



Published in final edited form as:

AIDS. 2011 January 14; 25(2): 143–152. doi:10.1097/QAD.0b013e328340fdaf.

Multiple independent lineages of HIV-1 persist in breast milk and plasma

Rebecca R. GRAY^{1,2}, Marco SALEMI^{1,2}, Amanda LOWE¹, Kyle J. NAKAMURA^{3,4}, William D. DECKER³, Moses SINKALA⁵, Chipepo KANKASA⁶, Connie J. MULLIGAN⁷, Don THEA⁸, Louise KUHN⁹, Grace ALDROVANDI^{3,4,*}, and Maureen M. GOODENOW^{1,10,*}

¹ Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL, USA

² Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA

³ Children's Hospital of Los Angeles, Los Angeles, CA, USA

⁴ USC Keck School of Medicine, Los Angeles, CA, USA

⁵ Lusaka District Health Management Team, Lusaka, Zambia

⁶ University Teaching Hospital, University of Zambia, Lusaka, Zambia

⁷ Department of Anthropology, University of Florida, Gainesville, FL, USA

⁸ Boston University School of Public Health, Boston, MA, USA

⁹ Columbia University, New York, NY, USA

¹⁰ Florida Center for AIDS Research, University of Florida, Gainesville, FL, USA

Abstract

Design—The origin and evolution of HIV-1 in breast milk is unclear, despite the continuing significance of this tissue as a transmitting compartment. To elucidate the evolutionary trajectory of viral populations in a transient mucosal compartment, longitudinal sequences of the envelope gp120 region from plasma and breast milk spanning the first year after delivery were analyzed in six women infected by HIV-1 subtype C.

Methods—Multiple phylogenetic algorithms were used to elucidate the evolutionary history and spatial structure of virus populations between tissues.

Results—Overall persistent mixing of viral sequences between plasma and breast milk indicated that breast milk is not a distinct genetic viral compartment. Unexpectedly, longitudinal phylogenies showed multiple lineages defined by long branches that included virus from both the breast milk and the plasma. Plasma was unlikely the anatomical origin of the most recent common ancestor (MRCA) in at least three of the subjects, while in other women, the temporal origin of the MRCA of the viral populations following delivery occurred well before the onset of breast milk production.

Corresponding Author: Maureen M. Goodenow 2033 Mowry Road PO Box 103633 Gainesville, FL, 32610-3633.

*These authors contributed equally to the manuscript.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusions—These findings suggest that during pregnancy/lactation, a viral variant distinct from the plasma virus initially seeds the breast milk, followed by subsequent gene flow between the plasma and breast milk tissues. This study indicates the potential for reactivation or re-introduction of distinct lineages during major immunological disruptions during the course of natural infection.

Keywords

HIV-1; breast milk; evolution; phylogeny; compartment; reservoir

INTRODUCTION

HIV-1 rapidly evolves as a quasispecies, targeting CD4⁺ cells, primarily macrophages and T-lymphocytes [1]. Because CD4⁺ cells are distributed throughout the body, viral variants may replicate and evolve independently in anatomical tissues or compartments without contributing to systemic flow of viral genes [2]. Tissues and cells may also act as viral reservoirs if viral variants circulating earlier in infection are archived and replication is restricted [2-4].

Breast milk is a transient mucosal tissue considered a potential compartment for virus based on physiology of mammary tissue and development of lactation. Breastfeeding remains a major route of maternal transmission of HIV-1 to infants in the developing world, infecting about 14% of infants born to antiretroviral therapy naïve mothers [5]. Breast tissue is not only an important transmitting tissue, but also provides a distinct biological site for milk production. Lactating mammary tissue is part of a common mucosal immune system and provides a source of cells, antibodies and diverse immune factors distinct from those in the circulation [6]. Mammary tissue remodeling begins during the second trimester of pregnancy when cells from the gut lymphoid system home to breast tissue (lactogenesis I). Initially, tight junctions between epithelial cells are permeable, allowing immune cells and plasma proteins from the blood to enter the breast milk directly via the paracellular pathway [7]. The paracellular pathway closes at the time of delivery, and remains blocked until weaning is initiated or inflammation occurs [8], which often corresponds with a concomitant rise in breast milk sodium levels. If HIV-1 infected cells were also blocked from passing between breast milk and plasma during exclusive breast feeding, the viral population in the breast milk may diverge from virus in the plasma, leading to the evolution of “breast milk variant[s]” of virus. The breast milk may also present a unique environment that could select for specific viral variants. Breast milk is an immunologically distinct tissue comprised mainly of lymphocytes and macrophages that likely derive from the gut lymphoid tissue [9-13] and appear at the highest concentration during early lactation [14,15]. Although T lymphocytes comprise less than 10% of the total cell population in breast milk, the magnitude and breadth of the breast milk T-cell repertoire are higher than in peripheral blood and include specificities distinct from circulating plasma T-cells [16].

Despite the biological plausibility of a breast milk compartment for virus, only a small number of studies have investigated HIV-1 compartmentalization in breast milk with conflicting results, perhaps reflecting complexities of interpretation of cross-sectional studies, analysis of several different subtypes (HIV-1A, G, B/D), a range of sampling times post-partum (1 to 12 wks or 1 year), and/or various analytical methods applied to different regions of the *env* gene [17-19]. A recent analysis, which was the most comprehensive to date, of cross-sectional breast milk and plasma samples found no evidence for virus compartmentalization immediately post-partum [20]. We designed a novel longitudinal study to examine the evolution of HIV-1 subtype C *env* gp120 gene in breast milk and plasma from six women who transmitted virus to their infants either perinatally or through

breastfeeding. The goal was to define the relationship of the circulating virus with that in milk and determine how viral populations changed over time.

METHODS

Subjects

Samples were from participants in the Zambia Exclusive Breastfeeding Study (ZEBS) who were therapy naïve except for a single dose of nevirapine at the time of labor [21]. Three right and left breast milk and plasma samples from the first year post-partum from six women (1-6) were analyzed. Three subjects presumably transmitted *in utero*, as their infants were PCR positive at birth (denoted as “IUT”), while three transmitted through breast milk (infants negative prior to PCR positive at 40 days, denoted as “BM”) [22]. Five women had low CD4 T cells (<200 per ml) and all had $4-5 \log_{10}$ plasma viral levels at the time of delivery (Table 1). All women reported exclusively breastfeeding for four months with no mastitis, elevated sodium levels, or other clinical problems during the study.

Sequence generation and analysis

Breast milk samples were processed at the site within 4 hrs of collection. Aliquots of whole breast milk (1 ml) were centrifuged twice at $1,600 \times g$ for 15 min to remove the cell fraction. Overall, breast milk samples had only <10 to 1000 cells per ml. Roche Amplicor Ultrasensitive HIV-1 Monitor assay was used as previously described, with an additional third spin performed to remove any remaining cells [23]. RNA was extracted from plasma and breast milk (including a DNase step), reverse transcribed to cDNA, and used as template for amplification of the V1-V5 region of the envelope gp120 gene as previously described [24]. This protocol ensured that no proviral DNA was present in the final template. Two independent amplifications were performed for most samples. Products were cloned and sequenced in the Genome Sequence Service Laboratory at the University of Florida. Sequences have been submitted to Genbank (accession numbers pending). Sequences [~1,000] were classified as subtype C using the Rega subtyping tool (www.bioafrica.net/subtypetool/html). Alignments of V1-V5 sequences were performed manually using BioEdit v7.0 and Mega v4.0 [25], gap-stripped, and analyzed to identify recombinant sequences [26] that were removed to reduce distortion of phylogenetic relationships. Co-receptor use was inferred using the PSSM algorithm specific for HIV-1 subtype C [27].

Phylogenetic analysis

Coalescent theory provides a solid theoretical framework to study the ancestral relationships of individuals sampled from a population [28], including gene flow patterns and time to most recent common ancestor (TMRCA). Bayesian methodology allows use of prior information in analyses, as well as estimation of probability distributions for each parameter of interest. In this study, the TMRCA (median and 95% high posterior density intervals, [HPDs]) were estimated in a Bayesian coalescent framework implemented using the BEAST software package 1.5 [29] with a relaxed molecular clock and a general time reversible with gamma distributed rate variation across sites (GTR+G) model of nucleotide substitution. A particular advantage of the BEAST algorithm is the use of heterogeneous samples to estimate evolutionary parameters under both a strict and relaxed molecular clock [30-32]. For each dataset, the maximum clade credibility (MCC) tree was selected from the posterior tree distribution using the program TreeAnnotator version 1.5. The programs are thoroughly described and available for download at: <http://tree.bio.ed.ac.uk>. Maximum likelihood trees were inferred using the GTR+G model, without the assumption of the molecular clock, using PhyML [33]. Genetic diversity and divergence calculations were performed with a

ML-correction model implemented in Mega v4.0 [25] and compared using a random block ANOVA in GraphPad.

Population structure among viruses from different tissues at different time points was assessed by a modified Hudson test [34] (implemented at <http://www.wabi.snv.jussieu.fr/achaz/hudsonstest.html>) with a Bonferroni-corrected alpha-value of 0.001. A parsimony analysis was used to infer the most likely tissue for each internal node of the MCC trees, and compartmentalization was tested by the Slatkin and Maddison test for gene flow [35] using the MacClade version 4 program [36].

Structure Index

A tree Structure Index (SI) was developed to quantify how close the observed tree topology is to a theoretical topology displaying perfect compartmentalization. Under complete structure: 1) sequences sampled from tissue t_i are not intermixed with sequences sampled from tissue t_j ($i \neq j$), and 2) all sequences sampled from t_i share the same common ancestor. Given n tissues, each sequence was assigned a state corresponding to the anatomical origin (i) ($i=1, 2, \dots, n$). A matrix with equal weights between n tissues was then constructed. Using the maximum clade credibility tree, the ancestral state of each internal node was estimated using maximum parsimony, and the number of steps (changed from one tissue to another) on the tree was counted (S_{obs}). Then, a set of 10,000 random trees was generated using the random-joining method, and for each random tree the number of steps was counted. The lower 95% count from this distribution was used as the upper limit of the expected number of steps on a non-random tree (S_{max}). The minimum number of steps (S_{min}) on a perfectly structured tree is simply $n-1$. The SI is: $(S_{\text{max}} - S_{\text{obs}})/(S_{\text{max}} - S_{\text{min}})$ and scaled from 0 (random structure) to 1 (complete structure) [manuscript in preparation].

RESULTS

Phylogenetic reconstruction of virus populations over time

Reconstruction of evolution by phylogenetic trees was performed to determine the population structure of virus in breast milk with respect to virus in peripheral blood. A total of 769 non-recombinant sequences from plasma, right and left breast milk sampled at three time points during the first year post-partum was analyzed for six women, three who transmitted *in utero* [IUT] (Fig. 1, A-C) and three by breast milk [BM] (Fig. 1, D-F). In all individuals, virus populations were organized into multiple major lineages with posterior branch probabilities >0.7 (branches are scaled in time). Moreover, major lineages in each tree were comprised of viral sequences from plasma and right and/or left breast milk (Fig. 1, A-F). To confirm the trends observed in the Bayesian MCC trees, maximum likelihood trees inferred under a non-clock model also revealed similar major phylogenetic patterns with deep lineages and clustering of sequences from different tissues within lineages (data not shown).

Even though shallow lineages comprised of sequences from only one type of tissue were identified occasionally in every subject, overall tree structures indicated that viruses in breast milk and plasma displayed little temporal structure, but were generally mixed over the course of time. No persistent independent evolution of breast milk virus, as predicted by a compartmentalization scenario, was apparent. Virus populations displayed no phylogenetic structure between breast milk and plasma, which is most consistent with a model of panmixia (complete gene flow).

Estimated time and tissue of origin of the Most Recent Common Ancestor (MRCA)

A particular strength of the Bayesian analysis used in this study is the ability to estimate the timing of internal nodes in the phylogeny, which by definition are ancestral and existed prior to the sampled sequences. This methodology therefore does not require the actual ancestral sequences to be included. A reasonable expectation for the origin of virus in breast milk was that virus(es) were introduced no earlier than lactogenesis, 4 to 5 months prepartum. The time of the MRCA (i.e. the root of the tree indicated by a circle in Fig. 1), estimated by using a relaxed molecular clock with the Bayesian framework, ranged from 9 months to 111 months before delivery (Fig. 1, A-F). The lower confidence limit was about 4 to 6 months, close to the onset of breast milk production for three patients (6IUT, 4BM, and 3BM). In contrast, the lower confidence limit in the other three subjects was between 17 to 45 months before delivery.

The anatomical origin for the MRCA of the divergent lineages within each individual was inferred. Among women who transmitted during labor and delivery, the anatomical origin of the divergent lineages was ambiguous for two women (1IUT and 6IUT), while plasma virus was the source in one case (2IUT). In contrast, among the three women who transmitted by BM, the most parsimonious tissue of origin for the MRCA was breast milk. Since the breast milk was not an anatomical site until a few months before birth, results suggest that the origin of the divergent viral lineages must have been in an unsampled tissue.

Spatial Structure

HIV-1 compartmentalization is defined by viral variants that replicate and evolve independently in anatomical tissues or compartments without contributing to systemic flow of viral genes [2]. Compartmentalization would be evident in the phylogeny as several distinct clades, each containing most (or all) the sequences from a specific tissue. Although trees failed to indicate that breast milk was a virus compartment, a suite of analytical tools to more rigorously test the hypothesis was applied [20]. A traditional phylogenetic measure of compartmentalization, the modified Slatkin-Maddison test, was applied to viruses in two tissues (breast milk vs. plasma) or three tissues (right breast milk, left breast milk, and plasma) (Table 2). In all subjects, the number of changes was less than the lower 5% value of expected changes under the random distribution model. Although a traditional interpretation of the result is that the sequences demonstrate significant compartmentalization, this statistic fails to provide any quantification on the extent of compartmentalization.

To quantify compartmentalization, a novel statistic, the tree Structure Index (SI), was developed. The SI quantifies how close the observed tree topology is to a theoretical topology displaying perfect compartmentalization. The SI is scaled from 0 (complete absence of compartmentalization) to 1 (perfect compartmentalization). When breast milk and plasma virus were compared, SI ranged from 14% to 50% among the subjects, indicating little evidence for compartmentalization (Table 2). When three tissues (left and right breast) were considered, no additional information was revealed. Overall, spatial structuring of sequences between different tissues that might indicate viral compartmentalization was weak and observed sporadically and stochastically, consistent with a largely panmictic population.

Population structure was subsequently evaluated by a non-phylogenetic method that considered both time and tissue of sampling. Even though significant population structure between breast milk and plasma viruses was identified occasionally in all subjects, no trend persisted over time (Table 2). Results suggest that a restricted virus population may develop transiently in one or the other breast milk sample, although viruses from each breast milk

sample are largely shared with contemporaneous plasma, similar to results from tree phylogenies.

Diversity and divergence of virus populations

Nucleotide diversity for all sequences in each tissue was similar in 1IUT, 6IUT, 4BM, but significantly different between plasma and breast milk in 2IUT, 5BM, and 3BM (Fig. 2A). Divergence from the MRCA was also different between plasma and breast milk for 2IUT and 5BM (Fig. 2B), reflecting deep lineages of multiple viral populations. Analysis of diversity/divergence for each timepoint uncovered no additional information (data not shown). Overall, using a simple measure of mean diversity/divergence failed to capture the complexity of the underlying population structure.

Population expansion

The number of lineages inferred to be present during the entire pre-partum period based on the MCC tree was calculated for each subject and reported as the proportion of the total number of lineages present at delivery (Fig. 3A). In five subjects, the number of lineages expanded exponentially during the six months immediately pre-partum, a time frame corresponding with the onset of lactogenesis I. In one subject (5BM), the period of expansion appears much earlier, although the confidence interval associated with the estimated time to the MRCA for this individual spanned 184 months; therefore, increase in the number of lineages could have occurred much closer to the time of pregnancy/lactation.

To assess the possibility that the population expansion resulted from a change in co-receptor usage, the PSSM algorithm was used to infer R5 or X4 coreceptor preference for each sequence. Envelopes predicted to use the X4 coreceptor were found in all subjects (Fig. 3B). In four subjects, viruses predicted to use X4 comprised <10% of viruses from only one time point. In contrast, 100% of viruses in both tissues from subject 4BM were classified as X4 at all time points, while in subject 2IUT, predicted X4 envelopes represented 90% of plasma sequences and 70% of breast milk sequences across time ($p < 0.05$). Overall the expansion of viral populations was independent of changes in coreceptor genotype.

DISCUSSION

Although numerous tissues, including semen [37,38], brain and cerebrospinal fluid [39-42], and genital tract [43-47], have been proposed to serve as viral compartments, cross-sectional study designs and low viral diversity in many tissues [48] confound interpretation and make conclusive assignment of viral population structure problematic. Novel aspects of our study of viral compartmentalization include longitudinal and simultaneously sampled populations of high levels of virus, which provide significant advantages to establishing the persistence of evolutionary trends. Results clearly show that any compartmentalization between breast milk and plasma is evident only sporadically and stochastically. Overall, the evolutionary pattern of viral populations in the breast milk fails to meet criteria for a virologic compartment. A model of panmixia (complete gene flow) between the breast milk and plasma over the first year post-partum best explains the results. The exchange of virus between tissues, which begins immediately after birth [20], thus persists throughout the first year. As none of the women in our study showed elevated sodium levels or evidence of epithelial barrier breakdown, on-going migration of the virus between the breast milk and the plasma by the paracrine pathway is unlikely, indicating that other intra-cellular pathways are involved.

Unexpected results from this longitudinal study include the persistence of strong population structure (multiple deep lineages) in both plasma and breast milk virus; the time of the

MRCA preceding the onset of lactogenesis I; and the origin of the MRCA in the breast milk for all of the BM transmitters. Because breast milk becomes an actual site for viral replication only during lactogenesis, the original infecting virus cannot have derived from this tissue; yet, results are inconsistent with a recent origin from the plasma virus. Taken together the findings implicate virus[es] from a different tissue (not sampled in this study) that migrate to the breast milk and establish an infection at this site.

One potential anatomical origin of the initial breast milk viral lineage is infected epithelial cells in the mammary gland [49], as in the case of mouse mammary tumor virus [50]; however, epithelial cells lack the CD4 receptor and probably do not comprise a significant portion of HIV-1 infected cells. A more likely origin of HIV-1 in breast milk is infected cells from gut-associated lymphoid tissue (GALT) [16,51,52]. Cellular markers indicate that T cells in human breast milk migrate from tissues other than the PBMC [53-55], while both rodent and human studies show that some lymphocytes in breast milk originate from the gut [9-13]. The predominant lymphocytes in the breast milk are CCR5-expressing memory CD4 T-cells [55], although macrophages are also a major cell type in breast milk [56], and a likely source in breast milk of HTLV-1 infection [57]. Even though a majority of infected lymphocytes turn over within days [58,59], macrophages survive HIV-1 infection for weeks to months [60], express CXCR4 and CCR5 co-receptors [61], and play a prominent role as a HIV-1 reservoir in the gastrointestinal mucosa [62]. Significantly, viruses predicted to use the CXCR4 coreceptor were found in breast milk tissue in five of the six subjects. Therefore, GALT-derived macrophages or T cells infected with an early HIV variant may migrate to the breast during lactogenesis and establish an infection. Subsequently, these newly introduced variants migrate to the periphery, in which a distinct set of viral variants is also circulating. The exchange of distinct viral variants between the breast milk and the plasma would result in the mixture of multiple divergent lineages with no compartmentalization.

Under this hypothesis, proviral DNA in breast milk cells, recently introduced from a different tissue, would be expected to show a high divergence with respect to the plasma virus. Unfortunately, the cellular fraction of breast milk was not available for the samples in this study. However, in a recent cross-sectional study of plasma, breast milk RNA and DNA population sequences of the reverse transcriptase and protease genes, median genetic distances between cell-free and cell-associated viruses in breast milk were greater than between plasma and cell-free virus [63]. These results are consistent with the hypothesis that distinct viral lineages in breast milk are introduced from unsampled anatomical sites via cellular migration.

The only difference between subjects who had transmitted through labor and delivery versus breastfeeding was the inferred tissue origin of the MRCA: breast milk for all three BM transmitters, and either plasma or undetermined for the IUT transmitters. However, distinctions between transmitting groups should be interpreted with caution due to the small number of individuals. Although a natural comparison population would be women who breastfed for a substantial period and failed to transmit during either IUT or breastfeeding, breast milk viral load [64] and viral DNA [65] are strongly correlated with transmission; therefore a lack of detectable virus limits the possibility to study viral evolution. Unfortunately, limited samples from the infants rendered the analysis of transmitted viral variants impractical. All women were at an advanced stage of disease, as evidenced by the detection of putative X4 viruses and low CD4 counts. The observation of predicted X4 viruses in all six women might reflect that the subjects were at an advanced stage of disease, as evidenced by low CD4 counts and high plasma viral levels. Although the reported frequency of infection by HIV-1 subtype C X4 variants varies from 0 to 30% depending on geographic location, the frequency appears to be higher in Africa in more recent studies, possibly suggesting an evolving epidemic on this continent [66-69].

Results from our study demonstrate that breast milk, similar to other transmitting tissues, does not provide a compartment for virus during the relatively short time of lactation. In general, evidence that any transmitting tissue provides a restricted source of viral genes/compartment is weak [48,70]. In fact, the brain is the only anatomical site clearly shown to act as a viral compartment/reservoir [39-42,71], although as a non-transmitting tissue the impact of this viral population on new infections is probably limited. While there is no doubt that long-lived cellular sources of virus must exist by virtue of the emergence of an early variant after cessation of antiretroviral therapy [72-74], conclusive identification of the source is unclear. The diverse lineages that emerge during lactation/pregnancy could arise from the same long-lived cellular population involved in post-therapy viral dynamics, or an entirely different anatomical site. Future studies including more individuals/anatomical sites are necessary to investigate the possibility that divergent virus populations emerge during immunological perturbations during the course of natural infection.

Acknowledgments

RRG, CJM, GA, and MMG conceived of the study design. RRG, AL, and KN generated data. WDD, MSinkala, CK, CJM, DMT, and LK assisted in interpretation of data. RRG and MSalemi designed the analysis plan, performed the analyses, and developed novel analysis methods. RRG and MMG wrote the manuscript. The study was supported by grants from the National Institutes of Health (NIH) Eunice Kennedy Shriver National Institute of Child Health and Human Development, (R01 HD39611, R01 HD40777), the National Institute of Allergy and Infectious Disease (R01 AI065265), the National Cancer Institute (T-32 CA09126) and the International Maternal Pediatric Adolescent AIDS Clinical Trials Group U01 AI068632; the University of Florida Stephany W. Holloway University Chair for AIDS Research; the Laura McClamma Endowment for Pediatric Immune Deficiency; the Florida Center for AIDS Research; and the Center for Research in Pediatric Immune Deficiency. GMA is recipient of the Elizabeth Glaser Pediatric AIDS Foundation (EGPAF) Scientist Award.

References

1. Goodenow M, Collman R. HIV-1 coreceptor preference is distinct from target cell tropism: a dual-parameter nomenclature to define viral phenotypes. *J Leukoc Biol.* 2006; 80:965–972. [PubMed: 16923919]
2. Nickle DC, Jensen MA, Shriner D, Brodie SJ, Frenkel LM, Mittler JE, et al. Evolutionary indicators of human immunodeficiency virus type 1 reservoirs and compartments. *J Virol.* 2003; 77:5540–5546. [PubMed: 12692259]
3. Kieffer T, Finucane M, Nettles R, Quinn T, Broman K, Ray S, et al. Genotypic analysis of HIV-1 drug resistance at the limit of detection: virus production without evolution in treated adults with undetectable HIV loads. *J Infect Dis.* 2004; 189:1452–1465. [PubMed: 15073683]
4. Persaud D, Pierson T, Ruff C, Finzi D, Chadwick K, Margolick J, et al. A stable latent reservoir for HIV-1 in resting CD4(+) T lymphocytes in infected children. *J Clin Invest.* 2000; 105:995–1003. [PubMed: 10749578]
5. Aldrovandi G, Kuhn L. What babies and breasts can teach us about natural protection from HIV infection. *J Infect. Dis.* in press.
6. Brandtzaeg P. The mucosal immune system and its integration with the mammary glands. *J Pediatr.* 2010; 156:S8–15. [PubMed: 20105666]
7. Neville, MC.; Neifert, MR. *Lactation: Physiology, Nutrition, and Breastfeeding.* Plenum Press; New York: 1983.
8. Neville MC, Allen JC, Archer PC, Casey CE, Seacat J, Keller RP, et al. Studies in human lactation: milk volume and nutrient composition during weaning and lactogenesis. *Am J Clin Nutr.* 1991; 54:81–92. [PubMed: 2058592]
9. Allardyce R, Shearman D, McClelland D, Marwick K, Simpson A, Laidlaw R. Appearance of specific colostrum antibodies after clinical infection with *Salmonella typhimurium*. *Br Med J.* 1974; 3:307–309. [PubMed: 4606237]
10. Fishaut M, Murphy D, Neifert M, McIntosh K, Ogra P. Bronchomammary axis in the immune response to respiratory syncytial virus. *J Pediatr.* 1981; 99:186–191. [PubMed: 7252673]

11. Goldblum R, Ahlstedt S, Carlsson B, Hanson L, Jodal U, Lidin-Janson G, et al. Antibody-forming cells in human colostrum after oral immunisation. *Nature*. 1975; 257:797–798. [PubMed: 1102990]
12. Roux M, McWilliams M, Phillips-Quagliata J, Weisz-Carrington P, Lamm M. Origin of IgA-secreting plasma cells in the mammary gland. *J Exp Med*. 1977; 146:1311–1322. [PubMed: 925605]
13. Weisz-Carrington P, Roux M, McWilliams M, PHILLIPS-Quagliata J, Lamm M. Organ and isotype distribution of plasma cells producing specific antibody after oral immunization: evidence for a generalized secretory immune system. *J Immunol*. 1979; 123:1705–1708. [PubMed: 479596]
14. Georgeson JC, Filteau SM. Physiology, immunology, and disease transmission in human breast milk. *AIDS Patient Care STDS*. 2000; 14:533–539. [PubMed: 11054937]
15. Goldman AS. The immune system of human milk: antimicrobial, antiinflammatory and immunomodulating properties. *Pediatr Infect Dis J*. 1993; 12:664–671. [PubMed: 8414780]
16. Sabbaj S, Edwards B, Ghosh M, Semrau K, Cheelo S, Thea D, et al. Human immunodeficiency virus-specific CD8(+) T cells in human breast milk. *J Virol*. 2002; 76:7365–7373. [PubMed: 12097549]
17. Becquart P, Chomont N, Roques P, Ayouba A, Kazatchkine MD, Belec L, et al. Compartmentalization of HIV-1 between breast milk and blood of HIV-infected mothers. *Virology*. 2002; 300:109–117. [PubMed: 12202211]
18. Becquart P, Courgnaud V, Willumsen J, Van de Perre P. Diversity of HIV-1 RNA and DNA in breast milk from HIV-1-infected mothers. *Virology*. 2007; 363:256–260. [PubMed: 17335864]
19. Henderson GJ, Hoffman NG, Ping LH, Fiscus SA, Hoffman IF, Kitrinos KM, et al. HIV-1 populations in blood and breast milk are similar. *Virology*. 2004; 330:295–303. [PubMed: 15527854]
20. Heath L, Conway S, Jones L, Semrau K, Nakamura K, Walter J, et al. Restriction of HIV-1 genotypes in breast milk does not account for the population transmission genetic bottleneck that occurs following transmission. *PlosOne*. 2010
21. Kuhn L, Aldrovandi G, Sinkala M, Kankasa C, Semrau K, Mwiya M, et al. Effects of early, abrupt weaning on HIV-free survival of children in Zambia. *N Engl J Med*. 2008; 359:130–141. [PubMed: 18525036]
22. Fox M, Brooks D, Kuhn L, Aldrovandi G, Sinkala M, Kankasa C, et al. Reduced mortality associated with breast-feeding-acquired HIV infection and breast-feeding among HIV-infected children in Zambia. *J Acquir Immune Defic Syndr*. 2008; 48:90–96. [PubMed: 18344878]
23. Ghosh M, Kuhn L, West J, Semrau K, Decker D, Thea D, et al. Quantitation of human immunodeficiency virus type 1 in breast milk. *J Clin Microbiol*. 2003; 41:2465–2470. [PubMed: 12791866]
24. Tuttle DL, Anders CB, Aquino-De Jesus MJ, Poole PP, Lamers SL, Briggs DR, et al. Increased replication of non-syncytium-inducing HIV type 1 isolates in monocyte-derived macrophages is linked to advanced disease in infected children. *AIDS Res Hum Retroviruses*. 2002; 18:353–362. [PubMed: 11897037]
25. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. 2007; 24:1596–1599. [PubMed: 17488738]
26. Salemi M, Gray R, Goodenow M. An exploratory algorithm to identify intra-host recombinant viral sequences. *Mol Phylogenet Evol*. 2008; 49:618–628. [PubMed: 18801446]
27. Jensen M, Coetzer M, van 't Wout A, Morris L, Mullins J. A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences. *J Virol*. 2006; 80:4698–4704. [PubMed: 16641263]
28. Kingman J. *The Coalescent. Stochastic Processes and Their Applications*. 1982; 13:235–248.
29. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007; 7:214. [PubMed: 17996036]
30. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*. 2002; 161:1307–1320. [PubMed: 12136032]

31. Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. Measurably evolving populations. *Trends in Ecology and Evolution*. 2003; 18:481–488.
32. Drummond A, Ho S, Phillips M, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006; 4:e88. [PubMed: 16683862]
33. Guindon S, Delsuc F, Dufayard J, Gascuel O. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol*. 2009; 537:113–137. [PubMed: 19378142]
34. Achaz G, Palmer S, Kearney M, Maldarelli F, Mellors JW, Coffin JM, et al. A robust measure of HIV-1 population turnover within chronically infected individuals. *Mol Biol Evol*. 2004; 21:1902–1912. [PubMed: 15215321]
35. Slatkin M, Maddison WP. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics*. 1989; 123:603–613. [PubMed: 2599370]
36. Maddison, WP.; Maddison, DR. *MacClade version 4: Analysis of phylogeny and character evolution*. Sinauer Associates; Sunderland, MA: 2003.
37. Kiessling A, Fitzgerald L, Zhang D, Chhay H, Brettler D, Eyre R, et al. Human immunodeficiency virus in semen arises from a genetically distinct virus reservoir. *AIDS Res Hum Retroviruses*. 1998; 14(Suppl 1):S33–41. [PubMed: 9581882]
38. Coombs R, Speck C, Hughes J, Lee W, Sampoleo R, Ross S, et al. Association between culturable human immunodeficiency virus type 1 (HIV-1) in semen and HIV-1 RNA levels in semen and blood: evidence for compartmentalization of HIV-1 between semen and blood. *J Infect Dis*. 1998; 177:320–330. [PubMed: 9466517]
39. Chang J, Jozwiak R, Wang B, Ng T, Ge Y, Bolton W, et al. Unique HIV type 1 V3 region sequences derived from six different regions of brain: region-specific evolution within host-determined quasispecies. *AIDS Res Hum Retroviruses*. 1998; 14:25–30. [PubMed: 9453248]
40. Donaldson Y, Bell J, Holmes E, Hughes E, Brown H, Simmonds P. In vivo distribution and cytopathology of variants of human immunodeficiency virus type 1 showing restricted sequence variability in the V3 loop. *J Virol*. 1994; 68:5991–6005. [PubMed: 7545945]
41. Salemi M, Lamers SL, Yu S, de Oliveira T, Fitch WM, McGrath MS. Phylodynamic analysis of human immunodeficiency virus type 1 in distinct brain compartments provides a model for the neuropathogenesis of AIDS. *J Virol*. 2005; 79:11343–11352. [PubMed: 16103186]
42. Ohagen A, Devitt A, Kunstman KJ, Gorry PR, Rose PP, Korber B, et al. Genetic and functional analysis of full-length human immunodeficiency virus type 1 env genes derived from brain and blood of patients with AIDS. *J Virol*. 2003; 77:12336–12345. [PubMed: 14581570]
43. Poss M, Rodrigo A, Gosink J, Learn G, de Vange Panteleeff D, Martin HJ, et al. Evolution of envelope sequences from the genital tract and peripheral blood of women infected with clade A human immunodeficiency virus type 1. *J Virol*. 1998; 72:8240–8251. [PubMed: 9733867]
44. Ping LH, Cohen MS, Hoffman I, Vernazza P, Seillier-Moisewitsch F, Chakraborty H, et al. Effects of genital tract inflammation on human immunodeficiency virus type 1 V3 populations in blood and semen. *J Virol*. 2000; 74:8946–8952. [PubMed: 10982338]
45. Pillai SK, Good B, Pond SK, Wong JK, Strain MC, Richman DD, et al. Semen-specific genetic characteristics of human immunodeficiency virus type 1 env. *J Virol*. 2005; 79:1734–1742. [PubMed: 15650198]
46. Kemal KS, Foley B, Burger H, Anastos K, Minkoff H, Kitchen C, et al. HIV-1 in genital tract and plasma of women: compartmentalization of viral sequences, coreceptor usage, and glycosylation. *Proc Natl Acad Sci U S A*. 2003; 100:12972–12977. [PubMed: 14557540]
47. Philpott S, Burger H, Tsoukas C, Foley B, Anastos K, Kitchen C, et al. Human immunodeficiency virus type 1 genomic RNA sequences in the female genital tract and blood: compartmentalization and intrapatient recombination. *J Virol*. 2005; 79:353–363. [PubMed: 15596829]
48. Bull M, Learn G, Genowati I, McKernan J, Hitti J, Lockhart D, et al. Compartmentalization of HIV-1 within the female genital tract is due to monotypic and low-diversity variants not distinct viral populations. *PLoS One*. 2009; 4:e7122. [PubMed: 19771165]
49. Toniolo A, Serra C, Conaldi PG, Basolo F, Falcone V, Dolei A. Productive HIV-1 infection of normal human mammary epithelial cells. *Aids*. 1995; 9:859–866. [PubMed: 7576319]

50. Minamishima I, Ueda K, Minematsu T, Minamishima Y, Umemoto M, Take H, et al. Role of breast milk in acquisition of cytomegalovirus infection. *Microbiol Immunol.* 1994; 38:549–552. [PubMed: 7968688]
51. Kourtis AP, Butera S, Ibegbu C, Beled L, Duerr A. Breast milk and HIV-1: vector of transmission or vehicle of protection? *Lancet Infect Dis.* 2003; 3:786–793. [PubMed: 14652204]
52. Manning LS, Parmely MJ. Cellular determinants of mammary cell-mediated immunity in the rat. I. The migration of radioisotopically labeled T lymphocytes. *J Immunol.* 1980; 125:2508–2514. [PubMed: 6968787]
53. Bertotto A, Castellucci G, Fabietti G, Scalise F, Vaccaro R. Lymphocytes bearing the T cell receptor gamma delta in human breast milk. *Arch Dis Child.* 1990; 65:1274–1275. [PubMed: 2147370]
54. Ullrich R, Schieferdecker HL, Ziegler K, Riecken EO, Zeitz M. gamma delta T cells in the human intestine express surface markers of activation and are preferentially located in the epithelium. *Cell Immunol.* 1990; 128:619–627. [PubMed: 2113432]
55. Bertotto A, Gerli R, Fabietti G, Crupi S, Arcangeli C, Scalise F, et al. Human breast milk T lymphocytes display the phenotype and functional characteristics of memory T cells. *Eur J Immunol.* 1990; 20:1877–1880. [PubMed: 2120066]
56. Goldman AS, Chheda S, Garofalo R. Evolution of immunologic functions of the mammary gland and the postnatal development of immunity. *Pediatr Res.* 1998; 43:155–162. [PubMed: 9475278]
57. Takeuchi H, Takahashi M, Norose Y, Takeshita T, Fukunaga Y, Takahashi H. Transformation of breast milk macrophages by HTLV-I: implications for HTLV-I transmission via breastfeeding. *Biomed Res.* 2010; 31:53–61. [PubMed: 20203420]
58. Ho D, Neumann A, Perelson A, Chen W, Leonard J, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature.* 1995; 373:123–126. [PubMed: 7816094]
59. Markert M, Hicks C, Bartlett J, Harmon J, Hale L, Greenberg M, et al. Effect of highly active antiretroviral therapy and thymic transplantation on immunoreconstitution in HIV infection. *AIDS Res Hum Retroviruses.* 2000; 16:403–413. [PubMed: 10772526]
60. Aquaro S, Bagnarelli P, Guenci T, De Luca A, Clementi M, Balestra E, et al. Long-term survival and virus production in human primary macrophages infected by human immunodeficiency virus. *J Med Virol.* 2002; 68:479–488. [PubMed: 12376954]
61. Goodenow MM, Rose SL, Tuttle DL, Sleasman JW. HIV-1 fitness and macrophages. *J Leukoc Biol.* 2003; 74:657–666. [PubMed: 12960234]
62. Smith P, Meng G, Salazar-Gonzalez J, Shaw G. Macrophage HIV-1 infection and the gastrointestinal tract reservoir. *J Leukoc Biol.* 2003; 74:642–649. [PubMed: 12960227]
63. Andreotti M, Galluzzo C, Guidotti G, Germano P, Altan A, Pirillo M, et al. 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.* 2009; 25:707–711. [PubMed: 19552594]
64. Semrau K, Ghosh M, Kankasa C, Sinkala M, Kasonde P, Mwiya M, et al. Temporal and lateral dynamics of HIV shedding and elevated sodium in breast milk among HIV-positive mothers during the first 4 months of breast-feeding. *J Acquir Immune Defic Syndr.* 2008; 47:320–328. [PubMed: 18398972]
65. Rousseau CM, Nduati RW, Richardson BA, John-Stewart GC, Mbori-Ngacha DA, Kreiss JK, et al. Association of levels of HIV-1-infected breast milk cells and risk of mother-to-child transmission. *J Infect Dis.* 2004; 190:1880–1888. [PubMed: 15499546]
66. Raymond S, Delobel P, Mavigner M, Ferradini L, Cazabat M, Souyris C, et al. Prediction of HIV type 1 subtype C tropism by genotypic algorithms built from subtype B viruses. *J Acquir Immune Defic Syndr.* 2010; 53:167–175. [PubMed: 19996764]
67. Cecilia D, Kulkarni S, Tripathy S, Gangakhedkar R, Paranjape R, Gadkari D. Absence of coreceptor switch with disease progression in human immunodeficiency virus infections in India. *Virology.* 2000; 271:253–258. [PubMed: 10860879]

68. Connell B, Michler K, Capovilla A, Venter W, Stevens W, Papathanasopoulos M. Emergence of X4 usage among HIV-1 subtype C: evidence for an evolving epidemic in South Africa. *AIDS*. 2008; 22:896–899. [PubMed: 18427209]
69. Ping L, Nelson J, Hoffman I, Schock J, Lamers S, Goodman M, et al. Characterization of V3 sequence heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: underrepresentation of X4 variants. *J Virol*. 1999; 73:6271–6281. [PubMed: 10400718]
70. Kemal K, Burger H, Mayers D, Anastos K, Foley B, Kitchen C, et al. HIV-1 drug resistance in variants from the female genital tract and plasma. *J Infect Dis*. 2007; 195:535–545. [PubMed: 17230413]
71. Zhang L, Rowe L, He T, Chung C, Yu J, Yu W, et al. Compartmentalization of surface envelope glycoprotein of human immunodeficiency virus type 1 during acute and chronic infection. *J Virol*. 2002; 76:9465–9473. [PubMed: 12186928]
72. Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio F, Yassine-Diab B, et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat Med*. 2009; 15:893–900. [PubMed: 19543283]
73. Chun T, Carruth L, Finzi D, Shen X, DiGiuseppe J, Taylor H, et al. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature*. 1997; 387:183–188. [PubMed: 9144289]
74. Finzi D, Hermankova M, Pierson T, Carruth L, Buck C, Chaisson R, et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science*. 1997; 278:1295–1300. [PubMed: 9360927]

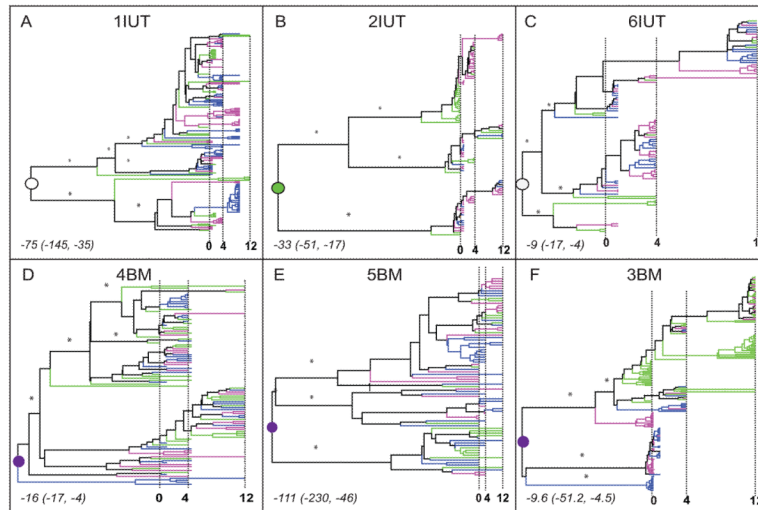


Figure 1. Phylogenetic history of plasma and breast milk virus

Six patients are designated based on mode of transmission to infants: IUT, *in utero*; BM, breast milk (panels A-F). A total of 769 non-recombinant sequences (1IUT: 163 sequences; 2IUT: 102; 3BM: 199; 4BM: 102; 5BM: 81; 6IUT: 122) were analyzed. Bayesian maximum clade credibility (MCC) phylogenies were inferred using the GTR+G model of nucleotide substitution, constant population size, and relaxed clock in BEASTv.1.4.7. Branches are scaled in time with the tips of branches sampled from the same date aligned according to the scale in months at the bottom of each panel. Vertical dashed lines indicate 0, 4, and 12 months after delivery. Terminal branches are colored according to the tissue of sampling (green = plasma, pink = right breast milk, blue = left breast milk). Internal branches are colored according to the same scheme when the tissue assignment was unambiguous; otherwise colored black. Asterisks designate posterior branch probabilities >0.7 for major internal branches. Circles indicate the internal node (root node) that is the most recent common ancestor (MRCA) of the entire population. The tissue of origin for the MRCA was determined by parsimony analysis (open = undetermined, green = plasma, purple = breast milk). The time of origin for the MRCA (the median number of months prior to delivery and 95% high posterior density intervals) are noted in italics in the lower left corner for each subject.

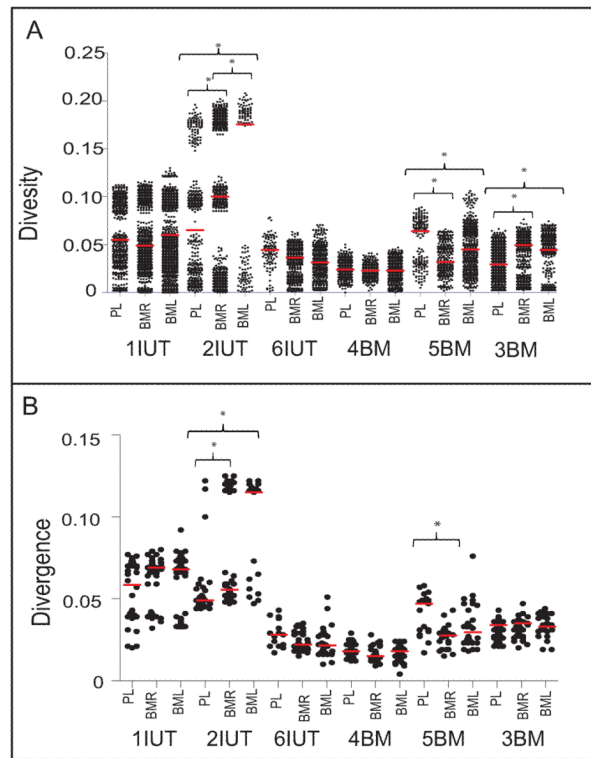


Figure 2. Diversity and divergence of plasma and breast milk virus are similar
A. Pairwise diversity and **B.** pairwise divergence from the inferred MRCA for plasma, left and right breast milk samples were calculated in MEGA v4.0 with the maximum likelihood + G correction. The median for each group of sequences is indicated with a horizontal red line. The y-axis is scaled in substitutions/site, defined as the number of nucleotide differences between two sequences divided by the sequence length. Asterisks indicate significantly different medians between tissues in the same patient ($p < 0.0019$).

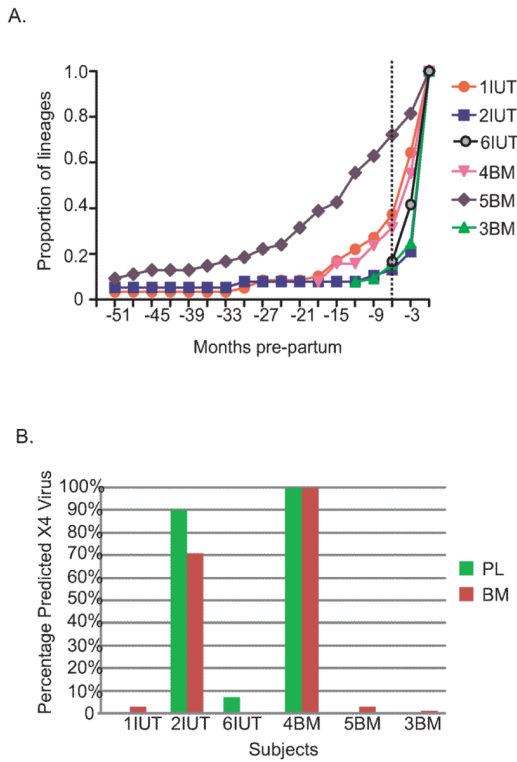


Figure 3. Exponential growth of viral lineages and predicted X4 usage
A. The proportion of lineages (based on the total number of lineages present at delivery, on y-axis) inferred from the MCC trees (Fig. 1) is plotted for each three month period pre-partum (x-axis) for six subjects. The timescale is the same as for the MCC trees in Figure 1. The dotted line corresponds with the onset of Lactogenesis I at -6 months pre-partum. **B.** The percentages of the total number of sequences for either plasma or breast milk sequences predicted to use the CXCR4 co-receptor based on the subtype C specific PSSM algorithm are shown as bars. Green bars indicates plasma, red bars indicates breast milk.

Table 1

Patient clinical characteristics for samples used in this study.

Patient ^a	Timepoint ^b	CD4 ^c	Viral Load ^d		
			plasma	BML	BMR
1IUT	Delivery	94	375,319		
	1 month			637	1,538
	2 months		57,162		
	4 months			7,845	3,865
	9 months			4,454	4,966
2IUT	12 months		49,333		
	Delivery	118	211,792		
	1 month			20,207	18,277
	4 months		83,274		257
	12 months		108,722	9,908	6,275
3BM	Delivery	134	21,148		
	1 week			NA ^e	NA
	1 month			159,012	80,925
4BM	4 months		235,196	3,987	10,570
	12 months		152,500	NA	
	Delivery	82	94,873		
5BM	1 month				18,645
	4 months		437,545	1,294	1,882
	12 months		115,517	8,468	NA
	Delivery	196			
6IUT	1 month		102,438	57,316	59,261
	4 months			28,980	15,837
	12 months		108,255	7,677	16,087
	Delivery	291	47,602		
	1 month			22,882	18,863
	4 months		38,083	4,767	11,967
	12 months		149,262	13,463	51,959

^a BM = breast milk transmitter, IUT = *in utero* transmitter

^b Times refer to sampling time (months post-delivery) at which samples were collected

^c CD4+ cell counts were measured at delivery for each patient

^d Viralload data from available samples used in this study; BML = left breast milk; BMR = right breast milk

^e NA = viral load was unavailable although sequences were generated and analyzed in this study

Table 2

Compartmentalization and Population Structure.

Subject	Compartmentalization ^a					Population Structure ^e			
	Plasma, BM		PL, BML, BMR		Timepoint	Plasma vs. BM	BML vs. BMR		
	#steps ^b	5% ^c	SI ^d	#steps					5%
1IUT	20	29	32%	33	64	50%	1st ^f	0.012	0.204
							2-4 months	<10 ⁻¹⁰	<10 ⁻¹⁰
							9-12 months	0.004	<10 ⁻¹⁰
2IUT	12	23	50%	21	36	44%	1st	<10 ⁻¹⁰	0.307
							4 months	0.023	NA ^g
							12 months	0.342	0.266
3BM	22	59	64%	33	68	53%	1st	<10 ⁻¹⁰	<10 ⁻¹⁰
							4 months	0.088	0.333
							12 months	0.023	NA
4BM	22	26	16%	36	43	17%	1st	0.434	NA
							4 months	0.123	0.001
							12 months	0.081	0.183
5BM	13	15	14%	26	32	20%	1st	0.073	0.005
							4 months	NA	0.029
							12 months	0.001	0.007
6IUT	10	12	18%	23	34	34%	1st	0.004	0.304
							4 months	<10 ⁻¹⁰	<10 ⁻¹⁰
							12 months	0.528	0.001

^aCompartmentalization was measured using the Slatkin-Maddison test and the SI for either plasma and breast milk (left and right) or plasma, left and right breast milk.

^bThe observed number of steps on the maximum clade credibility tree.

^cThe lower 5% cutoff of the null distribution.

^dStructure Index value for each patient.

^ePopulation structure assessed using the Hudson method in which values <0.001 are considered significant (in bold).

^fFirst timepoint includes delivery, 1 week and 1 month sequences.

^gNA = not applicable (sequences were not available for one tissue/timepoint).