JOURNAL OF VIROLOGY, Mar. 2011, p. 2751–2763 Vol. 85, No. 6
0022-538X/11/$12.00 doi:10.1128/JVI.02316-10

Origin and Evolution of HIV-1 in Breast Milk Determined by Single-Genome Amplification and Sequencing▽

Jesus F. Salazar-Gonzalez,[1]† Maria G. Salazar,[1]† Gerald H. Learn,[1] Genevieve G. Fouda,[2] Helen H. Kang,[3] Tatenda Mahlokozera,[3] Andrew B. Wilks,[3] Rachel V. Lovingood,[2] Andrea Stacey,[4] Linda Kalilani,[5] Steve R. Meshnick,[6] Persephone Borrow,[4] David C. Montefiori,[2] Thomas N. Denny,[2] Norman L. Letvin,[3] George M. Shaw,[1] Beatrice H. Hahn,[1] Sallie R. Permar,[3,7]* and Center for HIV/AIDS Vaccine Immunology A0167854

Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama 35294[1]; Duke Human Vaccine Institute, Duke University, Durham, North Carolina 27710[2]; Division of Viral Pathogenesis, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215[3]; Nuffield Department of Clinical Medicine, University of Oxford, The Jenner Institute, Compton, Newbury, Berkshire RG20 7NN, England[4]; College of Medicine, University of Malawi, Blantyre, Malawi[5]; Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill, North Carolina, 27599[6]; and Children's Hospital Boston, Harvard Medical School, Boston, Massachusetts 02115[7]

Received 5 November 2010/Accepted 17 December 2010

HIV transmission via breastfeeding accounts for a considerable proportion of infant HIV acquisition. However, the origin and evolution of the virus population in breast milk, the likely reservoir of transmitted virus variants, are not well characterized. In this study, HIV envelope (*env*) genes were sequenced from virus variants amplified by single-genome amplification from plasmas and milk of 12 chronically HIV-infected, lactating Malawian women. Maximum likelihood trees and statistical tests of compartmentalization revealed interspersion of plasma and milk HIV *env* sequences in the majority of subjects, indicating limited or no compartmentalization of milk virus variants. However, phylogenetic tree analysis further revealed monotypic virus variants that were significantly more frequent in milk (median proportion of identical viruses, 29.5%; range, 0 to 61%) than in plasma (median proportion of identical viruses, 0%; range, 0 to 26%) ($P = 0.002$), suggesting local virus replication in the breast milk compartment. Moreover, clonally amplified virus *env* genes in milk produced functional virus Envs that were all CCR5 tropic. Milk and plasma virus Envs had similar predicted phenotypes and neutralization sensitivities to broadly neutralizing antibodies in both transmitting and nontransmitting mothers. Finally, phylogenetic comparison of longitudinal milk and plasma virus *env* sequences revealed synchronous virus evolution and new clonal amplification of evolved virus *env* genes in milk. The limited compartmentalization and the clonal amplification of evolving, functional viruses in milk indicate continual seeding of the mammary gland by blood virus variants, followed by transient local replication of these variants in the breast milk compartment.

Breast milk transmission of human immunodeficiency virus (HIV) remains a significant mode of infant HIV acquisition globally. This mode of mother-to-child HIV transmission accounts for nearly half of the new infant HIV infections occurring in developing nations (36), where formula feeding is associated with increased morbidity and mortality from respiratory and diarrheal illnesses (51). Maternal (8, 24, 34, 48) and infant (8, 23, 27, 33) antiretroviral prophylaxis during breastfeeding is a promising strategy for reduction of breast milk transmission of HIV; however, implementation of this strategy is not widespread in areas of high HIV prevalence (33). Furthermore, the toxicities associated with provision of antiretroviral prophylaxis to mothers and infants during breastfeeding and the clinical impact of the development of antiretroviral-resistant viruses during breastfeeding prophylaxis are not well described for large populations.

While HIV transmission via breastfeeding accounts for a significant proportion of infant HIV acquisition, the mechanisms of this transmission and the pool of transmitted viruses in breast milk are not well characterized. The magnitude of the milk virus load has been associated with the risk of infant HIV acquisition via breastfeeding (18, 45, 47), and therefore viruses present in milk are the likely pool of viruses transmitted via breastfeeding. Notably, milk HIV RNA loads remain 1 to 2 log lower than those in plasma (18, 40), suggesting a lack of free virus mixing between these compartments. Defining the origin and genetic composition of viruses present in breast milk has important implications for strategies to prevent breast milk virus transmission.

Studies of HIV populations present in mucosal compartments suggest that virus variants in the genital tract may be distinct from the systemic virus population (2, 5, 21, 22). Divergent mucosal immune responses may shape compartment-specific viruses, as described for anatomic compartments such

* Corresponding author. Mailing address: Division of Viral Pathogenesis, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, E/CLS1028, Boston, MA 02115. Phone: (617) 735-4403. Fax: (617) 735-4527. E-mail: sallie.permar@childrens.harvard.edu.

† Jesus F. Salazar-Gonzalez and Maria G. Salazar contributed equally to this report.

▽ Published ahead of print on 29 December 2010.

as semen and cervicovaginal fluid (2, 6, 9, 11, 21, 22, 37). Early investigations of whether breast milk virus variants were compartmentalized from those in plasma were contradictory (5, 16). However, recent studies of HIV-infected lactating women (13, 15, 16) and simian immunodeficiency virus (SIV)-infected lactating monkeys (38) have suggested that the virus present in milk is not genetically distinct from that in plasma. In addition, recent human and primate studies of the virus population in milk reported that groups of identical or nearly identical viruses were frequently detected in milk, suggesting that a proportion of viruses present in milk are produced by infected cells which are resident in the mammary gland (13, 15, 38). However, the origin, persistence, and functionality of these monotypic viruses are unknown. Thus, the production site, virus trafficking patterns, and selection pressures that longitudinally shape the genetic composition of the transmitted pool of HIV variants remain to be characterized further.

In this study, full-length HIV *env* was amplified and sequenced from plasma and milk of HIV-infected, breast milk-transmitting and nontransmitting women by single-genome amplification, a method that reduces the possibility of resampling and PCR-associated error and virus recombination. Phylogenetic analysis and statistical assessment of compartmentalization were applied to the milk and plasma *env* sequences. Furthermore, HIV *env* sequences were cloned and used to make Env pseudoviruses for assessment of tropism and neutralization sensitivity. Phylogenetic and functional analyses of the full-length HIV *env* gene in these anatomically distinct compartments, both cross-sectionally and longitudinally, allow for insights into compartment-specific virus production and sequence evolution of the pool of virus variants present in the breast milk compartment.

MATERIALS AND METHODS

Subjects and sample collection and processing. Pregnant women testing HIV positive by a rapid antibody test were recruited into this study (CHAVI 009) from two rural health clinics outside Blantyre, Malawi. The study was approved by the College of Medicine Research and Ethics Committee in Malawi and by all participating institutions' institutional review boards. Women were enrolled at delivery if HIV infection was confirmed by HIV RNA and breastfeeding was initiated. All mothers and infants in this study received single-dose nevirapine at delivery. Maternal blood and left and right breast milk samples were collected 4 to 5 weeks and 3 months after delivery. Plasmas were isolated by Ficoll gradient centrifugation, and breast milk supernatants were isolated by centrifugation to separate the lipid and cellular portions. Twelve subjects who remained untreated (except for single-dose nevirapine at delivery) and had detectable milk virus loads (Table 1) were included in this substudy of HIV *env* diversity in breast milk and plasma. *env* amplicons were able to be generated from milk and plasma of each subject. None of the women included in the study had symptoms consistent with clinical mastitis; however, three women (subjects 0404, 1209, and 3404) had laboratory evidence of subclinical mastitis in the left breast, indicated by a milk supernatant sodium concentration of >12 mM and a sodium/potassium ratio of >1 (measured by a Gen2 ion-selective electrode on a Roche Cobas c501 platform [Roche Diagnostics]) (13, 47, 52). Three of the maternal subjects transmitted the virus to their infants via breastfeeding (subjects 0404, 1209, and 4403), as defined by an infant HIV DNA PCR result of negative at birth and 4 to 5 weeks but positive by 6 months of age. All infants were monitored every 3 months until the cessation of breastfeeding or 18 months of age.

Plasma and breast milk HIV loads. Plasma and milk virus loads were measured using the Roche Cobas Ampliprep/Cobas TaqMan 48 for HIV-1 viral load assay. Breast milk supernatant was diluted 1:5 in phosphate-buffered saline (PBS) prior to analysis. The minimum levels of detection for this assay are 48 viral RNA copies/ml of plasma and 240 viral RNA copies/ml of breast milk. The laboratory performing these assays was enrolled in the National Institute of Allergy and Infectious Diseases Division of AIDS Virology Quality Assessment Program and certified for HIV load determinations.

TABLE 1. Clinical characteristics of chronically HIV-infected, lactating Malawian subjects[b]

Subject	CD4 count (cells/μl)	Plasma virus load (copies/ml)	Milk virus load in left/right breast at 4 to 5 wk postpartum (milk virus load in left breast at 12 wk postpartum) (copies/ml)
0301	406	159,000	1,315/265
0404[a]	80	22,600	29,650/29,200
0601	295	346,000	478/1,290
0702	523	35,200	31,900/2,530
1209[a]	153	3,070	635/615 (460)
3009	273	202,000	770/489
3305	166	359,000	3,335/16,150
3404	548	72,400	102,500/860
3902	213	19,900	27,950/28,500 (1,095)
4403[a]	208	100,892	101,500/30,450 (<240)
4707	247	83,400	10,500/22,500
5807	188	138,000	2,760/1,970

[a] Breast milk-transmitting subject.
[b] Subjects with laboratory evidence of subclinical mastitis are indicated with shading.

Single-genome amplification and sequencing of the HIV *env* gene. Plasma aliquots containing approximately 10,000 RNA copies were extracted using a QIAamp Viral RNA Mini kit (Qiagen). Plasma samples with low viral loads and all breast milk supernatant 1-ml aliquots were concentrated by centrifugation at 23,600 × *g* for 1 h at 4°C, and the virus pellet in approximately 100 μl of leftover plasma or milk supernatant was used for the same RNA extraction procedure. Depending on the virus RNA load in milk (Table 1), between 200 and 10,000 RNA copies were reverse transcribed. Single-strand cDNA synthesis was performed by using the SuperScript III protocol according to the manufacturer's instructions (Invitrogen Life Technologies). RNA, deoxynucleoside triphosphates (0.5 mM [each]), and 0.25 μM primer OFM19 (5′-GCACTCAAGGCAAGCTTTATTGAGGCTTA-3′) (corresponding to nucleotides [nt] 9604 to 9632 of the HXB2 sequence) were incubated for 5 min at 65°C to denature the secondary structure of the RNA. First-strand cDNA synthesis was carried out in 20-μl to 60-μl reaction mixtures with 1× reverse transcriptase buffer containing 5 mM dithiothreitol, 2 U/μl of an RNase inhibitor (RNaseOut), and 10 U/μl SuperScript III. The reaction mixture was incubated at 50°C for 60 min, followed by an additional hour at 55°C. Next, reactions were heat inactivated at 70°C for 15 min, followed by RNase H digestion at 37°C for 20 min (Invitrogen Life Technologies). The resulting cDNA was used immediately for PCR or stored frozen at −80°C.

Full-length *env* genes were PCR amplified by single-genome amplification methods as previously described (20, 46). Briefly, cDNA was titrated by endpoint dilution (undiluted to 1:40 dilution for breast milk cDNA) in 96-well PCR plates to a concentration that yielded no more than 30% PCR-positive wells and conformed to a Poisson distribution of a single template per reaction mix. PCR conditions for first-round PCR consisted of 1× High Fidelity Platinum PCR buffer, 2 mM $MgSO_4$, a 0.2 mM concentration of each deoxynucleoside triphosphate, 0.2 μM (each) primers Vif1 (5′-GGGTTTATTACAGGGACAGCAGA G-3′; nt 4900 to 4923) and OFM19, and 0.025 U/μl Platinum *Taq* High Fidelity polymerase (Invitrogen) in a 20-μl reaction mixture. The following PCR conditions were used: 94°C for 2 min, followed by 35 cycles of 94°C for 15 s, 55°C for 30 s, and 68°C for 4 min, with a final extension of 68°C for 10 min. Second-round PCR was carried out using 2 μl of the first-round product and 0.2 μM (each) primers EnvA (5′-GGCTTAGGCATCTCCTATGGCAGGAAGAA-3′; nt 5954 to 5982) and EnvN (5′-CTGCCAATCAGGGAAGTAGCCTTGTGT-3′; nt 9145 to 9171) with the same PCR mixture as in the first round. The PCR conditions included the following: 94°C for 2 min, followed by 45 cycles of 94°C for 15 s, 55°C for 30 s, and 68°C for 4 min, with a final extension at 68°C for 10 min. The amplicons were sized in a precast 1% agarose E-gel 96 (Invitrogen Life Technologies). Amplicons derived from cDNA dilutions yielding less than 30% PCR positivity were sequenced using BigDye Terminator v.3.1 chemistry and the protocols recommended by the manufacturer (Applied Biosystems). The sequences were determined by using an ABI 3730xl DNA analyzer (Applied Biosystems) and were edited by using the Sequencher program, version 4.9 (Gene Codes). Both strands of DNA were sequenced. All chromatograms were carefully inspected for sites of ambiguous sequence (double peaks). A total of 702 *env* sequences were inferred to be derived from a single cDNA template; sequences

that displayed multiple double peaks in sequence chromatograms were excluded from further analyses.

Phylogenetic sequence analysis. Sequences showing significant evidence of APOBEC-driven G-to-A hypermutation, according to the Los Alamos National Laboratory (LANL) HIV Sequence Database tool Hypermut 2.0 (44; http://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html), were excluded from subsequent analyses. Sequence alignments were produced using ClustalW (29) and were subsequently adjusted to optimize codon alignment. Regions that could not be aligned unambiguously were omitted in further phylogenetic analyses but were included in all other genetic and phenotypic analyses of the virus populations. A neighbor-joining phylogenetic tree including sequences from all 12 subjects indicated that the sequences corresponding to each subject clustered with high bootstrap support (100% [not shown]) (29). Additionally, the sequences from each subject were included in an analysis with 500 subtype C *env* sequences to determine that in each case the subject sequences were monophyletic, to the exclusion of the other sequences (not shown). For each subject, a maximum likelihood phylogenetic tree was inferred using PhyML, version 3.0 (14), and the results of approximate likelihood ratio tests (≥0.95) were used to infer phylogenetic support (4). Recombination breakpoints in the sequence alignment for each subject were detected using GARD (25).

Calculation of the proportion of clonally amplified virus variants, genetic diversity, N-linked glycosylation site number, and variable region length. Clonally amplified virus *env* sequences were defined as viruses that were exactly identical to each other (0 nucleotide differences). Sequence diversity was calculated using PAUP* (50). Sequence variability plots were generated using the LANL Highlighter web tool (http://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter.html), using a plasma-derived sequence nearest to the center node of the phylogenetic tree as the master sequence to enhance detection of unique virus sequences or deletions in breast milk virus variants. Potential N-linked glycosylation sites were enumerated in amino acid sequences inferred from the nucleotide sequences (54). Amino acid assignments to gp120 variable (V) regions were based on the LANL HXB2 spreadsheet (30; http://www.hiv.lanl.gov/content/sequence/HIV/MAP/hxb2.xls) for determination of V region length.

HIV *env* cloning and pseudovirus production. HIV *env* genes amplified from milk and plasma were reamplified from the first-round PCR product by use of blunt-end primers and a low PCR cycle number to reduce PCR-induced sequence errors. Amplicon size was confirmed on an agarose gel, and the amplicon was gel purified according to the manufacturer's instructions (Qiagen). Purified amplicons were ligated into pcDNA3.1 directional TOPO vectors (Invitrogen) and used to transform TOP10 *Escherichia coli* cells per the manufacturer's instructions. Cultures were grown at 30°C, and selected colonies were screened by miniprep and restriction digestion with XhoI and BamHI (New England Biolabs). Plasmids with fragments of the correct size after restriction digestion were sequenced to confirm sequence identity to the original *env* amplicon. Confirmed *env* clones were grown in large cultures at 30°C. Plasmids were prepared by megapreps (Qiagen) and resequenced to again confirm sequence identity to the original *env* amplicon. Env pseudoviruses were prepared by transfection into 293T cells of 4 μg of *env* plasmid DNA and 8 μg of *env*-deficient HIV plasmid DNA, using the FuGENE 6 transfection reagent (Roche Diagnostic). Two days after transfection, the culture supernatant containing pseudoviruses was harvested, filtered, aliquoted, and stored at −80°C. An aliquot of stored pseudovirus was used to measure infectivity in TZM-bl cells. Serial fivefold dilutions of pseudovirus were distributed in quadruplicate in 96-well flat-bottom plates (Costar) in a total volume of 100 μl per well. Freshly trypsinized TZM-bl cells were then added (10,000 cells/well in Dulbecco's modified Eagle's medium [DMEM]–10% fetal bovine serum [FBS] containing HEPES and 25 μg/ml of DEAE-dextran). After 48 h of incubation, 100 μl of cells/well was transferred to a 96-well black solid plate (Costar), and the luminescence was measured using a Britelite Plus luminescence reporter gene assay system (Perkin-Elmer Life Sciences). Wells producing relative luminescence units of >3× the background were scored as positive, and the 50% tissue culture infective dose ($TCID_{50}$) was calculated by the method of Reed and Muench, as previously described (19, 31, 35). The concentration of the pseudoviruses was measured using a commercial p24 enzyme-linked immunosorbent assay (ELISA) (Perkin-Elmer Life Sciences) per its protocol. The number of infectious units of each pseudovirus preparation is reported as the $TCID_{50}$/ng of p24 as a measure of infectivity per unit of virus (26).

Coreceptor usage and neutralization phenotype of HIV Env pseudoviruses. Coreceptor usage was determined in TZM-bl cells by using TAK-779 (NIH HIV reagent repository), a CCR5 receptor inhibitor, and AMD 3100 (NIH HIV reagent repository), a CXCR4 antagonist. Freshly trypsinized TZM-bl cells were distributed into 96-well plates (10,000 cells/well in DMEM–10% FBS containing HEPES and 25 μg/ml of DEAE-dextran) and incubated for 1 hour with either TAK-779 (10 μM), AMD 3100 (1.3 μM), a combination of the two reagents, or no inhibitor (control wells). Pseudoviruses were diluted to a predetermined concentration based on the $TCID_{50}$ (a concentration yielding the equivalent of 150,000 relative light units or at least 10 times the background level). Diluted pseudoviruses were added to wells with the different treatments in triplicate, and the plates were incubated for 2 days. Cells were lysed, and luciferase units were measured. Wells containing coreceptor inhibitors were compared to control wells to determine if either agent led to a reduction in infectivity. As controls, viruses known to use either CCR5 or CXCR4 were included.

The neutralization phenotypes of the milk and plasma Env pseudoviruses were determined in TZM-bl cells. The reagents used included soluble CD4 (sCD4; Progenics, Tarrytown, NY); broadly neutralizing monoclonal antibodies 4E10, 2F5, 2G12 (all from Polymun Scientific, Inc., Vienna, Austria), and IgG1 B12 (Quality Biological, Gaithersburg, MD); HIV immunoglobulin (HIVIg) (obtained from NVITAL, Gaithersburg, MD); and 5 polyclonal IgGs purified from clade C HIV-infected sera. The HIV-infected sera were the top 5 neutralizers among a large panel of HIV-infected sera from the South African National Blood Services and were obtained from Lynn Morris (National Institute of Communicable Diseases, Johannesburg, South Africa). Neutralization of the Env-pseudotyped virus was measured as a reduction in luciferase gene expression, as previously described (35).

Statistical analysis. The proportions of genetically identical virus variants in milk and plasma for each subject were compared by Fisher's exact test with Prism, version 5. Furthermore, the proportions of clonally amplified viruses in milk and plasmas of all subjects were compared by the paired, nonparametric Wilcoxon signed rank test. Evidence of compartmental structure was assessed using programs implemented by HyPhy software (41), including the Slatkin-Maddison (SM) test, which compares the minimum number of intercompartmental migration events to the frequency distribution of migration events in 1,000 randomized trees (42, 49), and the nearest neighbor statistic (S_{nn}), which compares how often the nearest sequence pairs were isolated from the same or different compartments (17, 53) in 1,000 permutations. All other comparisons of intra- and intersubject plasma and virus *env* sequence variables and milk factors, such as genetic diversity, N-linked glycosylation site number, and cytokine concentration, were performed using the paired, nonparametric Wilcoxon signed rank test with Prism, version 5. Correlations of milk sodium concentration and cytokine concentration with breast milk virus phenotypes were performed using Spearman's rank test. Nonparametric statistics were employed due to the small number of subjects and virus variants that could be amplified from the milk of some subjects. Statistical significance was defined for P values of <0.05.

Nucleotide sequence accession numbers. All *env* sequences determined in this study were deposited in GenBank under accession numbers HM070449 to HM070824 and HQ595810 to HQ596189.

RESULTS

Limited phylogenetic compartmentalization of HIV variants from the systemic virus population in the breast milk. Full-length HIV *env* genes were amplified and sequenced from plasma and breast milk supernatants collected 4 to 5 weeks after delivery from 12 chronically HIV-infected, lactating women (Table 1). All subjects received single-dose nevirapine at delivery but otherwise remained off antiretroviral therapy. Between 20 and 38 amplicons were generated from plasma, and 9 to 79 amplicons were generated from breast milk samples (see Table 3), by single-genome amplification. Three subjects with laboratory evidence of subclinical mastitis (sodium concentration of >12 mM and a sodium/potassium ratio of >1) and three subjects who transmitted the virus to their infants via breastfeeding (infant was HIV DNA negative at birth and 4 to 5 weeks of age but was confirmed as being HIV DNA positive at >6 weeks of age) were included in the study. The median virus loads trended toward being significantly higher in mastitic milk (median, 29,650 copies/ml; range, 635 to 102,500 copies/ml) than in nonmastitic milk (median, 3,335 copies/ml; range, 478 to 101,500 copies/ml) ($P = 0.06$) but were

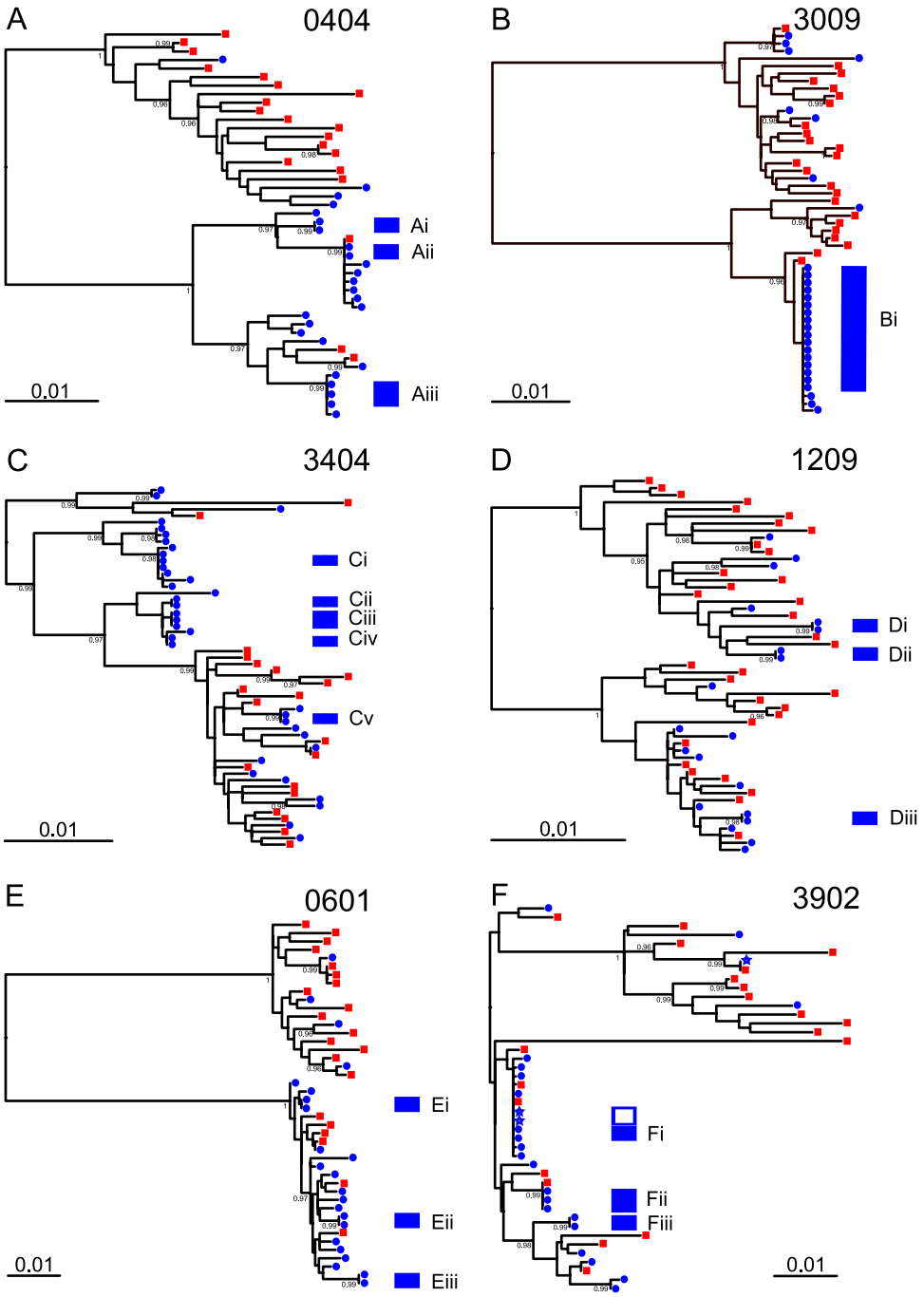


FIG. 1. HIV variants in milk do not appear to be compartmentalized from plasma virus variants in phylogenetic analyses yet display a higher frequency of genetically identical virus variants than that of plasma virus variants. Trees from maximum likelihood analyses are shown for full-length HIV *env* RNA nucleotide sequences amplified by single-genome amplification from milk (dark blue circles) and plasma (red circles) from 12 chronically HIV-infected Malawian women (A to L). HIV *env* sequences amplified from milk from the left breast are identified with filled circles, and those amplified from milk from the right breast are noted with stars. Open blue bars represent groups of identical viruses from the right breast, and filled blue bars represent groups of identical viruses from the left breast. Numerals at nodes indicate approximate likelihood ratio test values of ≥0.95. The scale bar represents 0.01 nuclear substitution per site. Groups of identical *env* sequences in plasma (red bars) or breast milk (blue bars) are indicated and numbered.

not different in milk of transmitting (median, 29,650 copies/ml; range, 635 to 101,500 copies/ml) and nontransmitting (median, 3,335, copies/ml; range, 478 to 102,500 copies/ml) subjects ($P = 0.72$). Interestingly, two of the three transmitting subjects had laboratory evidence of subclinical mastitis, consistent with the association of mastitis and the risk of breast milk transmission (47, 52).

Breast milk and plasma HIV *env* sequences were compared by maximum likelihood phylogenetic analysis (Fig. 1). Importantly, breast milk HIV *env* gene sequences were interspersed

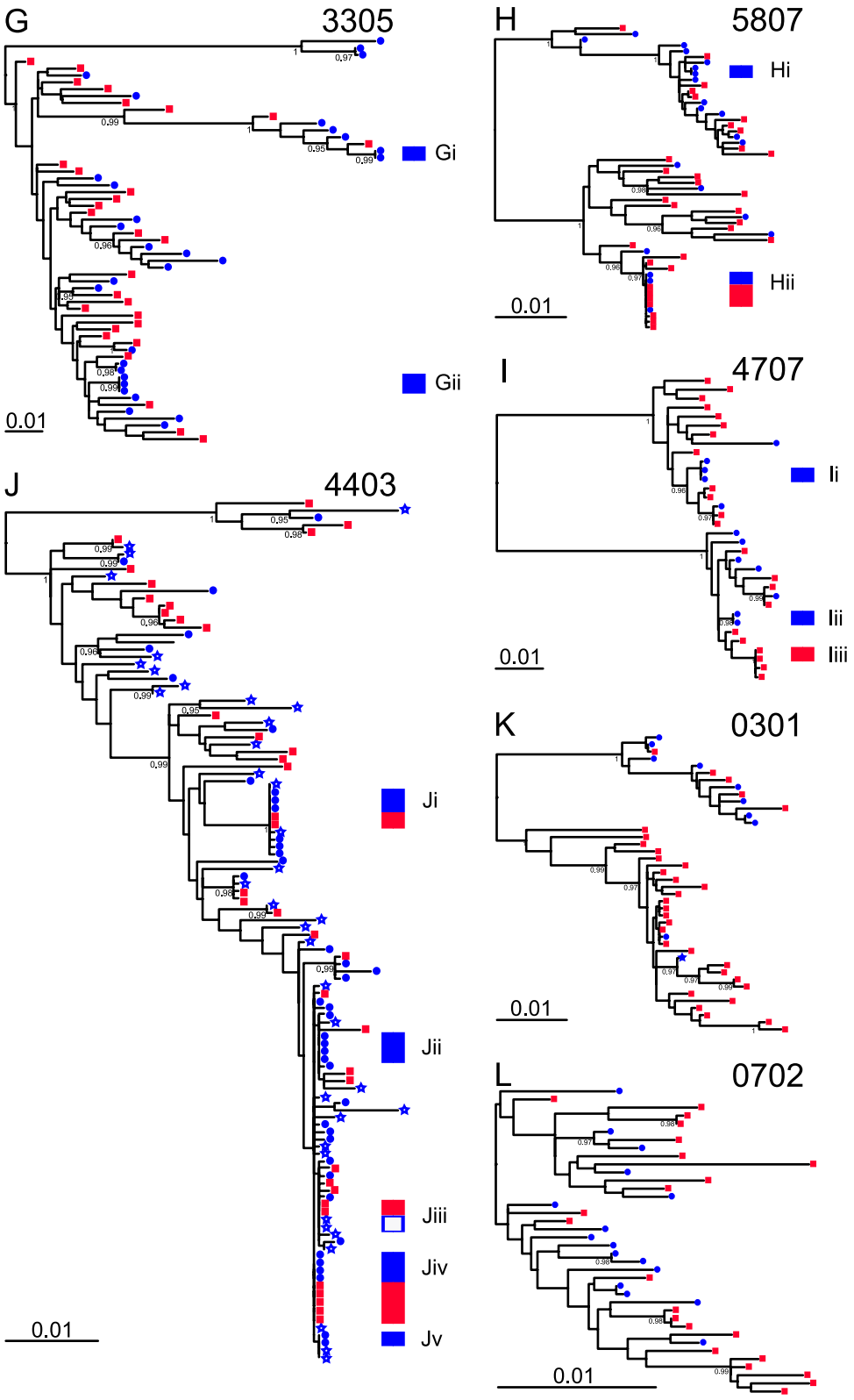


FIG. 1—*Continued*.

TABLE 2. Results of SM and S_{nn} tests of compartmentalization of plasma and breast milk virus populations from HIV-infected, lactating mothers[b]

Subject	*P* value			
	SM test	S_{nn} test	SM test without identical sequences	S_{nn} test without identical sequences
0301	0.19	0.05	0.19	0.05
0404[a]	**<0.0001**	**0.007**	**<0.0001**	**0.003**
0601	**0.001**	0.16	**0.02**	0.48
0702	0.23	0.14	0.23	0.14
1209[a]	**0.007**	0.19	**0.02**	0.32
3009	**<0.0001**	**<0.0001**	**0.04**	**0.02**
3305	0.05	0.07	0.09	0.19
3404	**0.002**	**<0.0001**	**0.03**	**0.001**
3902	0.24	0.62	0.37	0.32
4403[a]	0.08	0.07	**0.04**	0.15
4707	0.10	0.85	0.19	0.27
5807	0.21	0.53	0.43	0.72

[a] Breast milk-transmitting subject.
[b] Subjects with laboratory evidence of subclinical mastitis are indicated with shading. Significant *P* values (<0.05) are shown in bold.

among the plasma HIV *env* gene sequences on the phylogenetic trees comparing virus variants amplified from milk and plasmas of all HIV-infected, lactating women (Fig. 1). A portion of the milk virus variants clustered on an isolated branch of the phylogenetic tree in three subjects (subjects 0404, 3009, and 3404) (Fig. 1A, B, and C). Many of the milk virus variants that clustered on a single branch in these subjects were identical or nearly identical sequences. The milk virus *env* sequences were not confined to a specific branch in the remaining majority of the phylogenetic trees, indicating limited compartmentalization of the breast milk virus population from the circulating virus population in the majority of subjects. Furthermore, the HIV *env* gene sequences of virus variants amplified from milk collected from the right breast of three subjects did not cluster separately from those amplified from milk of the left breast from the same subjects (subjects 3902, 0301, and 4403) (Fig. 1F, K, and J). Moreover, the HIV *env* sequences amplified from right breast milk clustered within groups of identical viruses that included sequences amplified from left breast milk for two subjects (subjects 3902 and 4403) (Fig. 1F and J).

We then investigated the degree of compartmentalization of the milk and blood sequences by both the SM test of the minimum possible number of intercompartmental migration events compared to the distribution of migration events in 1,000 randomized trees (39, 41, 49) and Hudson's nearest-neighbor statistic (S_{nn}), which measures how often the nearest neighbors of each sequence were isolated from the same or different compartments (17, 53) (Table 2). Five of 12 subjects had evidence of compartmentalization of milk and blood virus sequences by either test ($P < 0.05$). However, only the three subjects with sequences from milk visibly isolated on a separate branch of the maximum likelihood phylogenetic tree (subjects 0404, 3009, and 3404) (Fig. 1A, B, and C) exhibited significant compartmentalization between blood and milk virus variants by both tests. Interestingly, there was significant compartmentalization by phylogenetic and statistical analyses of viruses in two of the three milk samples from subjects with laboratory evidence of subclinical mastitis (subjects 0404 and 3404), a finding that may be consistent with the high level of genetic diversity in mastitic milk that has been described previously (13). Furthermore, there was evidence of compartmentalization by phylogenetic and statistical tests of compartmentalization for one of the three breast milk-transmitting subjects (subject 0404).

Each of the subjects with evidence of significant compartmentalization by both tests had a large number of identical or nearly identical sequences in milk (Table 3), which could contribute to an overestimation of potential compartmentalization (6, 7, 15). Therefore, the identical sequences within the same compartment were removed from the data set subjected to the SM and S_{nn} analyses for all subjects. Significant compartmentalization between blood and milk virus variants remained evident for 6 of 12 subjects by the SM test, for 3 of 12 subjects by the S_{nn} test, and for 3 subjects (0404, 3009, and 3404) by both tests (Table 2). However, the degree of significance was marginal for many of the subjects after removal of the repeated identical sequences ($P = 0.02$ to 0.04) (Table 2). Therefore, we assessed the contribution of recombination to the tests of compartmentalization (1) for the three subjects with significant compartmentalization in two statistical tests by repeating the statistical tests on 4 or 5 segments of the genome that did not contain a recombination breakpoint detected using GARD (25). Both tests of compartmentalization for all nonrecombined *env* segments remained significant for subject 0404, whereas 2 of 5 *env* segments from subject 3009 and 1 of 4 *env* segments from subject 3404 virus populations were no longer significantly compartmentalized by the SM test (data not shown). Therefore, there is evidence that the breast milk and blood represent distinct anatomic compartments of HIV evolution in only a minority of chronically HIV-infected, lactating women.

Higher frequency of functional, clonally amplified virus variants in breast milk virus populations than in plasma virus populations. Notably, groups of identical or nearly identical viruses appeared more commonly among breast milk virus variants than among plasma virus variants in the majority of subjects (Fig. 1; Table 3). In fact, groups of exactly identical viruses (*env* sequences with 0 nucleotide differences) were significantly more frequent among milk virus variants than among plasma virus variants for 6 of 12 subjects by Fisher's exact test (Fig. 1; Table 3). Moreover, the median frequency of genetically identical virus variants in milk among the 12 subjects (median, 28.5%; range, 0 to 61%) was significantly higher than that in plasma (median, 0%; range, 0 to 26%) ($P = 0.002$). Interestingly, two subjects exhibited groups of identical *env* sequences that included both plasma and breast milk virus variants (subjects 4403 and 5807) (Fig. 1J and H). Furthermore, both right and left breast milk virus variants populated the major cluster of genetically identical and nearly identical virus variants in subject 4403 (Fig. 1J). The frequent occurrence of clonally amplified viruses in milk that are not phylogenetically distinct from those in plasma or the contralateral breast suggests that cell-free or cell-associated virus in blood traffics to the mammary gland and transiently replicates locally, contributing to a proportion of the virus pool in the breast milk compartment.

To explore the factors in the breast milk compartment that

TABLE 3. Proportions of clonally amplified virus variants and genetic diversity of plasma and breast milk virus populations from HIV-infected, lactating mothers[b]

Subject	Sample type	No. of amplicons from left breast milk (no. of amplicons from right breast milk)	No. of clonally amplified variants[c] (proportion [%])	*P* value[d]	Median percent genetic diversity (range)	*P* value
0301	Plasma	32	0 (0)	NC	1.42 (0.04–6.03)	0.08
	Milk	9 (1)	0 (0)		1.78 (0.08–3.39)	
0404[a]	Plasma	20	0 (0)	**0.012**	2.21 (0.12–5.90)	0.33
	Milk	25	7 (28)		2.47 (0–5.77)	
0601	Plasma	21	0 (0)	**0.02**	1.78 (0.08–13.7)	**<0.0001**
	Milk	23	6 (26)		2.47 (0–5.78)	
0702	Plasma	21	0 (0)	NC	1.54 (0.04–2.96)	**<0.0001**
	Milk	16	0 (0)		1.21 (0.04–2.22)	
1209[a]	Plasma	33	0 (0)	**0.002**	2.1 (0.08–3.84)	**0.03**
	Milk	20	6 (30)		1.66 (0–3.23)	
3009	Plasma	24	0 (0)	**<0.0001**	2.1 (0.1–8.8)	**<0.0001**
	Milk	28	17 (61)		0.16 (0–8.65)	
3305	Plasma	29	0 (0)	0.05	2.01 (0.86–6.2)	**<0.0001**
	Milk	30	5 (17)		2.5 (0–15.68)	
3404	Plasma	20	0 (0)	**0.004**	1.16 (0.16–5.07)	**<0.0001**
	Milk	36	11 (31)		2.43 (0–4.04)	
3902	Plasma	20	0 (0)	**0.009**	3.8 (0.04–9.46)	**<0.0001**
	Milk	21 (3)	7 (33)		1.34 (0–5.57)	
4403[a]	Plasma	38	9 (24)	0.81	1.42 (0–4.88)	**<0.0001**
	Milk	42 (37)	13 (31)		1.25 (0–5.53)	
4707	Plasma	22	2 (9)	0.15	9.3 (0–11.35)	0.68
	Milk	12	4 (33)		9.83 (0–11.24)	
5807	Plasma	32	4 (12)	0.7	2.62 (0–6.53)	0.41
	Milk	21	4 (19)		3.39 (0–6.63)	
All subjects (n = 12)	Plasma	344	Median proportion, 0; range, 0–26	**0.002**	2.1 (0–13.7)	0.83
	Milk	317	Median proportion, 29; range, 0–61		1.78 (0–15.7)	

[a] Breast milk-transmitting subject.
[b] Subjects with laboratory evidence of subclinical mastitis are indicated with shading. Significant *P* values (<0.05) are shown in bold.
[c] Calculated by including sequences from left breast milk only.
[d] NC, Fisher's exact test is not calculable due to the presence of 0 monotypic viruses in both plasma and milk.

may be associated with clonal amplification of milk virus, we investigated the association of breast inflammation with the frequency of genetically identical virus variants. Three of three subjects with laboratory evidence of subclinical mastitis had significantly higher proportions of genetically identical viruses in milk than in blood, but the sodium concentration in milk did not correlate with the proportion of monotypic viruses in milk ($r^2 = 0.4$; $P = 0.19$). In addition, the two transmitting subjects with mastitic milk in this cohort had significantly more monotypic viruses in milk than in plasma.

Consistent with the high frequency of identical viruses in milk, the median percent genetic diversity in the HIV *env* genes of the breast milk virus population was significantly lower than that for the plasma virus population for five subjects (0702, 1209, 3009, 3902, and 4403) (Table 3). This low genetic diversity among breast milk virus variants compared to plasma virus variants of these subjects likely reflects the large number of genetically identical or nearly identical viruses in the milk of these subjects (Fig. 1). The median percent genetic diversity in the milk for the remaining subjects was either equal to (subjects 0301, 0404, 4707, and 5807) or greater than (subjects 0601, 3305, and 3404) (Table 3) that in the plasma virus population, despite statistically greater proportions of genetically identical virus *env* genes in milk of subjects 0601, 3404, and 0404 than those for plasma. Overall, there was no difference in the median percent genetic diversity in the *env* gene for milk (1.78%) and plasma (2.1%) viruses among all 12 subjects ($P = 0.83$) (Table 3). In addition, we did not find any association between milk sodium concentration and the genetic distance of the breast milk virus variants ($r^2 = 0.08$; $P = 0.82$). Notably, the genetic diversity of milk virus variants was not consistently lower than that of plasma virus variants, as might be expected with divergent evolution of a small pool of anatomically compartmentalized virus variants.

We then assessed the functionality and coreceptor usage of the clonally amplified virus *env* sequences identified in milk of four HIV-infected, lactating women to assess if the monotypic virus variants were functional and shared a consistent phenotype with that of transmitted/founder virus variants (20). *env* sequences within groups of identical milk virus variants were cloned and used to make pseudoviruses by transfection of 293T cells. All of the clonally amplified milk *env* sequences tested produced functional pseudoviruses that were able to infect $CD4^+$ $CCR5^+$ TZM-bl cells with variable infectivity, measured as the $TCID_{50}$/ng p24 (Table 4). Furthermore, all of the clonally amplified milk *env* pseudoviruses used CCR5, not CXCR4, as the virus coreceptor (Table 4). Therefore, the clonally amplified milk virus variants identified by *env* ampli-

TABLE 4. Monotypic virus *env* genes in milk produce functional virus envelopes[a]

Subject	*env* amplicon	Monotypic virus group[b]	Pseudovirus titer ($TCID_{50}$/ng of p24) in TZM-bl cells
3305	bmC9	Gii	1,777.1
3305	bmF4	Gi	101.1
3902	bmG14	Fi	171.7
3902	bmH9	Fii	324.8
3902	bmD7	Fiii	14.1
4403	bmC5	Jiv	8.7
4403	bmB6	Ji	19.4
4707	bmF8	Ii	65.8

[a] All viruses demonstrated CCR5 coreceptor usage.
[b] Monotypic virus groups are indicated in Fig. 1.

fication and sequencing are likely functional and have a cellular tropism consistent with transmitted virus variants (20, 43).

Absence of an *env* genotype or phenotype unique to breast milk virus variants. We next sought to determine if viruses found in breast milk carry a unique HIV *env* genotype that could differentiate breast milk viruses from circulating plasma viruses. We first employed Highlighter plot sequence analysis to compare breast milk HIV *env* gene sequences to plasma HIV *env* gene sequences from each subject (data not shown). The Highlighter plot analysis revealed no specific virus *env* mutation or deletion that appeared consistently among the milk virus variants and not in plasma virus variants. Therefore, we did not identify an HIV *env* gene sequence that defined a milk virus variant from a plasma virus variant in these subjects.

We then used the virus *env* sequences to predict potential Env phenotypic differences in breast milk and plasma viruses. A larger number of N-linked glycosylation sites and deletions in the *env* variable (V) loop region may confer virus neutralization resistance, a phenotype that has been associated with vertically transmitted viruses (10, 43). The median numbers of N-linked glycosylation sites in the entire *env* gene product for milk virus variants were significantly greater than those for plasma virus variants for two subjects (3009 and 3902) and were lower than those for plasma virus variants for two subjects (0301 and 0601) (Table 5). While the median N-linked glycosylation numbers were not vastly different in plasmas and milk of those subjects with statistically significant differences in N-linked glycosylation number, the ranges of the numbers of predicted glycosylation sites in the large number of virus variants analyzed in those subjects were minimally overlapping, leading to a statistical difference by the paired Wilcoxon rank sum test. The remaining subjects had no significant difference in the number of N-linked glycosylation sites of plasma and virus variants. Furthermore, the median numbers of N-linked glycosylation sites were similar in milk (30 sites) and plasma (31 sites) virus *env* gene sequences among all 12 subjects ($P = 0.14$) (Table 5). Therefore, no consistent, distinct pattern in the number of *env* glycosylation sites existed among milk virus variants compared to plasma viruses of our subjects.

We next compared the V region lengths of milk and plasma

TABLE 5. Numbers of N-linked glycosylation sites and V region lengths for breast milk and plasma virus variants from chronically HIV-infected, lactating women[b]

Subject	Sample type	Median no. of gp160 N-glycosylation sites (range)	*P* value	Median V region length (range)	*P* value
0301	Plasma	32 (31–33)	**0.04**	148 (143–152)	**0.01**
	Milk	31 (29–32)		146.5 (139–148)	
0404[a]	Plasma	31 (28–33)	0.07	144.5 (141–149)	0.9
	Milk	30 (28–33)		145 (141–146)	
0601	Plasma	33 (30–36)	**0.005**	148 (138–157)	**0.008**
	Milk	30 (30–36)		142 (142–156)	
0702	Plasma	25 (24–27)	0.68	128 (127–131)	0.08
	Milk	25 (25–27)		128 (127–132)	
1209[a]	Plasma	29 (24–35)	0.32	149 (138–161)	0.08
	Milk	25 (24–32)		143 (138–160)	
3009	Plasma	27 (26–34)	**0.03**	144 (142–160)	**0.002**
	Milk	32 (26–34)		160 (142–160)	
3305	Plasma	33 (29–34)	0.10	165 (146–173)	**0.006**
	Milk	31 (29–33)		161 (134–167)	
3404	Plasma	30.5 (29–33)	0.07	218 (218–235)	0.09
	Milk	30 (27–31)		218 (214–236)	
3902	Plasma	32.5 (27–33)	**0.006**	161 (137–167)	**0.03**
	Milk	33 (31–34)		163 (151–167)	
4403[a]	Plasma	28 (27–33)	0.06	141 (139–149)	0.12
	Milk	28 (25–32)		141 (122–153)	
4707	Plasma	32 (25–38)	0.84	150 (130–161)	0.62
	Milk	27 (25–35)		130 (130–158)	
5807	Plasma	31 (27–34)	0.21	151 (128–156)	0.12
	Milk	29 (27–34)		132 (128–155)	
All subjects	Plasma	31 (24–38)	0.14	148 (127–235)	0.19
	Milk	30 (24–36)		143 (122–236)	

[a] Breast milk-transmitting subject.
[b] Subjects with laboratory evidence of subclinical mastitis are indicated with shading. Significant *P* values (<0.05) are shown in bold.

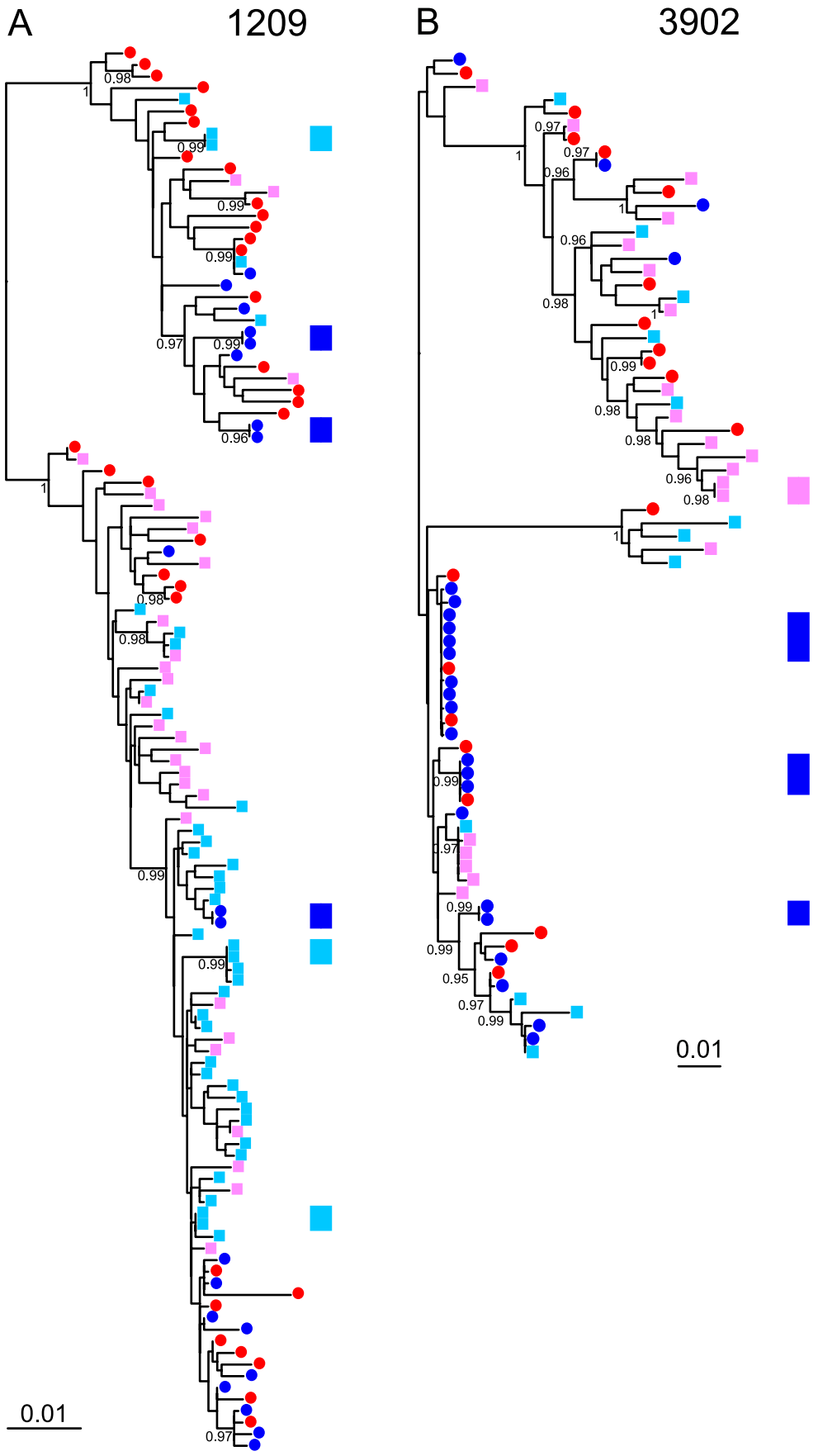
TABLE 6. Milk- and plasma-derived envelope pseudoviruses display similar sensitivities to neutralization by sCD4, broadly neutralizing monoclonal antibodies, and plasmas of chronically HIV-infected subjects[a]

Pseudovirus	Sample type	Titer (50% infective dose)										
		SA C3	SA C9	SA C26	SA C32	SA C76	HIVIg	1B12	2G12	2F5	4E10	sCD4
3902.A7	Plasma	>625	239.5	153.6	>625	373.0	857.0	>25	>25	>25	4.0	4.0
3902.B10	Plasma	>625	111.8	>625	131.4	221.0	>2,500	24.9	>25	24.6	7.6	5.2
3902.G3	Plasma	>625	196.7	>625	>625	123.2	1,670.0	>25	>25	14.2	7.7	14.1
3902.G10	Plasma	>625	>625	>625	>625	>625	2,399.0	>25	>25	20.1	6.1	24.8
3902.G12	Plasma	461.4	122.8	31.1	138.4	197.3	1,504.0	>25	>25	>25	8.0	3.8
3902.BMD7	Milk	197.5	131.4	31.5	126.3	231.0	>2,500	>25	>25	>25	24.1	4.4
3902.BME6	Milk	>625	198.5	39.7	427.6	295.3	1,848.0	>25	>25	>25	18.0	6.2
3902.BMG14	Milk	>625	80.8	86.6	164.0	202.7	1,971.0	13.6	>25	19.4	11.5	1.9
3902.BMG24	Milk	>625	62.7	23.0	176.4	149.9	1,598.0	>25	>25	15.0	7.1	4.3
3305.A7	Plasma	>625	>625	>625	>625	>625	>2,500	>25	>25	>25	>25	>25
3305.A10	Plasma	>625	>625	276.6	>625	>625	>2,500	>25	>25	>25	6.7	>25
3305.C4	Plasma	>625	>625	>625	>625	>625	>2,500	>25	>25	>25	11.9	20.2
3305.E9	Plasma	>625	>625	255.3	>625	406.7	>2,500	>25	>25	>25	14.0	23.2
3305.BMC1	Milk	>625	>625	414.0	>625	446.6	>2,500	>25	>25	>25	5.8	>25
3305.BMC9	Milk	>625	>625	271.6	>625	476.3	>2,500	>25	>25	>25	17.1	>25
3305.BMC11	Milk	>625	>625	40.3	>625	>625	>2,500	>25	>25	>25	13.2	15.8
3305.BMF4	Milk	>625	>625	>625	>625	>625	>2,500	>25	>25	>25	8.4	>25
3305.BMF6	Milk	>625	>625	381.9	>625	575.2	>2,500	>25	>25	>25	14.0	>25
4403.A18	Plasma	195.4	49.5	>625	32.2	138.5	1,612.7	>25	>25	>25	>25	8.2
4403.D1	Plasma	184.4	41.5	>625	38.9	145.5	679.9	>25	>25	>25	>25	5.0
4403.H2	Plasma	112.8	44.6	>625	38.5	111.9	1,064.1	>25	>25	>25	>25	4.0
4403.H7	Plasma	346.7	71.4	>625	49.0	181.7	1,380.3	>25	>25	>25	>25	5.1
4403.BMB6	Milk	272.4	58.9	>625	45.8	158.3	523.6	>25	>25	>25	>25	3.1
4403.BMB10	Milk	204.4	44.7	>625	33.7	125.0	1,320.3	>25	>25	>25	>25	4.3
4403.BMC5	Milk	302.6	58.8	>625	48.7	154.8	1,090.1	>25	>25	>25	>25	4.8
0301.A8	Plasma	310.6	16.6	133.3	48.4	55.6	632.1	>25	>25	9.3	11.1	3.5
0301.D1	Plasma	197.4	30.7	119.8	53.2	62.2	984.7	>25	>25	5.6	7.7	7.8
0301.E8	Plasma	323.4	73.2	149.0	69.7	111.0	1,800.7	>25	>25	20.3	18.3	6.4
0301.BMA2	Milk	230.2	15.3	104.6	36.5	45.0	1,229.7	>25	>25	10.0	16.9	2.2
0301.BMA6	Milk	40.4	5.2	>625	24.8	18.9	576.8	>25	>25	9.8	12.2	5.7
0301.BMA12	Milk	111.3	22.1	499.1	40.4	44.3	1,133.6	>25	>25	8.3	10.1	3.8
47707.E1	Plasma	>625	293.9	>625	79.0	277.1	>2,500	>25	>25	>25	>25	>25
47707.G3	Plasma	365.2	293.9	511.0	281.8	>625	>2,500	>25	>25	>25	6.9	1.2
47707.H8	Plasma	>625	>625	>625	>625	>625	>2,500	>25	>25	>25	>25	10.3
47707.BMF8	Milk	>625	>625	>625	549.2	>625	>2,500	>25	>25	>25	>25	6.6
47707.BMAH2	Milk	>625	139.1	>625	105.1	171.8	>2,500	>25	>25	>25	>25	10.5
0702.A5	Plasma	>625	368.4	249.2	234.2	107.8	1,635.2	>25	>25	>25	3.9	>25
0702.B12	Plasma	>625	521.0	>625	388.6	141.7	944.9	>25	>25	>25	10.2	>25
0702.F21	Plasma	248.6	564.6	177.6	463.2	122.9	1,408.5	>25	>25	>25	5.2	>25
0702.G6	Plasma	185.3	507.0	82.2	217.8	83.2	>2,500	>25	>25	>25	5.4	11.8
0702.BMB4	Milk	>625	336.5	>625	175.2	80.4	1,998.3	>25	>25	>25	5.6	>25
0702.BMB9	Milk	>625	537.6	338.3	308.8	116.7	1,897.2	>25	>25	>25	3.6	13.4
0702.BMH12	Milk	>625	544.4	>625	333.6	97.3	1,368.6	>25	>25	>25	4.8	11.1

[a] Shading represents detectable neutralization.

virus variants. The V region lengths of breast milk virus variants were significantly shorter than those of plasma virus variants for three subjects (0301, 0601, and 3305) and were significantly longer than those for plasma virus variants for two subjects (3009 and 3902) (Table 5). The median V region lengths were not significantly different among plasma and milk virus variants from the other subjects. Moreover, the overall median V region lengths of milk (143 nucleotides) and plasma (148 nucleotides) virus *env* sequences from all 12 subjects did not differ significantly ($P = 0.19$) (Table 5).

Finally, to determine if functionally distinct virus variants are harbored in the breast milk compartment, suggesting distinct virus evolution in this compartment, we compared the neutralization sensitivities of two to five plasma and milk Env pseudoviruses from seven subjects. The milk and plasma Env variants of each subject displayed comparable sensitivities to plasma IgGs of chronically clade C HIV-infected patients with broad neutralizing activity (SA C3, SA C6, SA C32, SA C26, and SA C76) and to pooled HIVIg (Table 6). Typically, the majority of the Env pseudoviruses from both compartments of a single subject were uniformly resistant or susceptible to the reagent. Occasionally, only one or two viruses from either the milk or plasma compartment were resistant to the reagent. Furthermore, the neutralization sensitivities of the breast milk and plasma viruses from all subjects to HIVIg and plasma IgG of the chronically HIV-infected subjects are consistent with a tier 2 neutralization phenotype (31). Similarly, the milk and plasma Env pseudoviruses within each subject displayed similar sensitivities to well-characterized broadly HIV-neutralizing antibodies. The majority of the Env pseudoviruses from both the milk and plasma of each subject were resistant to monoclonal antibodies 1B12 and 2G12. The sensitivities of the Env



pseudoviruses to 2F5 were either uniform in both the milk and plasma compartments or unique to a portion of the variants from both compartments. The sensitivities of the Env pseudoviruses to 4E10 were typically uniform in both the milk and plasma compartments or unique to just a single variant from the plasma compartment. Finally, all or a portion of the Env pseudoviruses from each compartment within each subject were susceptible to neutralization by sCD4.

Based on these data, breast milk viruses do not appear to have a distinct *env* gene sequence, predicted *env* gene product phenotype, or functional neutralization phenotype compared to plasma viruses. This uniform appearance of the milk and plasma virus *env* genotype and phenotype further supports the lack of compartmentalization of virus in milk and the continual seeding of the breast milk compartment by circulating virus variants. Moreover, the similar predicted Env phenotypes and neutralization sensitivities of milk and plasma virus Envs indicate that milk and plasma viruses evolve under similar immune pressures. Interestingly, the breast milk virus *env* genotype, N-linked glycosylation number, and V region length were not distinct from those in plasma in the three transmitting subjects included in this study.

Continual seeding of the breast milk compartment from the blood virus population. Finally, we sought to determine whether the breast milk virus population undergoes distinct virus evolution from that in plasma over time. We amplified and sequenced the virus *env* genes from blood and breast milk virus variants of two subjects at 4 to 5 weeks and 3 months postpartum and phylogenetically compared the compartment-specific longitudinal *env* sequences (Fig. 2). The breast milk and plasma sequences evolved synchronously between 4 to 5 weeks and 3 months postpartum in both subjects and continued to intermingle on the phylogenetic tree at 3 months postpartum. Therefore, breast milk virus variants do not appear to evolve distinctly over time. Importantly, we did not see persistence of the same groups of genetically identical viruses present at 4 to 5 weeks postpartum. Furthermore, groups of identical viruses appeared in milk at 3 months postpartum that were not present at 4 to 5 weeks postpartum in subject 1209, indicating local virus replication in the breast milk compartment of newly evolved virus variants recently seeded from the circulation. The appearance of new groups of identical viruses longitudinally in milk that are closely related to concurrent blood virus variants is evidence of ongoing seeding of the breast milk compartment by the systemic virus population, followed by repeated, transient bursts of virus replication in milk.

DISCUSSION

The statistically significantly larger proportion of genetically identical viruses in breast milk than in plasma for 6 of 12 subjects in this cohort, identified by a sequencing method that limits the possibility of virus resampling (20, 46), is striking. Moreover, the milk virus *env* sequences from these groups of identical viruses produced functional pseudovirions, indicating that the groups of identical virus *env* sequences represent replication of functional virus variants. This clonal amplification of virus variants has been described for milk of chronically SIV-infected, lactating monkeys (38) and HIV-infected, lactating women (13, 15). However, the presence of genetically identical viruses in HIV-infected, lactating women has not been confirmed previously by strict single-virus RNA genome amplification or assessed for functionality and persistence. While we used only genetically identical *env* genes for statistical analysis of monotypic viruses in milk and plasma in this study, the nearly identical viruses that cluster together on the phylogenetic trees certainly also represent clonally amplified variants, given the large amount of sequence diversity within a single chronically HIV-infected individual. Therefore, our report of the frequency of only the genetically identical viruses in milk likely underestimates the true frequency of clonally amplified viruses in milk. The high frequency of genetically identical and nearly identical clonally amplified virus variants in milk indicates that infected cells that have trafficked to or are resident in the mammary gland may produce at least a portion of the viruses present in the milk.

Despite evidence that breast milk virus may be produced by infected cells resident in the breast milk compartment, we found limited evidence of phylogenetic compartmentalization and distinct evolution of viruses in breast milk from that of circulating viruses by maximum likelihood tree analysis and phylogenetic and genetic distance-based tests of compartmentalization for the majority of subjects. Furthermore, longitudinal *env* sequencing of milk and plasma virus variants revealed analogous virus evolution over time, indicating ongoing seeding of the breast milk virus population by blood variants. This lack of compartmentalization of virus in milk in the majority of subjects is in contrast to studies of virus populations in the male and female genital tracts, where compartmentalization of virus in genital secretions is frequently observed in the majority of subjects studied (1, 2, 6, 9, 11, 21, 22, 37). These contrasting results may indicate a difference in virus trafficking from the systemic circulation to the genital tract and the breast milk compartment (38). Moreover, comparison of the levels of genetic diversity of the HIV *env* sequences of breast milk and plasma virus populations did not reveal consistently low genetic diversity among breast milk virus variants, which might be expected of a compartmentalized, small pool of virus that evolves distinctly from circulating virus. However, the women included in this study were selected for detectable virus loads in milk, and therefore, it is possible that the degree of compartmentalization of milk virus is substantial in women with

FIG. 2. Milk and plasma HIV variants evolve similarly over time, with clonal amplification of new virus variants in milk. Trees from maximum likelihood analyses are shown for full-length HIV *env* RNA nucleotide sequences amplified by single-genome amplification from milk (dark blue circles) and plasmas (red circles) obtained at 4 to 5 weeks and from milk (light blue squares) and plasmas (pink squares) obtained 3 months after delivery from two chronically HIV-infected, lactating women. Numerals at nodes indicate approximate likelihood ratio test values of ≥0.95. The scale bar represents 0.01 nuclear substitution per site. Groups of identical sequences in plasma at 4 to 5 weeks (red bars) and 3 months (pink bars) or in breast milk at 4 to 5 weeks (dark blue bars) and 3 months (light blue bars) are indicated.

low milk virus loads. Although early studies employing phylogenetic analysis of the variable region of HIV *env* after bulk cloning of breast milk and plasma viruses were contradictory as to whether these virus populations were compartmentalized (5, 16), the application of single-genome amplification of full-length HIV *env* and the consistency with data from nonhuman primate models (38) and other human cohorts (13, 15) increase the strength of our finding that breast milk viruses are not compartmentalized phylogenetically from plasma viruses in the majority of subjects.

Interestingly, two subjects displayed groups of identical and nearly identical viruses that included both blood and milk virus sequences (subjects 4403 and 5807). Furthermore, both right and left breast milk virus variants populated the major cluster of genetically identical and nearly identical virus variants for subject 4403. Moreover, the genetic diversity of virus *env* sequences in milk samples that displayed a large proportion of clonally amplified virus variants was not always more restricted than that in plasma (subjects 0601, 3305, and 3404). These findings, paired with the appearance of new clusters of clonally amplified virus variants that are closely related to concurrent blood virus variants in longitudinal milk virus sequencing, indicate ongoing bursts of local virus replication after seeding of the milk compartment by circulating virus variants. In fact, a recent comparison of breast milk HIV *env* genetic diversity in antiretroviral-treated women revealed virus evolution under selective drug pressure in milk that was distinct from that in plasma, implicating local replication of breast milk virus (3). This insight into the location of milk virus replication is important to the design of immunologic interventions to reduce virus content in the milk, as decreasing milk virus loads would likely lead to a reduction in the risk of transmission of HIV via breastfeeding.

Finally, full-length HIV *env* sequencing of the milk virus population allowed us to assess whether a distinct virus *env* sequence, predicted virus Env phenotype, or neutralization sensitivity was unique to breast milk virus *env* genes. Analysis of milk virus *env* mutations and deletions compared to those in plasma HIV *env* sequences did not reveal a signature sequence or deletion unique to breast milk viruses. Furthermore, milk virus *env* sequences were not consistently predicted to be glycosylated more or less than those of plasma viruses, did not consistently have a longer or shorter V region length than that of plasma viruses, and were not distinctly resistant to neutralization by plasmas of chronically HIV-infected subjects or broadly HIV-neutralizing antibodies. Therefore, while neutralization-resistant viruses have been associated with breast milk transmission of HIV (43), a neutralization-resistant *env* phenotype consisting of a small number of N-linked glycosylation sites and deletions in the variable loops (28, 32) was not more common in breast milk viruses than in plasma viruses. This lack of a predicted Env phenotype or neutralization phenotype unique to breast milk virus variants suggests that plasma and breast milk viruses are not compartmentalized from each other and evolve under similar humoral immune pressures.

Importantly, we did not find evidence of distinct breast milk virus evolution, a distinct degree of local virus replication, or a predicted virus Env phenotype in milk of transmitting subjects compared to nontransmitting subjects. This finding suggests that the degree of local breast milk virus replication or the virus Env phenotype may not contribute to the risk of virus transmission via breastfeeding. Subclinical mastitis, which is known to increase the risk of breast milk virus transmission (12, 18, 47, 52), was associated with long genetic distances between plasma and milk virus variants in a recent report (13). In our study, subclinical mastitis was associated with a large proportion of clonally amplified viruses in milk for three of three subjects with mastitic milk (subjects 0404, 1209, and 3404), with evidence of compartmentalization between plasma and milk virus variants for two of three subjects with mastitic milk (subjects 0404 and 3404), and with higher genetic diversity of milk viruses than plasma viruses for one of three subjects with mastitic milk (subject 3404). These findings may suggest increased kinetics of virus replication in the breast milk compartment during subclinical mastitis and breast inflammation. Therefore, the possibility of increased local virus replication as a result of inflammatory processes in the mammary gland leading to a high risk of virus transmission via breast milk should be investigated further.

The absence of phylogenetic compartmentalization, distinct virus evolution, or unique virus *env* genetic or phenotypic features of milk viruses compared to those in plasma suggests that viruses in breast milk are replenished from the pool of circulating viruses. However, the high frequency of clonally amplified virus variants in milk suggests that a relatively small number of productively infected cells in mammary tissue transiently contribute a disproportionately large number of virions to the virus pool in milk. Therefore, there may be two mechanisms by which virus populates the breast milk compartment. The first is continual trafficking of cell-free or cell-associated virus from blood into the breast milk compartment. The second is transient local production of virus recently trafficked from the blood by HIV-infected cells in the mucosa of the mammary gland or in the milk. Therefore, immunologic and antiretroviral drug interventions to decrease breast milk virus load as a strategy to reduce transmission of HIV via breastfeeding may need to target both blood-derived virus trafficking into the mammary gland and local replication of HIV within the breast milk compartment.

ACKNOWLEDGMENTS

We are grateful to Ebbie Chalaluka, the study nurses, the staff at the Mpemba and Madziabango Health Centers, the Blantyre District Health Office, Alfred Mali, James Bunn, Victor Mwapasa, Bonus Makanani, Nicole Carpenetti, Newton Kumwenda, Taha Taha, Deborah Hilgenberg, Missie Allen, Joann Kuruc, and the study participants for their contributions to subject enrollment, clinical site management, and laboratory specimen collection for this study.

This work was supported by the Center for HIV/AIDS Vaccine Immunology (S.R.P., N.L.L., B.H.H., G.M.S., and P.B.) (R01AI067854), NIH grant K08AI087992 (S.R.P.), a Children's Hospital Boston Faculty Development Award (S.R.P.), and a Pediatric Infectious Disease Society/St. Jude Children's Hospital Basic Science Research Award (S.R.P.). P.B. is a Jenner Institute Investigator.

No conflicts of interest exist among the authors.

REFERENCES

1. **Anderson, J. A., et al.** 2010. HIV-1 populations in semen arise through multiple mechanisms. PLoS Pathog. **6:**e1001053.
2. **Andreoletti, L., et al.** 2007. Genetic and phenotypic features of blood and genital viral populations of clinically asymptomatic and antiretroviral-treatment-naive clade A human immunodeficiency virus type 1-infected women. J. Clin. Microbiol. **45:**1838–1842.
3. **Andreotti, M., et al.** 2009. Comparison of HIV type 1 sequences from

plasma, cell-free breast milk, and cell-associated breast milk viral populations in treated and untreated women in Mozambique. AIDS Res. Hum. Retroviruses **25:**707–711.
4. **Anisimova, M., and O. Gascuel.** 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst. Biol. **55:**539–552.
5. **Becquart, P., et al.** 2002. Compartmentalization of HIV-1 between breast milk and blood of HIV-infected mothers. Virology **300:**109–117.
6. **Bull, M., et al.** 2009. Compartmentalization of HIV-1 within the female genital tract is due to monotypic and low-diversity variants not distinct viral populations. PLoS One **4:**e7122.
7. **Bull, M. E., et al.** 2009. Monotypic human immunodeficiency virus type 1 genotypes across the uterine cervix and in blood suggest proliferation of cells with provirus. J. Virol. **83:**6020–6028.
8. **Chasela, C. S., et al.** Maternal or infant antiretroviral drugs to reduce HIV-1 transmission. N. Engl. J. Med. **362:**2271–2281.
9. **Delwart, E. L., et al.** 1998. Human immunodeficiency virus type 1 populations in blood and semen. J. Virol. **72:**617–623.
10. **Dickover, R., et al.** 2006. Role of maternal autologous neutralizing antibody in selective perinatal transmission of human immunodeficiency virus type 1 escape variants. J. Virol. **80:**6525–6533.
11. **Ellerbrock, T. V., et al.** 2001. Cellular replication of human immunodeficiency virus type 1 occurs in vaginal secretions. J. Infect. Dis. **184:**28–36.
12. **Embree, J. E., et al.** 2000. Risk factors for postnatal mother-child transmission of HIV-1. AIDS **14:**2535–2541.
13. **Gantt, S., et al.** 2010. Genetic analyses of HIV-1 env sequences demonstrate limited compartmentalization in breast milk and suggest viral replication within the breast that increases with mastitis. J. Virol. **84:**10812–10819.
14. **Guindon, S., F. Delsuc, J. F. Dufayard, and O. Gascuel.** 2009. Estimating maximum likelihood phylogenies with PhyML. Methods Mol. Biol. **537:**113–137.
15. **Heath, L., et al.** Restriction of HIV-1 genotypes in breast milk does not account for the population transmission genetic bottleneck that occurs following transmission. PLoS One **5:**e10213.
16. **Henderson, G. J., et al.** 2004. HIV-1 populations in blood and breast milk are similar. Virology **330:**295–303.
17. **Hudson, R. R.** 2000. A new statistic for detecting genetic differentiation. Genetics **155:**2011–2014.
18. **John, G. C., et al.** 2001. Correlates of mother-to-child human immunodeficiency virus type 1 (HIV-1) transmission: association with maternal plasma HIV-1 RNA load, genital HIV-1 DNA shedding, and breast infections. J. Infect. Dis. **183:**206–212.
19. **Johnson, V. A., et al.** 1991. Two-drug combinations of zidovudine, didanosine, and recombinant interferon-alpha A inhibit replication of zidovudine-resistant human immunodeficiency virus type 1 synergistically in vitro. J. Infect. Dis. **164:**646–655.
20. **Keele, B. F., et al.** 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc. Natl. Acad. Sci. U. S. A. **105:**7552–7557.
21. **Kemal, K. S., et al.** 2003. HIV-1 in genital tract and plasma of women: compartmentalization of viral sequences, coreceptor usage, and glycosylation. Proc. Natl. Acad. Sci. U. S. A. **100:**12972–12977.
22. **Kiessling, A. A., et al.** 1998. Human immunodeficiency virus in semen arises from a genetically distinct virus reservoir. AIDS Res. Hum. Retroviruses **14**(Suppl. 1)**:**S33–S41.
23. **Kilewo, C., et al.** 2008. Prevention of mother-to-child transmission of HIV-1 through breast-feeding by treating infants prophylactically with lamivudine in Dar es Salaam, Tanzania: the Mitra Study. J. Acquir. Immune Defic. Syndr. **48:**315–323.
24. **Kilewo, C., et al.** 2009. Prevention of mother-to-child transmission of HIV-1 through breastfeeding by treating mothers with triple antiretroviral therapy in Dar es Salaam, Tanzania: the Mitra Plus Study. J. Acquir. Immune Defic. Syndr. **52:**406–416.
25. **Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. Frost.** 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. **23:**1891–1901.
26. **Kraus, M. H., et al.** 2010. A rev1-vpu polymorphism unique to HIV-1 subtype A and C strains impairs envelope glycoprotein expression from rev-vpu-env cassettes and reduces virion infectivity in pseudotyping assays. Virology **397:**346–357.
27. **Kumwenda, N. I., et al.** 2008. Extended antiretroviral prophylaxis to reduce breast-milk HIV-1 transmission. N. Engl. J. Med. **359:**119–129.
28. **Laakso, M. M., et al.** 2007. V3 loop truncations in HIV-1 envelope impart resistance to coreceptor inhibitors and enhanced sensitivity to neutralizing antibodies. PLoS Pathog. **3:**e117.
29. **Larkin, M. A., et al.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:**2947–2948.
30. **Leonard, C. K., et al.** 1990. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. J. Biol. Chem. **265:**10373–10382.
31. **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.
32. **Li, Y., et al.** 2008. Removal of a single N-linked glycan in human immunodeficiency virus type 1 gp120 results in an enhanced ability to induce neutralizing antibody responses. J. Virol. **82:**638–651.
33. **Luo, C., et al.** 2007. Global progress in PMTCT and paediatric HIV care and treatment in low- and middle-income countries in 2004–2005. Reprod. Health Matters **15:**179–189.
34. **Mofenson, L. M.** 2008. Antiretroviral prophylaxis to reduce breast milk transmission of HIV type 1: new data but still questions. J. Acquir. Immune Defic. Syndr. **48:**237–240.
35. **Montefiori, D. C.** 2005. Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. Curr. Protoc. Immunol. **2005:** 12.11.
36. **Nduati, R., et al.** 2000. Effect of breastfeeding and formula feeding on transmission of HIV-1: a randomized clinical trial. JAMA **283:**1167–1174.
37. **Overbaugh, J., R. J. Anderson, J. O. Ndinya-Achola, and J. K. Kreiss.** 1996. Distinct but related human immunodeficiency virus type 1 variant populations in genital secretions and blood. AIDS Res. Hum. Retroviruses **12:**107–115.
38. **Permar, S. R., et al.** Local replication of simian immunodeficiency virus in the breast milk compartment of chronically-infected, lactating rhesus monkeys. Retrovirology **7:**7.
39. **Pillai, S. K., et al.** 2006. Genetic attributes of cerebrospinal fluid-derived HIV-1 env. Brain **129:**1872–1883.
40. **Pillay, K., A. Coutsoudis, D. York, L. Kuhn, and H. M. Coovadia.** 2000. Cell-free virus in breast milk of HIV-1-seropositive women. J. Acquir. Immune Defic. Syndr. **24:**330–336.
41. **Pond, S. L., S. D. Frost, and S. V. Muse.** 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics **21:**676–679.
42. **Poss, M., et al.** 1998. Evolution of envelope sequences from the genital tract and peripheral blood of women infected with clade A human immunodeficiency virus type 1. J. Virol. **72:**8240–8251.
43. **Rainwater, S. M., et al.** 2007. Cloning and characterization of functional subtype A HIV-1 envelope variants transmitted through breastfeeding. Curr. HIV Res. **5:**189–197.
44. **Rose, P. P., and B. T. Korber.** 2000. Detecting hypermutations in viral sequences with an emphasis on G→A hypermutation. Bioinformatics **16:** 400–401.
45. **Rousseau, C. M., et al.** 2004. Association of levels of HIV-1-infected breast milk cells and risk of mother-to-child transmission. J. Infect. Dis. **190:**1880–1888.
46. **Salazar-Gonzalez, J. F., 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.
47. **Semba, R. D., et al.** 1999. Human immunodeficiency virus load in breast milk, mastitis, and mother-to-child transmission of human immunodeficiency virus type 1. J. Infect. Dis. **180:**93–98.
48. **Shapiro, R. L., et al.** Antiretroviral regimens in pregnancy and breast-feeding in Botswana. N. Engl. J. Med. **362:**2282–2294.
49. **Slatkin, M.** 1989. Detecting small amounts of gene flow from phylogenies of alleles. Genetics **121:**609–612.
50. **Swofford, D. L.** 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods), 4.0b10. Sinauer Associates, Sunderland, MA.
51. **WHO Collaborative Study Team on the Role of Breastfeeding on the Prevention of Infant Mortality.** 2000. Effect of breastfeeding on infant and child mortality due to infectious diseases in less developed countries: a pooled analysis. Lancet **355:**451–455.
52. **Willumsen, J. F., et al.** 2000. Subclinical mastitis as a risk factor for mother-infant HIV transmission. Adv. Exp. Med. Biol. **478:**211–223.
53. **Zarate, S., S. L. Pond, P. Shapshak, and S. D. Frost.** 2007. Comparative study of methods for detecting sequence compartmentalization in human immunodeficiency virus type 1. J. Virol. **81:**6643–6651.
54. **Zhang, M., et al.** 2004. Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. Glycobiology **14:**1229–1246.