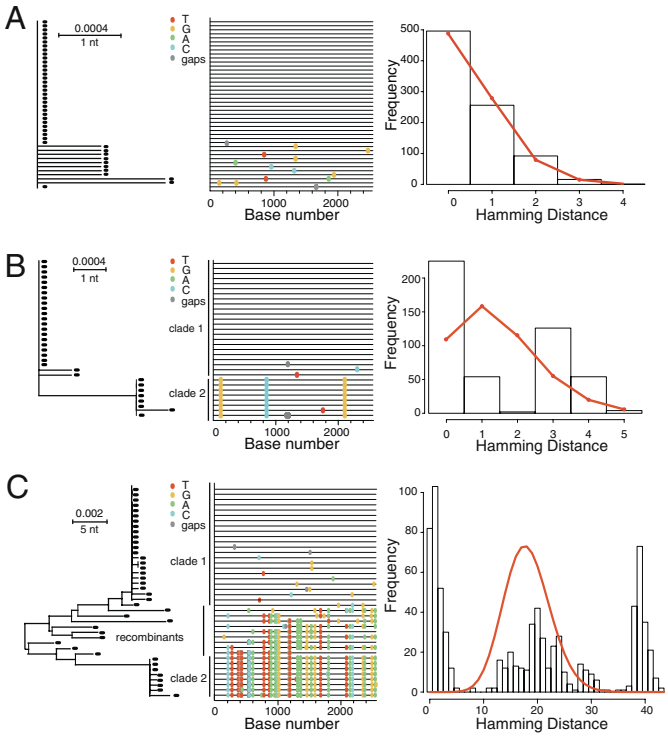

Fig. 2. NJ, *Highlighter*, and HD analyses of *env* diversity. (*A*) Subject 1006. NJ and *Highlighter* show infection by a single virus with HD frequencies, conforming precisely to model predictions of a single virus infection (red line). (*B*) Subject 6247. (*C*) Subject CAAN5342. NJ and *Highlighter* show infection by two closely (*B*) or distantly (*C*) related viruses (clades 1 and 2) with HD frequencies that do not conform to model predictions of a single virus infection (red line). Subject CAAN5342 has multiple sequences representing interlineage *env* recombination after transmission.

uals had productive clinical infections originating from a single virus or from more than one very closely related virus.

Identification and Enumeration of Transmitted or Early Founder Viruses. Sequences from all 102 subjects were analyzed by using neighbor-joining (NJ) phylogenetic tree methods together with a novel sequence visualization tool, *Highlighter* (www.HIV.lanl.gov), that allows tracing of common ancestry between sequences based on individual nucleotide polymorphisms. In 98 subjects, we identified one or more distinct, low diversity monophyletic *env* lineage. Each lineage contained a unique set of identical or near identical sequences. Seventy-eight of the 81 subjects with more homogeneous *env*s (Fig. 1*C*) had sequences that formed single lineages in NJ trees (Fig. 2*A* and Fig. S1). Three other subjects had sequences that exhibited low overall diversity but were composed of two distinct lineages distinguished by sets of three or four nucleotide polymorphisms (Fig. 2*B* and Fig. S2*A*). Model projections suggested that these three subjects were infected by very closely related viruses, most likely from a sexual partner who himself or herself was recently infected (29) as opposed to a single virus that evolved into two distinct lineages in the brief period preceding peak viremia (*SI Text*). Among the 21 subjects with greater *env* diversity (Fig. 1*B*), 17 had sequences represented by two to five discernible lineages, frequently accompanied by sequences showing interlineage recombination (Fig. 2*C* and Figs. S2–S4). We could infer that the recombination events resulting in most *env* chimeras that we identified occurred within the acutely infected patient after virus transmission (and not in the transmitting partner) because each recombinant contained long stretches of sequences that were identical or nearly identical to those of the principal transmitted viral lineages (Fig. 2*C* and Fig. S4). Four other subjects with heterogeneous infections had samples available only from later Fiebig stages, and in these individuals, discrete *env* lineages were not discernible because of larger numbers of nucleotide substitutions confounded by viral recombination. From the combined NJ and *Highlighter* analyses and modeling, we concluded that 78 of the 102 subjects (76%) had been productively infected by a single virus (or virus-infected cell) and 24 others (24%) had been infected by at least two to five infectious units. Among the latter subjects, 16 (67%) had clear evidence of recombination based on visual inspection of *Highlighter* tracings, which was confirmed by statistical analyses by using GARD or Recco recombination identification tools (*SI Text*).

Model Testing and Analysis of HIV-1 Evolution. To explore how *env* sequences sampled near peak viremia conformed to model assumptions, we first examined sequences from the 81 subjects with low *env* diversity. For each subject, we obtained 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. We then investigated whether or not the observed sequences evolved under a star-phylogeny model (i.e., all evolving sequences are equally likely and all coalesce at the founder) in the expected time frame based on Fiebig stage. Fifty-three (of 81) samples were consistent with both the Poisson model and a star phylogeny (Fig. 2*A*, Fig. S1, and Dataset S1). Among the samples that deviated in their mutational patterns from a Poisson distribution, we found 13 subjects had a striking overall enrichment for G-to-A hypermutation patterns with APOBEC3G/F signatures (30). In six of these 13 cases, the evidence for hypermutation was isolated to a single sequence (Fig. 3*A*), whereas in seven others, G-to-A hypermutation was evident across multiple sequences (Fig. 3*B*). When these APOBEC-mediated mutations were excluded from the analysis, a good fit to the Poisson model was restored (Fig. 3 and Dataset S2). Two other subjects could be resolved by excluding APOBEC signature patterns from the analysis despite the sample not being overtly hypermutated. The remaining 13 samples had apparent branching structures based on sublineages with a small number of shared mutations. Based on the temporal appearance and patterns of these mutations, we could explain these sublineages as resulting from transmission of closely related variants of a donor quasispecies (Fig. 2*B* and Fig. S2*A*), from stochastic mutations generated shortly after transmission (Fig. S5), or from HLA-restricted cytotoxic T lymphocyte (CTL) escape mutations that accumulated in patients sampled at later Fiebig stages (Fig. S6 and Dataset S3). Sequences from subjects with heterogeneous sequences resulting from infection by more than one virus violated model expectations for Poisson distribution and star phylogeny of mutations (Fig. 2*C* and Fig. S3) but conformed when identifiable *env* sublineages were analyzed individually (Dataset S4).

We also examined early virus diversification directly by studying 10 subjects sampled longitudinally (Dataset S5). This analysis included a total of 1,045 complete *env* sequences (median of 89 per subject; range, 57–203). Seven subjects had evidence of infection by one virus and three subjects by more than one virus. The model assumes that before the onset of immune selection, virus evolves randomly with the proportion of sequences identical to the transmitted virus(es) declining as depicted by the blue symbols in Fig. 1*A*. For subject 1058, whose plasma was sampled three times between Fiebig stages I and IV when viral loads were increasing from 2,737 to 26,162 to 550,000 RNA copies/ml, the proportion of identical viral sequences declined from 89% to 59% to 45% (Fig. 4), consistent with model projections. In all ten subjects, we found the proportion of identical *env* sequences to decline in a manner that closely approximated the model until Fiebig stages V-VI, when an abrupt decline in the proportion of identical sequences coincided

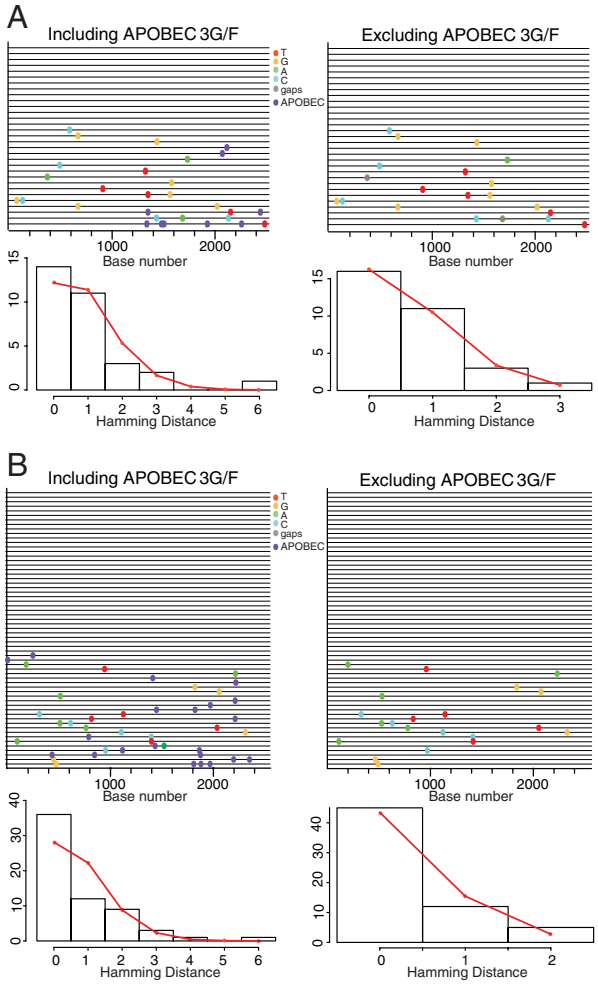

Fig. 3. Effect of APOBEC3G/F-mediated G-to-A hypermutation on HD frequency distribution. HD (*x* axis) is plotted versus frequency (*y* axis) as in Fig. 2. G-to-A hypermutations primarily in one sequence (*A*) or distributed in multiple sequences (*B*). Hypermutated sequences do not conform to model predictions for HD frequency distribution (red lines) but do conform if APOBEC3G/F related G-to-A mutations are eliminated.

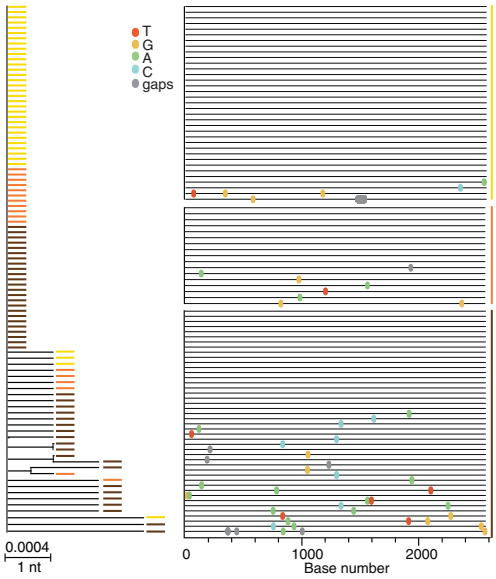

Fig. 4. NJ and *Highlighter* analysis of sequential *env* sequences from subject 1058. Sequences depicted by yellow, orange, and brown bars correspond to sample dates March 8, 1998 (Fiebig stage I), March 11, 1998 (Fiebig stage II), and March 18, 1998 (Fiebig stage IV) (Dataset S5). At these time points, plasma vRNA levels were 2,737 copies per ml, 26,162 copies per ml, and 550,000 copies per ml, respectively. The proportion of identical sequences corresponding to the transmitted or early founder *env* decreased from 89% to 59% to 45%, consistent with model projections illustrated by the blue symbols in Fig. 1*A*.

with selection for CTL escape mutants. This is illustrated for subject WEAU0575 (Fig. S6), where 29 of 30 *env* sequences mutated in a single confirmed HLA-restricted CTL epitope over a period of 29 days as the patient progressed from Fiebig stage II to V. Importantly, in none of these 10 individuals did we find evidence of a transmitted virus lineage that was lost during the acute infection period before peak viremia, nor did we find evidence, aside from CTL escape variants, of a predominant viral lineage that appeared *de novo*.

Although the empirical results suggested that HIV-1 *env* sequences sampled near peak viremia coalesce to virus(es) at or near transmission, we considered alternate explanations. We first considered limitations imposed by virus sampling. With a sample of at least 20 plasma vRNA sequences (which was the case for 81 of 102 subjects), we could be 95% confident that a given missed variant comprised less than 15% of the virus population (*SI Text*). Sampling biases were further minimized by sequential analyses in 10 subjects and by additional SGAs by using different primer sets that amplified *env*, a 7kb *gag–env* fragment, or the complete 9-kb viral genome (G.M.S., unpublished data), all of which gave identical results. Thus, sampling biases are unlikely to affect substantially our conclusions regarding the minimum number or identity of the viral lineages that established productive clinical infection. We also considered the possibility that *env* sequences sampled near the time of peak viremia might not coalesce to a transmitted or founder virus but instead to a more recent common ancestor that evolved from this virus. Several lines of evidence suggest this was not the case. First, in all but five subjects, the estimated time to the MRCA (±95% C.I.) of sequences analyzed by Poisson and BEAST models overlapped the estimated durations of infection based on the Fiebig stages (Dataset S1, Dataset S2, Dataset S3, and Dataset S4); shorter estimates of time to MRCA in three of these five subjects could be explained by overcompensation for G-to-A hypermutation. Second, the frequency of HIV-1 RT-mediated nucleotide misincorporation affecting a gene the length of *env* (≈2,600 bp) in a single infection cycle is small ($2 \times 10^{-5} \times 2{,}600 = 0.05$, or one *env* mutation in every 20 virus infection events), and most mutations would be expected to be neutral or deleterious. We found evidence for the latter in a statistically significant trend for lower than expected dN/dS ratios in *env* genes of viruses from Fiebig stage I and II subjects ($P < 0.01$ by Wilcoxon signed rank test with continuity correction). Third, the relatively long HIV-1 generation time (≈2 days), low R_0 (<10), and brief eclipse period (≈10 days) provide little time or opportunity for generation and outgrowth of a selected variant, a conclusion supported by model projections (*SI Text*). We recognize that virus recombination at the initial infection event is possible, but this would have negligible impact if vRNA is homozygous; if the transmitted virus was heterozygous (heterodimeric) and recombination did occur, then the progeny from that cell would represent the founder virus just one step removed from the transmitted virus. Finally, we used SGA to obtain 908 full-length *env* sequences from 43 chronically infected, treatment-naïve subjects (Dataset S6). Maximum within-subject *env* diversity among the chronic subjects ranged from 0.77 to 5.4%, essentially overlapping the range of HDs found in the samples from the 21 acute patients infected by more than one divergent virus (Fig. 1*B*). In none of the chronically infected subjects did we find predominant low diversity lineages comparable to those found in acutely infected subjects. Thus, we conclude that in the 98 subjects in whom we identified discrete low

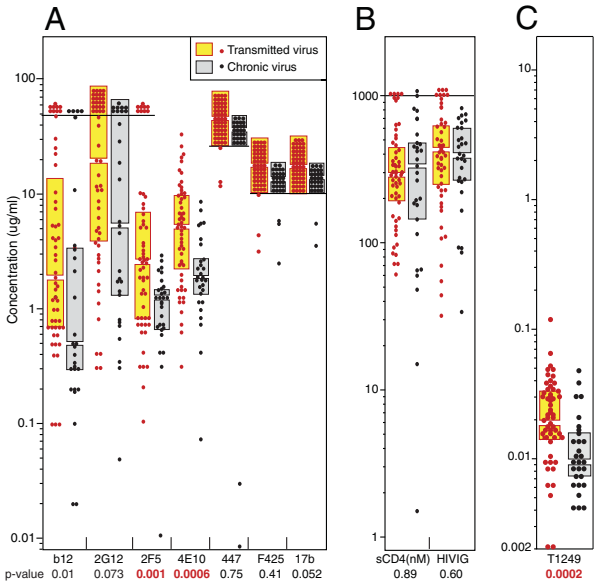

Fig. 5. Neutralization sensitivity of transmitted Envs compared with chronic Envs. Red dots represent the IC_{50} of the transmitted viral Envs, and yellow bars represent the interquartile range, with the median value indicated as the gap between the bars. Black dots and gray bars represent chronic Envs. The dots clustered above the thin black lines are Envs that were not neutralized at the highest concentration of ligand tested. Results for anti-Env mAbs (*A*), sCD4 and HIVIG (*B*), and the fusion inhibitor T1249 (*C*) are depicted. Uncorrected *P* values at the base of the figure were calculated by using a Wilcoxon rank-sum statistic comparing transmitted and chronic Envs. A total of 20 comparisons were made (Dataset S7); hence, to be significant after consideration of multiple tests, a significance level of 0.0025 was required. *P* values of <0.0025 are in red.

diversity viral *env* lineages, these lineages most often coalesce to *env* sequences of actual transmitted viruses. In instances where they do not, then they coalesce to early founder viruses only one or few replication steps removed from the transmitted virus.

Phenotypic Analysis of Transmitted or Early Founder Envs. The identification of viral *env* genes responsible for productive clinical infection enabled us to examine the phenotypic properties of Env glycoproteins most relevant to natural virus infection. We tested 55 Envs from 45 acutely infected subjects compared with 29 SGA-derived Envs from 13 chronically infected clade B patients for their coreceptor usage and sensitivity to human mAbs specific for the HIV-1 Env CD4 receptor and coreceptor binding sites (b12 and 17b, respectively); surface glycans (2G12); V3 loop (447-52D, F425-B4e8); membrane proximal external region of gp41 (2F5, 4E10, Z13e1); a polyclonal human HIV immune globulin (HIVIG); soluble CD4 (sCD4); the fusion inhibitor T1249; coreceptor inhibitors TAK779 (CCR5) and AMD3100 (CXCR4); and the anti-CD4 mAb RPA-T4. Each of the 55 Envs corresponding to transmitted or early founder viruses was biologically functional in mediating CD4 dependent virus entry. Fifty-four Envs were CCR5 tropic, and one was CCR5/X4 dual-tropic (Dataset S7), indicating that the R5 phenotype is a property of the transmitted virus *per se* and not one that evolves during acute or early infection (14, 31). All transmitted or founder Envs were sensitive to one or more of the broadly neutralizing antibodies IgG1b12, 2G12, 4E10 or 2F5 and most were neutralized by HIVIG at concentrations less than 1,000 μg/ml (equivalent to a 1:10 plasma dilution) (Fig. 5, Fig. S7, and Dataset S7). Transmitted or founder Envs were uniformly resistant to 17b, and all but three were resistant to the V3 loop mAbs 447-52D and F425-B4e8. Evidence that V3 epitopes in some Envs were conformationally masked and not absent came from experiments showing that sCD4 at subinhibitory concentrations (equal to the IC_{50} of sCD4 for each virus) could trigger a conformational change in gp120 and render the virus susceptible to neutralization by 447-52D and/or F425-B4e8. As a group, transmitted Envs were different from chronic Envs in that three chronic Envs were X4-dependent and two were CCR5/X4 dual-tropic, whereas only 1 of 55 transmitted Envs was R5/X4 tropic (Fisher exact test, $P = 0.017$). In addition, Envs from three chronically infected subjects were unusually sensitive to sCD4 and to the mAb 17b and to certain other gp120 or gp41 ligands; such globally heightened neutralization sensitivity was not seen in any of the transmitted Envs. Finally, we observed evidence of a modest (2- to 4-fold) heightened resistance of acute Envs compared with chronic Envs to T1249, 4E10, and 2F5 but not to HIVIG or other Env ligands (Fig. 5 and Dataset S7). We note that some of the Envs that we studied were derived from the same patients and thus were not independent. Trends for greater neutralization resistance in transmitted Envs were still observed after using a resampling scheme incorporating a single sequence from each patient for the ligands T1249, 4E10, and 2F5 (*SI Text*). These findings are different from those reported by Rusert *et al.* (13), possibly owing to the timing of sample acquisition or the effects of *in vitro* virus cultivation on HIV-1 Env receptor binding in the latter study. We caution that our study was not designed prospectively or powered statistically to make a robust comparison of the sensitivity of transmitted/founder Envs versus chronic Envs to broadly neutralizing envelope ligands, and thus these findings should be considered exploratory.

Discussion

The present study builds on a large body of published work on acute HIV-1 infection (2–14) by shedding new light on the genetic and phenotypic properties of HIV-1 at or near the moment of transmission and in the critical period of virus replication and diversification leading to peak viremia and antibody seroconversion. Among 102 subjects, we found that 78 (76%) had evidence of infection by a single virus or virus-infected cell and that 24 others (24%) had evidence of infection by at least two to five viruses. Aside from early selection of CTL escape variants found in several subjects, there was no suggestion of virus adaptation to a more replicative variant or bottlenecking in virus diversity (7) preceding peak viremia. Our findings regarding the number of viruses leading to productive clinical infection are minimal estimates, and additional viruses could conceivably have been transmitted but not sufficiently propagated *in vivo* to allow detection within the scope or timing of our sampling. We note, however, that our observation of a low number of transmitted viruses (range, 1–5; median 1) is consistent with epidemiological observations of the relative inefficiency of virus transmission by most sexual routes (1, 29). These considerations notwithstanding, we interpret the findings of low multiplicity infection and limited viral evolution preceding peak viremia to suggest a crucial but finite window of potential vulnerability of HIV-1 to vaccine-elicited immune responses.

Previous studies characterized HIV-1 *env* sequences in acute and early infection by using methods that did not involve SGA and direct sequencing, and, as a result, they did not allow for the precise identification or enumeration of transmitted or early founder *env*s, their pathways of diversification as genetic units, or their corresponding phenotypes (2–14). The extent to which differences in experimental approach can impact the interpretation of virus complexity was recently highlighted in a study of acute and early HIV-1 clade C infection (17) and was seen again in the present study in 10 subjects who had been analyzed previously by the HTA method (10). In this latter example, SGA and HTA methods were concordant in identifying five acute subjects (Z02, Z05, Z13, Z20, and Z27) infected by a single virus but were discordant in defining the number and identity of transmitted viruses in five others (Z30, Z18, Z03, Z10, and Z29). Reanalysis of all 26 subjects in that study (10) led to the conclusion that 80% of subjects (not 50%, as

suggested by HTA) had been productively infected by a single virus (G. Schnell and R.S., unpublished observations), a finding more in keeping with the present report.

Differences in risk behaviors, routes of virus transmission, clinical stage and viral load in the infected partner, and comorbid clinical conditions influence the frequency of HIV-1 transmission (1, 29) and likely could affect the numbers of viruses transmitted and subsequent disease natural history. All are factors relevant to vaccine design and evaluation. In the present study, a substantial proportion of the subjects had limited behavioral information available for analysis; therefore, we could draw no firm conclusions regarding particular risk factors and the likelihood of single versus multiple virus transmission. However, we did note that among men who acknowledged sex with men (MSM), half had been acutely infected by more than one virus strain from their HIV-infected partner. This was also true for the one injecting drug user who acknowledged sharing needles with an HIV-positive partner. We are currently examining this question of risk factors for multiple virus acquisition in additional well-defined acute infection cohorts by using SGA-based strategies.

Our finding in 98 of 102 subjects that the SGA-direct sequencing approach could unambiguously identify transmitted or early founder *env*s provided an unparalleled opportunity to characterize those envelope genes and proteins most relevant to virus transmission and vaccine development. When such *envs* were cloned, expressed, and analyzed phenotypically, we found all to be biologically functional in conferring CD4-dependent and CCR5-dependent cell entry. We explored the possibility that viruses responsible for productive clinical infection in naïve individuals who are devoid of preexisting neutralizing antibodies may rapidly adapt (or be preselected) for preferential replication in this environment, but we found no genetic or phenotypic evidence for this. To the contrary, most viral *env* sequences coalesced to a MRCA near the time estimated for transmission based on Fiebig staging; coreceptor binding surfaces on transmitted/founder Envs (probed by 17b, 447-52D, and F425-B4e8) were conformationally masked; and transmitted/founder Envs were comparable to primary virus strains (32) in their susceptibility to broadly neutralizing antibodies. Studies aimed at further analyzing Env glycoproteins of transmitted or early founder viruses may help to identify unique features and potential vulnerabilities relevant to vaccine design. Beyond this, the present study illustrates a strategy for identifying and characterizing full-length genomes and proteomes of transmitted or early founder viruses including, but not limited to, HIV-1. Such analyses may facilitate a better understanding of virus natural history and virus-specific cellular and humoral immune responses in naïve and vaccinated individuals.

Methods

Plasma HIV-1 RNA sequences were derived by SGA-direct amplicon sequencing (17) and can be accessed at GenBank (EU574937–EU579293) with edited *env* alignments at www.hiv.lanl.gov/content/sequence/hiv/user_alignments/keele. A model of random virus diversification and power calculations for estimating the likelihood of detecting infrequent transmitted *env* variants are presented in *SI Text* (see also Figs. S8 and S9). *Env* genes were molecularly cloned and analyzed by Env pseudotype cell entry assay (24).

ACKNOWLEDGMENTS. We thank D. Burton and M. Zwick (Scripps Research Institute, La Jolla, CA), J. Mascola (National Institutes of Health, Bethesda, MD), S. Zolla-Pazner (New York University, New York), H. Katinger (Polymun Scientific, Vienna, Austria), L. Cavacini (Harvard University, Boston, MA) and J. Robinson (Tulane University, New Orleans, LA), for monoclonal antibodies; P. Sharp and J. Sodroski for critical input; D. McPherson, Y. Chen, and B. Cochran for technical assistance; the clinical core of the Center for HIV/AIDS Vaccine Immunology; and J. White for manuscript preparation. This work was supported by the Center for HIV/AIDS Vaccine Immunology; National Institutes of Health Grants AI67854, AI61734, AI27767, AI50410, AI64518, and AI41530; and Bill and Melinda Gates Foundation Grants 37874 and 38619.

1. Shattock RJ, Moore JP (2003) Inhibiting sexual transmission of HIV-1 infection. *Nat Rev Microbiol* 1:25–34.
2. Wolfs TF, Zwart G, Bakker M, Goudsmit J (1992) HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. *Virology* 189:103–110.
3. Wolinsky SM, *et al*. (1992) Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. *Science* 255:1134–1137.
4. Zhu T, *et al*. (1993) Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science* 261:1179–1181.
5. Poss M, *et al*. (1995) Diversity in virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1. *J Virol* 69:8118–8122.
6. Long EM, *et al*. (2000) Gender differences in HIV-1 diversity at time of infection. *Nat Med* 6:71–75.
7. Learn GH, *et al*. (2002) Virus population homogenization following acute human immunodeficiency virus type 1 infection. *J Virol* 76:11953–11959.
8. Grobler J, *et al*. (2004) Incidence of HIV-1 dual infection and its association with increased viral load set point in a cohort of HIV-1 subtype C-infected female sex workers. *J Infect Dis* 190:1355–1359.
9. Derdeyn CA, *et al*. (2004) Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* 303:2019–2022.
10. Ritola K, *et al*. (2004) Multiple V1/V2 env variants are frequently present during primary infection with human immunodeficiency virus type 1. *J Virol* 78:11208–11218.
11. Sagar M, *et al*. (2004) Human immunodeficiency virus type 1 (HIV-1) diversity at time of infection is not restricted to certain risk groups or specific HIV-1 subtypes. *J Virol* 78:7279–7283.
12. Frost SD, *et al*. (2005) Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. *Proc Natl Acad Sci USA* 102:18514–18519.
13. Rusert P, *et al*. (2005) Virus isolates during acute and chronic human immunodeficiency virus type 1 infection show distinct patterns of sensitivity to entry inhibitors. *J Virol* 79:8454–8469.
14. Margolis L, Shattock R (2006) Selective transmission of CCR5-utilizing HIV-1: The 'gatekeeper' problem resolved? *Nat Rev Microbiol* 4:312–317.
15. Simmonds P, Balfe P, Ludlam CA, Bishop JO, Leigh Brown AJ (1990) Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J Virol* 64:5840–5850.
16. Palmer S, *et al*. (2005) Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J Clin Microbiol* 43:406–413.
17. Salazar-Gonzalez JF, *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.
18. Meyerhans A, Vartanian JP, Wain-Hobson S (1990) DNA recombination during PCR. *Nucleic Acids Res* 18:1687–1691.
19. Shriner D, Rodrigo AG, Nickle DC, Mullins JI (2004) Pervasive genomic recombination of HIV-1 *in vivo*. *Genetics* 167:1573–1583.
20. Liu SL, *et al*. (1996) HIV quasispecies and resampling. *Science* 273:415–416.
21. Ho DD, *et al*. (1995) Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 373:123–126.
22. Markowitz M, *et al*. (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.
23. Wei X, *et al*. (1995) Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 373:117–122.
24. Wei X, *et al*. (2003) Antibody neutralization and escape by HIV-1. *Nature* 422:307–312.
25. Stafford MA, *et al*. (2000) Modeling plasma virus concentration during primary HIV infection. *J Theor Biol* 203:285–301.
26. Mansky LM, Temin HM (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.
27. Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129:555–562.
28. Fiebig EW, *et al*. (2003) Dynamics of HIV viremia and antibody seroconversion in plasma donors: Implications for diagnosis and staging of primary HIV infection. *AIDS* 17:1871–1879.
29. Wawer MJ, *et al*. (2005) Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J Infect Dis* 191:1403–1409.
30. Simon V, *et al*. (2005) Natural variation in Vif: Differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. *PLoS Pathog* 1:e6.
31. Hladik F, *et al*. (2007) Initial events in establishing vaginal entry and infection by human immunodeficiency virus type-1. *Immunity* 26:257–270.
32. Li M, *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.