

## HIV-1 populations in blood and breast milk are similar

Gavin J. Henderson<sup>a,b</sup>, Noah G. Hoffman<sup>a,b</sup>, Li Hua Ping<sup>a</sup>, Susan A. Fiscus<sup>a,b</sup>,  
Irving F. Hoffman<sup>a,c</sup>, Kathryn M. Kitrinos<sup>a,d,1</sup>, Topia Banda<sup>e</sup>, Francis E.A. Martinson<sup>c,e</sup>,  
Peter N. Kazembe<sup>e</sup>, David A. Chilongozi<sup>c,e</sup>, Myron S. Cohen<sup>a,c</sup>, Ronald Swanstrom<sup>a,b,\*</sup>

<sup>a</sup>UNC Center for AIDS Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7295, United States

<sup>b</sup>Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7295, United States

<sup>c</sup>Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7295, United States

<sup>d</sup>Curriculum of Genetics and Molecular Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7295, United States

<sup>e</sup>Lilongwe Central Hospital, Lilongwe, Malawi

Received 18 June 2004; returned to author for revision 19 August 2004; accepted 8 September 2004

Available online 6 October 2004

### Abstract

Mother-to-child transmission (MTCT) of human immunodeficiency virus type 1 (HIV-1) through breast milk is a significant mechanism of infection in many regions of the world. We compared the HIV-1 populations in paired blood and breast milk samples using a heteroduplex tracking assay (HTA) for the V1/V2 regions of *env* (V1/V2-HTA). V1/V2-HTA patterns were similar in the eight pairs of samples for which adequate template sampling could be demonstrated. No unique variants existed in either compartment, and differences detected in the relative abundance of variants between compartments were small, occurred among low abundance variants, and were not statistically significant. We also documented the impact of template sampling as a limiting feature in comparing two viral populations. The absence of unique variants and the lack of significant differences in the relative abundance of variants between these compartments support the conclusion that viruses in the blood plasma and breast milk are well equilibrated.

© 2004 Elsevier Inc. All rights reserved.

**Keywords:** HIV; Compartmentalization; Breast milk

### Introduction

Vertical transmission of human immunodeficiency virus type-1 (HIV-1) from mother to child is a common source of pediatric HIV infection (United Nations Children's Fund Mother-to-Child Transmission of HIV Fact Sheet [[http://www.unicef.org/publications/index\\_4414.html](http://www.unicef.org/publications/index_4414.html)]), with reported frequencies of 13–48% in the absence of antire-

troviral prophylaxis (Thorne and Newell, 2000). Mother-to-child transmission (MTCT) can occur transplacentally during pregnancy, during delivery by exposure to infectious maternal blood or mucosal surfaces, or after delivery through HIV-containing breast milk. Breast feeding has been identified as an important route for vertical transmission of HIV-1, particularly in regions of the world where interventions such as antiretroviral therapy and infant formula are not widely available (Dunn et al., 1992; John et al., 2001; Nduati et al., 2000; Oxtoby, 1988; Van de Perre et al., 1993). Both cell-free virus particles and HIV-1-infected cells have been found in breast milk (Lewis et al., 1998; Van de Perre et al., 1993) and have been implicated in oral HIV-1 transmission (Baba et al., 1996; Baron et al., 2000).

Tissue-specific HIV-1 variants have been identified in various anatomical compartments, including the central

\* Corresponding author. Mailing address: CB7295, Room 22-062 Lineberger Building, UNC Center for AIDS Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7295. Fax: +1 919 966 8212.

E-mail address: [risunc@med.unc.edu](mailto:risunc@med.unc.edu) (R. Swanstrom).

<sup>1</sup> Current address: Department of International Clinical Virology, Glaxo-Smith-Kline, RTP, NC 27709-3398, United States.

nervous system, the genitourinary tract, and the gastrointestinal tract (for review, see McGrath et al., 2001). Compartmentalization of viral genotypic variants is thought to occur due to constraints on viral entry and replication, target cell differences, and differing immune pressures in distinct anatomical regions that result in the independent selection of subsets of the virus population for continued replication. The potential exists for HIV-1 to exhibit compartmentalization between breast milk and other compartments as well. A recent study by Becquart et al. (2002) suggested that distinct variants do exist in comparing blood plasma and breast milk. Confirming whether this is the case is important to understanding the virologic and immunologic factors that affect postnatal MTCT.

As part of an effort to understand the determinants of postnatal vertical transmission, we have examined HIV-1 populations in blood plasma and breast milk in a cross-sectional analysis to assess whether compartmentalization occurs. In order to determine whether tissue-specific viral variants are present, we have compared HIV-1 populations from these compartments using the heteroduplex tracking assay (HTA), a sensitive and quantitative method to detect nucleotide sequence variability in complex populations (Delwart et al., 1993, 1994; Kitrinis et al., 2003; Nelson et al., 1997). HTA evaluation of HIV-1 populations is performed by creating heteroduplexes between a PCR product generated from viral RNA and a radiolabeled probe DNA. Nucleotide and length mismatches between the two strands induce bends in the duplex that alter its migration rate during native polyacrylamide gel electrophoresis. Genotypic variants, represented as discrete heteroduplex bands in the gel, can then be quantified to compare their relative abundances.

We have used an HTA specific for variable regions 1 and 2 (V1/V2) of the *env* gene (V1/V2-HTA; Kitrinis et al., 2003) to compare HIV-1 populations in paired blood plasma and breast milk samples. The V1/V2 region was chosen as a marker for genetic diversity because of its high degree of variability (Starcich et al., 1986) and the presence of multiple coexisting variants (Kitrinis et al., 2003). The presence of multiple variants and their relative abundance provide a sensitive approximation of the virus population composition when comparing two compartments, and also represent a stringent tool for assessing the quality of template sampling of the virus population. Compartmentalization can be recognized by the appearance of distinct genotypic variants or the selective outgrowth of shared variants, reflecting the phenotypic differences that are selected for during compartmentalization. The high degree of similarity in V1/V2-HTA patterns representing the virus from blood plasma and breast milk suggests a lack of compartmentalization between these two compartments. Quantitative statistical analysis of V1/V2-HTA patterns further demonstrate that any real differences in the

composition of virus populations of blood plasma and breast milk are at most small.

## Results

### *Samples and virus loads*

The women who participated in this study were part of a group of 124 lactating women attending a postnatal clinic at Lilongwe Central Hospital, Lilongwe, Malawi (Hoffman et al., 2003). From this cohort, 33 women were identified with anti-HIV-1 antibodies in their blood. The virus loads of the blood plasma and the skim milk fraction of the breast milk samples from these subjects were determined using the Roche Amplicor HIV-1 Monitor and the NucliSens HIV-1 QT assays, respectively (Fig. 1). Previous comparison suggests little difference in viral RNA measurements by these assays (Dyer et al., 1999). Of these samples, 42% (14 of 33) had no detectable virus in the skim breast milk, while all but one of the plasma samples had detectable levels of virus. A subset of 28 skim breast milk samples and their matched plasma samples was analyzed. We were able to generate a V1/V2 reverse-transcriptase (RT)-PCR product from all 28 of the blood plasma samples, but from only 17 of the 28 breast milk samples. Of the 17 breast milk samples that were able to be amplified, only eight generated reproducible V1/V2-HTA patterns, a measure of efficient RNA template sampling. The ability to generate RT-PCR products demonstrating adequate template sampling largely correlated with the sample virus load. The inconsistent amplification of some breast milk samples raised sampling error as a major challenge in evaluating virus populations,

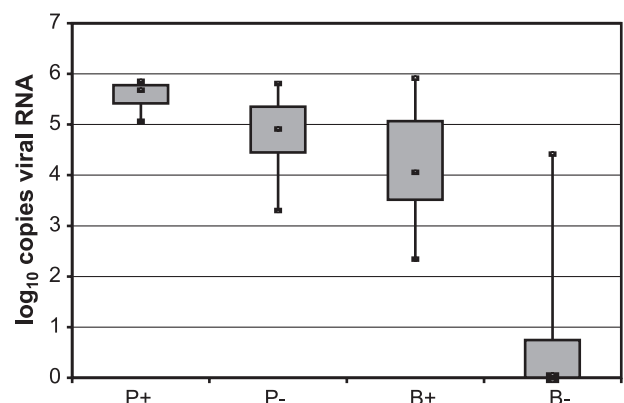


Fig. 1. Subject samples and virus loads. Paired blood and breast milk samples were obtained with informed consent from women attending a postnatal clinic at Lilongwe Central Hospital, Lilongwe, Malawi. Virus load values are plotted as the  $\log_{10}$  of the viral RNA copies per milliliter. The boxes encompass the 25th–75th percentiles with the median values denoted by a dash (–); bars denote the outer quartiles. Virus load values for blood plasma (P) and breast milk (B) samples are separated into the eight paired samples successfully validated for template sampling (+) and the remaining samples (–).

even in those samples where a PCR product could be obtained.

*V1/V2-HTA*

V1/V2-HTA was performed on RT-PCR products from paired blood plasma and breast milk samples using a heterologous probe to increase the sensitivity to small differences between genotypic variants within a population (Kitrinis et al., 2003). Comparison of the viral variants in blood plasma and skim breast milk by V1/V2-HTA showed that all of the major variants could be detected in both compartments (Fig. 2). It is important to note that the low intensity of some lanes does not indicate homogeneity, but rather that different amounts of total PCR product are present in the samples, which is compensated for by normalizing to the total signal in each lane (see below). Samples contained from four to eight heteroduplex bands representing at least 5% of the signal present in the respective sample; other bands comprising minor (less than 5%) populations were also present in some samples. Each subject's heteroduplex pattern was unique, representing viral sequences unique to that subject. As a control to confirm adequate template sampling, duplicate RT-PCR reactions of each sample were compared. Because the reproduction of complex HTA patterns by two independent RT-PCR amplifications is unlikely to occur by chance, we interpret the reproduction of these patterns as an indication that all of the major genotypic variants present were adequately sampled. Only those samples that could be validated in this manner were included in our analysis.

*Heteroduplex quantitation and statistical analysis*

Phosphorimager analysis of V1/V2-HTAs was used to quantitate the relative abundance of each heteroduplex variant in a sample. The relative abundance values thus measured are directly related to the abundance of each genotypic variant present in a sample and therefore represent their frequency in the virus population of that sample. The difference between any two samples was

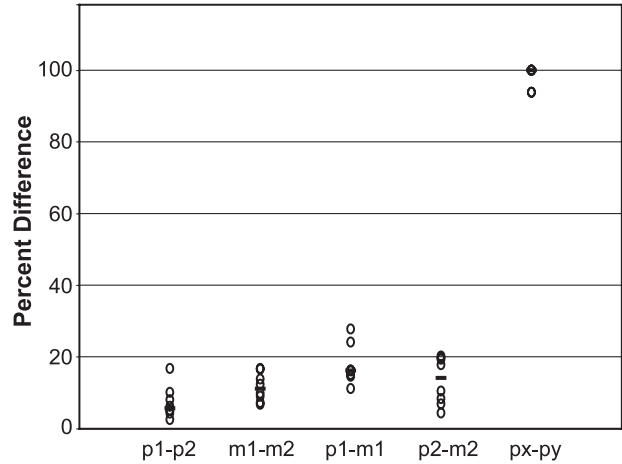


Fig. 3. Percent difference comparison of blood plasma and breast milk samples. Phosphorimager analysis of V1/V2-HTAs was used to determine the relative abundance of each heteroduplex variant in a sample. The difference between any two samples was assessed using the percent change calculation as described. The p1–p2 and m1–m2 values are the difference calculated between separate amplifications of the same plasma or breast milk samples, respectively. The p1–m1 and p2–m2 values are the calculated differences in the abundance of variants between the plasma and breast milk samples from either the first or the second amplification. The px–py values represent the difference between pairs of plasma samples from different subjects as a reference. The median value for all comparisons is denoted by a dash (–).

assessed using the percent change calculation, a method used to characterize the difference between two populations in the context of HTA analysis (Kitrinis et al., 2003; Nelson et al., 2000), with bands present at less than 5% relative abundance being discarded.

To assess the variability inherent in our analysis, the percent difference between the repeated measurements of the plasma samples (p1–p2) and the repeated measurements of the breast milk (m1–m2) was calculated (Fig. 3). Because each sample is taken from the same source, the theoretical difference between the repeated experiments is zero. The median percent difference for the repeated analysis of breast milk samples was 11%, and this was significantly higher than the percent difference for blood plasma (6%) ( $P = 0.032$ ). The greater variability observed in the repeated

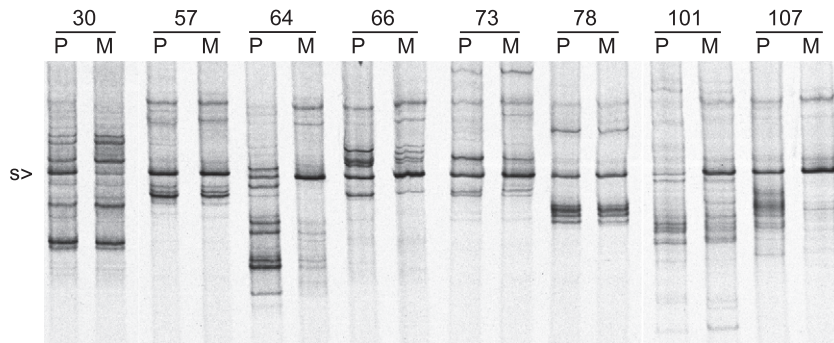


Fig. 2. V1/V2-HTA comparing paired blood plasma and breast milk samples. V1/V2-HTA was performed on HIV-1 sequences from paired blood plasma (P) and breast milk (M) samples using a heterologous HIV-1 clade C V1/V2 env probe. Paired samples are in adjacent lanes, with the subject identifier above them. The single-stranded radiolabeled probe (ss) is present in each lane.

analysis of breast milk samples (reflecting less complete sampling of the viral genotypes) is likely due to lower virus load or reduced efficiency in the RT-PCR step with RNA from this source. As a control, the comparison of two samples from different populations (px–py) gave percent difference values near 100%.

The percent difference between the virus populations in the blood plasma and breast milk was calculated in the same manner for both analyses of these samples (p1–m1, p2–m2; Fig. 3). These differences were small, generally less than 20%, and were not significantly different from one another by a rank sum test ( $P = 0.221$ ), indicating that the percent difference measurement between the two compartments was made consistently.

Finally, to determine whether the difference in the abundance of variants between compartments was significant, the rank sum test was used to compare the two measures of assay variability (p1–p2, m1–m2) and the potential compartmental difference (p1–m1, p2–m2). A significant difference was found between the compartmental percent difference (p1–m2, p2–m2) and the repeated plasma analysis (p1–p2) ( $P = 0.002$ ). However, the percent difference between the repeated breast milk analysis (m1–m2) and the compartmental percent difference measurements was not significant ( $P = 0.065$ ). Therefore, the differences in the relative abundance of HIV-1 variants present in the blood and the breast milk were not significantly greater than those measured for the repeated analysis of the breast milk samples alone. Furthermore, the magnitude of the differences observed are relatively small, with median values in the two plasma-breast milk comparisons of approximately 15% difference, of which at least 10% is likely contributed by assay variability.

As a second measure of the difference in the relative abundance of variants, we compared the ratio of the repeated breast milk measurements (m1:m2) and the averaged ratio of the blood plasma and breast milk measurements (p/m) to the relative abundance of each variant (Fig. 4). If the relative abundance of a variant was fully reproduced between the two analyses, this value would be 1. We found that the largest variability in abundance measurements occurred for rare variants. This is likely due to the fact that variants present at lower relative abundances are more difficult to sample consistently, such that sampling error is more pronounced in measuring these variants. We also note that the distribution of the differences between the plasma and breast milk variants (p:m) and those of the repeated breast milk samples (m1:m2) are similar. This suggests that the small differences we measured between these two compartments are related to those of the repeated measurements and likely stem from experimental variability.

Together, the absence of a significant difference in the identity and relative abundance of variants between blood and breast milk, and the similar distribution of the differences that were detected to those that occur due to experimental variability, suggest an absence of compart-

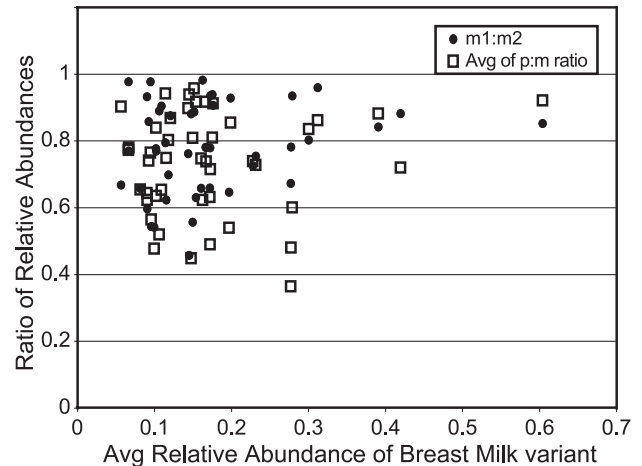


Fig. 4. Sampling reproducibility is reduced for less abundant variants. As a second measure of the difference in the relative abundance of variants, the ratio of the repeated breast milk measurements (m1/m2) for each variant and the average ratio of the blood plasma and breast milk abundances for each variant (p/m) were compared to the average relative abundance of that variant. In this comparison, those variants present at high relative abundances ( $>0.3$ ) are measured consistently (the ratio of measurements approaches 1.0), while those variants present at lower relative abundances are measured with less consistency (ranging from near 1.0 to approximately 0.5). This pattern suggests that the largest degree of variability in measuring the relative abundance of a variant is for those that exist at lower relative abundances. The distribution pattern of the differences between the plasma and breast milk variants (p/m) is similar to that for repeated breast milk samples. This suggests that differences in the abundance of variants between these two compartments are largely related to those of the repeated measurements and likely stem from experimental variability.

mentalization as defined by the absence of a compartment-specific outgrowth of either unique or shared variants.

#### *The effects of sampling error on the inference of population composition*

To assess the effects of sampling error on evaluations of virus population composition, we examined the effects of sample size in three settings: (i) the ability to accurately assign a relative abundance to a variant within a population; (ii) the apparent difference measured between two identical populations; and (iii) the impact of undersampling in the context of an HTA experiment.

To simulate the process of sampling a virus population, we created a large starting population composed of four “genotypic variants” having the frequencies of 50%, 25%, 15%, and 10%. Samples of  $N$  elements (genomes) were chosen at random from this population and the frequency of each variant calculated, simulating the process of template utilization during RT-PCR amplification. This process was repeated  $10^4$  times for values of  $N$  ranging from  $10^1$  to  $10^3$ , and 95% confidence intervals were determined for each distribution (Fig. 5A). This simulation illustrates the effects of small sample sizes on estimates of population structure: the three variants displayed in Fig. 5A cannot be shown to exist at different

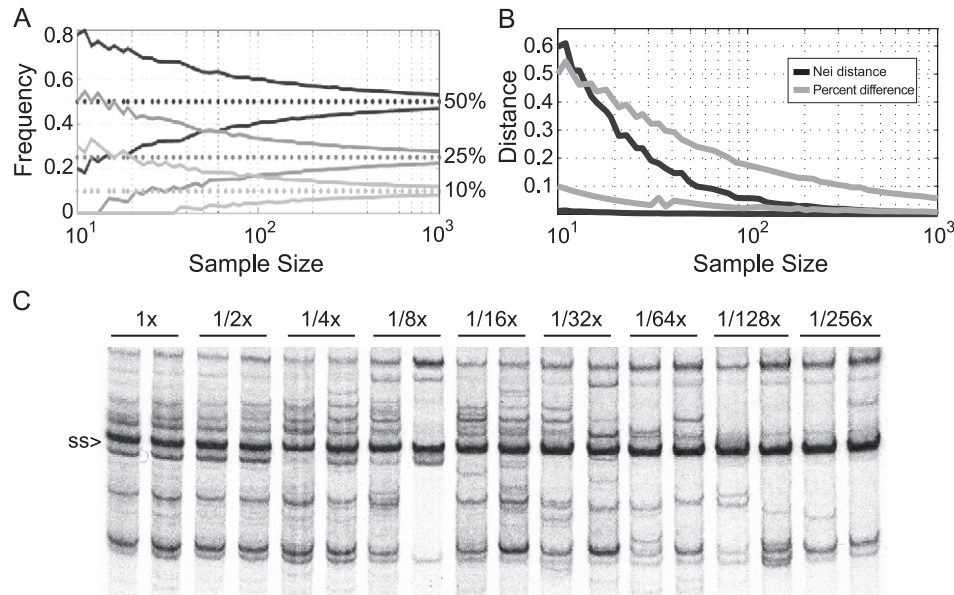


Fig. 5. Effects of sample size on evaluations of population composition. (A) A population of  $10^7$  elements composed of four “variants” at frequencies of 50%, 25%, 15%, and 10% was constructed from which samples of size  $N$  ranging from  $10^1$  to  $10^3$  (indicated by circles along the  $x$ -axis) were randomly chosen. By iterating this process  $10^4$  times for each value of  $N$ , a distribution of values for the frequency of each variant was generated. Solid lines encompass 95% of the values measured for each variant by this process (the 15% species is not shown). (B) To examine the effects of sampling error on the difference measured between two populations,  $10^3$  pairs of frequency measurements were randomly chosen with replacement from the pool of  $10^4$  samples generated above for each value of  $N$ . The difference between each pair of samples was calculated using both the percent difference and Nei distance calculations. Lines indicate the upper and lower limits bounding 95% of the values calculated for each of the distance measures over a range of sample sizes. (C) RT-PCR was performed on 2-fold serial dilutions of template viral RNA from the plasma of subject 30 (VL = 492 732 copies/ml; 2464 RNA molecules were added to the 1× reaction) and the products were analyzed by V1/V2-HTA. Duplicate sets of RNA dilutions were examined to compare the reproducibility of the HTA patterns at each dilution. Similar dilutions are in adjacent lanes with the dilution factor noted above.

frequencies in at least 95% of trials until the sample size is larger than 100, while the variant present at a frequency of 10% cannot be reliably detected until  $N$  is between 30 and 40. Presumably, the same sampling limitations would be at work in sampling a virus population using RT-PCR, with the added caveat that the sample size would be limited by the number of viral genomes that actually contributed to the PCR reaction, not simply the number that was added.

This simulation was extended to examine the effects of sample size on measuring the apparent difference between two identical samples. For each value of  $N$ ,  $10^3$  pairs of frequency distributions were chosen with replacement from the pool of  $10^4$  distributions described above. Because both frequency distributions of each pair are drawn from the same starting population, the expected value of the difference between any pair is zero. However, the actual value of this difference is determined by the extent of sampling error. The “distance” between two samples can be quantified based on the frequency (i.e., relative abundance) of the distinct variants in each sample. This distance was calculated using both the percent difference and Nei’s genetic distance (Nei, 1972) calculations. Fig. 5B demonstrates that the distance between two identical populations as measured by percent difference will remain nonzero even for sample sizes of 1000; Nei distance approaches zero for sample sizes over 200–300. It is notable that even for

sample sizes as large as 100 the apparent percent difference between two samples drawn from the same population averages 10%, and sample sizes of 10 and 20 have average percent differences averaging in the 25–30% range for identical populations.

The findings of this simulation were tested by performing RT-PCR on 2-fold serial dilutions of viral RNA. RNA isolated from the plasma of subject 30 was diluted in PBS, subjected to RT-PCR, and analyzed by V1/V2-HTA. Duplicate sets of dilutions and RT-PCR reactions were performed separately to compare the reproducibility of the HTA patterns at each dilution. Similar heteroduplex patterns were present in duplicate reactions at high RNA concentrations (1× through 1/4×), indicating that good sampling had occurred. However, this reproducibility began to deteriorate from the 1/8× dilution onward. This demonstrates that inadequate sampling can give the appearance of differences in virus populations where none exists. Furthermore, at more extreme dilutions, only one or a few variants are detected, demonstrating that inadequate sampling can give the appearance of homogeneity where a heterogeneous population exists. These results demonstrate that a rigorous assessment of the quality of sampling is critical to any analysis that suggests either a difference between two virus populations or detection of a homogeneous population.

## Discussion

We present here a comparison of the HIV-1 populations in paired blood and breast milk samples by V1/V2-HTA, which demonstrates a lack of compartmentalization. V1/V2-HTA patterns were similar in the eight pairs of samples for which adequate sampling could be demonstrated, with no statistical differences in the relative abundance of variants in these compartments using the percent difference calculation. Evaluations of measurement consistency indicated that the majority of the measured differences occurred for variants that existed at low relative abundances and that the distribution of these differences was the same for repeated breast milk measurements when compared to differences between blood and breast milk samples. Furthermore, differences in measurement given by these methods, as well as computer simulation and template dilution RT-PCR experiments, illustrate how essential it is that sampling error be considered in comparing virus populations. These examples further illustrate the impact of template sampling in the context of PCR as previously discussed by Liu et al. (1996).

The simplest means of minimizing the effects of sampling error is to increase the sample size examined. Because HTA is able to sample all of the viral genomes that contributed to the RT-PCR product, minimizing sampling error is one of its strengths, compared, for example, to cloning and sequencing PCR products. By including only those samples that can be validated for adequate template sampling, our analysis is limited to samples with higher viral loads (Fig. 1), which may bias our results if compartmentalization exists only in samples with low viral loads. We have examined the V1/V2 region of *env* because the high degree of variability in this region typically identifies a greater number of genotypic variants within an individual, allowing a more sensitive assessment of potential population differences between compartments. As has been seen previously (Kitrinis et al., 2003), the use of a molecular clone probe (Du151) increased the sensitivity of the HTA to small genotypic differences in this highly variable region, demonstrated by the large number of variants detected in these samples. It is worth noting that the bending of DNA as recorded by the HTA is not dependent on an absolute percent difference but rather on the presence of clustered nucleotide differences or length differences. Thus querying the variable regions of *env* enhances the sensitivity of the HTA as mutations are frequently concentrated over a short stretch of sequence. When taken with previous similar evaluations (Kitrinis et al., 2003), this validates the use of the HTA as a tool to assess the relative abundance of genotypic variants within complex populations.

In comparing this study to a recent contradictory report by Becquart et al. (2002), the sampling issues we have raised must be considered. In their study, they concluded that compartmentalization of HIV-1 does exist within breast milk based on approximately 15 C2–V3 sequences cloned

from each compartment of three individuals. Referring again to our simulation, Fig. 5B shows that at small sample sizes it is unlikely that a difference of zero will be measured between two identical populations. The large degree of heterogeneity observed among the HIV-1 *env* sequences in both this and the Becquart studies makes sample size even more critical for accurate comparisons of sample populations. We were able to show that not only are major variants present in both compartments at equivalent frequencies, but also many minor variants. Therefore, we propose that the discrepancy between this study and that of Becquart et al. is the result of differences in sampling. Because we have validated the template sampling for the samples presented here, we are confident in concluding that little difference exists between the HIV-1 populations of blood plasma and breast milk in our specimens.

HIV-1 in breast milk may originate from either the systemic compartment or from local HIV-1-infected macrophages, lymphocytes, and ductal epithelial cells. Virus production by infected mammary ductal epithelial cells or lymphoid cells could contribute to compartmentalization via a founder effect. However, breast milk lymphocytes are thought to be an unlikely source of virus due to factors in breast milk believed to inhibit virus replication (McNeely et al., 1995; Van de Perre et al., 1993) and the observation that most infected cells in milk are unable to produce infectious virus (Boulerice et al., 1990). Furthermore, while the presence of mastitis can increase the virus load of breast milk (Semba et al., 1999) and thereby influence the composition of HIV-1 populations in this compartment, the degree of similarity measured between virus populations isolated from the two subjects identified as having breast abnormalities (subjects 57 and 64) is comparable to those from the remaining six subjects, suggesting that the presence of mastitis is not responsible for our observations. Similarly, high titers of anti-HIV-1 antibodies in breast milk may represent a differential selective pressure in this compartment. However, the majority of these antibodies are of the IgG isotype (Van de Perre et al., 1993), likely of systemic origin, and therefore are unlikely to present a unique immune selection in the breast milk compartment. Finally, a direct correlation between plasma virus load and the risk of HIV-1 infection via breast feeding (John et al., 2001) strongly supports a relationship between the viruses in these two compartments.

In summary, this assessment of HIV-1 variants by V1/V2-HTA demonstrates that both the number and relative abundance of HIV-1 V1/V2 variants in blood plasma and breast milk are similar. We have also demonstrated that the quality of viral template sampling influences the accuracy of assessing complex virus populations and that sampling must therefore be validated to support conclusions of difference or homogeneity. Using the ability of the HTA to validate template sampling, we can conclude that blood plasma and breast milk are well equilibrated with little evidence of compartmentalization.

## Materials and methods

### Subject samples

Human blood and breast milk samples were obtained with informed consent from women attending a postnatal clinic at Lilongwe Central Hospital, Lilongwe, Malawi as part of a larger study described in greater detail by Hoffman et al. (2003). Breast milk samples were manually expressed into sterile containers and separated into lipid, skim milk, and cellular fractions by centrifugation at  $1000 \times g$  for 10 min. The virus loads of the blood plasma and the skim milk fraction of the breast milk samples were determined using the Roche Amplicor HIV-1 Monitor and the NucliSens HIV-1 QT assays, respectively. The average CD4<sup>+</sup>T-cell count for the eight subjects analyzed was 340 cells/ $\mu$ l (range 227–468), the average time since delivery was 12.6 months (range 3–26), and all women were antiretroviral therapy naive. Two of the eight subjects (57 and 64) were identified as having breast abnormalities (defined as either rash, cracked nipples, or mastitis).

### RNA extraction and RT-PCR

To remove previously reported inhibitors of the reverse-transcriptase (RT) reaction (Ghosh et al., 2003; Shepard et al., 2000) and to concentrate the virus in the breast milk samples, the skim milk fractions, whose volumes varied from 0.25 to 1.5 ml, were centrifuged at  $25000 \times g$  and 4 °C for 1.5 h. The virus pellet was resuspended in 140  $\mu$ l of phosphate-buffered saline (PBS) prior to viral RNA isolation. Blood plasma samples were not concentrated prior to viral RNA isolation. Viral RNA was extracted from 140  $\mu$ l of plasma or 140  $\mu$ l of PBS using a QIAmp viral RNA kit (Qiagen) according to the manufacturer's instructions. Viral RNA was reverse transcribed using the Titan One-Tube RT-PCR system (Roche), modified as previously described (Kitrinis et al., 2003). To further increase the input number of viral RNA templates from the breast milk samples, 10  $\mu$ l of eluted viral RNA and 24 U of avian myeloblastosis virus RT (Roche Molecular Biochemicals) were used for the cDNA reactions, as compared to 5  $\mu$ l and 12 U, respectively, for plasma samples.

### V1/V2-HTA probe

An HIV-1 clade C V1/V2 *env* probe was constructed using previously described V1 and V2 primers (Kitrinis et al., 2003) to amplify this region from an isolate from Durban, South Africa (Du151; Williamson et al., 2003). The PCR product was cloned into the pT7Blue vector using the Perfectly Blunt cloning kit (Novagen). The Du151 V1/V2-HTA probe was labeled by digesting 6  $\mu$ g of plasmid DNA with *Nde*I at 37 °C for 1 h and subsequently filling in the *Nde*I overhangs by the addition of 50  $\mu$ Ci of [<sup>35</sup>S]dATP (1250 Ci/mmol; NEN Life Science Products), 50  $\mu$ M dTTP,

and 10 U of Klenow fragment of DNA polymerase I, as described (Kitrinis et al., 2003). After removing unincorporated nucleotides using a QIAquick PCR purification kit (Qiagen), the probe was released from the plasmid by digestion with *Kpn*I at 37 °C for 1 h.

### V1/V2-HTA

Heteroduplex annealing reactions, consisting of 8  $\mu$ l of RT-PCR product, 1  $\mu$ l of 10 $\times$  annealing buffer (10 $\times$  = 1 M NaCl, 100 mM Tris-HCl [pH 7.5], 20 mM EDTA), 0.1  $\mu$ M V1 and V2 primer, and 1  $\mu$ l of labeled Du151 probe in a total volume of 10  $\mu$ l, were denatured at 95 °C for 2 min followed by annealing at room temperature for 5 min. The heteroduplexes were then separated by electrophoresis in a 6% polyacrylamide gel (acrylamide-bisacrylamide, 37.5:1) in 1 $\times$  Tris-borate-EDTA buffer. The gels were dried and exposed to X-ray film or a phosphorimaging screen (Molecular Dynamics).

### Quantitative and statistical analysis of V1/V2-HTA

The signal intensity of each HTA band was measured using ImageQuant software (Molecular Dynamics) to analyze captured phosphorscreen images. From these values the relative abundance of each heteroduplex variant was calculated as a percent of the total signal of that HTA lane. The difference between any two samples was assessed using the percent change calculation (Kitrinis et al., 2003; Nelson et al., 2000), with bands comprising less than 5% relative abundance being discarded. Statistical comparisons were made between groups of percent difference values using Wilcoxon's rank sum test.

### Template sampling simulation

Simulations of virus population sampling by RT-PCR were performed using MATLAB (MathWorks, Inc.) as follows. A starting population of  $10^7$  elements composed of four "variants" having the frequencies of 50%, 25%, 15%, and 10% was created to represent the viral genotypic variants circulating in an infected subject. The process of sampling from a population of circulating virus was simulated by randomly choosing a sample of  $N$  elements with replacement from the above starting population and calculating the frequency of each variant in the sample. We repeated this process  $10^4$  times for each of 50 values of  $N$ , ranging logarithmically from  $10^1$  to  $10^3$ . A distribution of  $10^4$  values for the frequency of each variant at each sample size was generated and 95% confidence intervals for each distribution were determined.

The effects of sampling error on measuring the difference between two populations were estimated as follows: for each value of  $N$ ,  $10^3$  pairs were chosen with replacement from the pool of  $10^4$  samples described above. The "distance" between each of the pairs of samples was

calculated based on the relative abundance of each of the four distinct variants using both the percent difference and Nei distance calculations. The percent distance calculation has been used previously to characterize the difference between populations in the context of HTA analysis (Kitrinos et al., 2003; Nelson et al., 2000) and is calculated as:

$$100 \times \frac{1}{2} \sum_{i=1}^n |A_i - B_i|$$

Nei's standard genetic distance is used to define the distance between populations given the allelic frequencies at multiple loci (Nei, 1972). In this analysis, we used a simplified form for a single "locus", where each discrete band of an HTA is analogous to an allele:

$$D_s = \ln(J_{AB}/\sqrt{J_A J_B})$$

where

$$J_A = \sum_{i=1}^n A_i^2, J_B = \sum_{i=1}^n B_i^2, \text{ and } J_{AB} = \sum_{i=1}^n A_i B_i$$

In both of these formulae,  $n$  is the number of variants (i.e., HTA bands), and  $A_i$  and  $B_i$  are the relative abundances of variant  $i$  in populations  $A$  and  $B$  (i.e., HTA lanes), where  $\sum A_i = \sum B_i = 1$ . Using these methods, we generated a distribution of  $10^3$  values for each distance measure at each value of  $N$  and 95% confidence intervals were calculated.

## Acknowledgments

This work was supported by NIH grants R01-AI44667 (R.S.) and R01-DK49381 (M.S.C.). G.J.H. was supported by NIH Training Grant T32-AI07419. K.M.K. was supported by NIH Training Grant T-32-GM07092. We acknowledge the support of the UNC Center for AIDS Research (P30-AI50410), and the UNC Project in Lilongwe, Malawi.

## References

Baba, T.W., Trichel, A.M., An, L., Liska, V., Martin, L.N., Murphey-Corb, M., Ruprecht, R.M., 1996. Infection and AIDS in adult macaques after nontraumatic oral exposure to cell-free SIV. *Science* 272, 1486–1489.

Baron, S., Poast, J., Richardson, C.J., Nguyen, D., Cloyd, M., 2000. Oral transmission of human immunodeficiency virus by infected seminal fluid and milk: a novel mechanism. *J. Infect. Dis.* 181, 498–504.

Becquart, P., Chomont, N., Roques, P., Ayoub, A., Kazatchkine, M.D., Belec, L., Hocini, H., 2002. Compartmentalization of HIV-1 between breast milk and blood of HIV-infected mothers. *Virology* 300, 109–117.

Boulerice, F., Bour, S., Geleziunas, R., Lvovich, A., Wainberg, M.A., 1990. High frequency of isolation of defective human immunodeficiency virus type 1 and heterogeneity of viral gene expression in clones of infected U-937 cells. *J. Virol.* 64, 1745–1755.

Delwart, E.L., Shpaer, E.G., Louwagie, J., McCutchan, F.E., Grez, M., Rubsamén-Waigmann, H., Mullins, J.I., 1993. Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 env genes. *Science* 262, 1257–1261.

Delwart, E.L., Sheppard, H.W., Walker, B.D., Goudsmit, J., Mullins, J.I., 1994. Human immunodeficiency virus type 1 evolution in vivo tracked by DNA heteroduplex mobility assays. *J. Virol.* 68, 6672–6683.

Dunn, D.T., Newell, M.L., Ades, A.E., Peckham, C.S., 1992. Risk of human immunodeficiency virus type 1 transmission through breast-feeding. *Lancet* 340, 585–588.

Dyer, J.R., Pilcher, C.D., Shepard, R., Schock, J., Eron, J.J., Fiscus, S.A., 1999. Comparison of NucliSens and Roche Monitor assays for plasma HIV-1 RNA quantitation. *J. Clin. Microbiol.* 37, 447–449.

Ghosh, M.K., Kuhn, L., West, J., Semrau, K., Decker, D., Thea, D.M., Aldrovandi, G.M., 2003. Quantitation of human immunodeficiency virus type 1 in breast milk. *J. Clin. Microbiol.* 41, 2465–2470.

Hoffman, I.F., Martinson, F.E., Stewart, P.W., Chilongozi, D.A., Leu, S.Y., Kazembe, P.N., Banda, T., Dzinyemba, W., Joshi, P., Cohen, M.S., Fiscus, S.A., 2003. Human immunodeficiency virus type 1 RNA in breast-milk components. *J. Infect. Dis.* 188, 1209–1212.

John, G.C., Nduati, R.W., Mbori-Ngacha, D.A., Richardson, B.A., Panteleeff, D., Mwatha, A., Overbaugh, J., Bwayo, J., Ndinya-Achola, J.O., Kreiss, J.K., 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.

Kitrinos, K.M., Hoffman, N.G., Nelson, J.A., Swanson, R., 2003. Turnover of env variable region 1 and 2 genotypes in subjects with late-stage human immunodeficiency virus type 1 infection. *J. Virol.* 77, 6811–6822.

Lewis, P., Nduati, R., Kreiss, J.K., John, G.C., Richardson, B.A., Mbori-Ngacha, D., Ndinya-Achola, J., Overbaugh, J., 1998. Cell-free human immunodeficiency virus type 1 in breast milk. *J. Infect. Dis.* 177, 34–39.

Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., MacDonald, H., Stuhlmann, M.E., Koup, R.A., Landau, N.R., 1996. Homozygous defect in HIV-1 Coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 86, 367–377.

McGrath, K.M., Hoffman, N.G., Resch, W., Nelson, J.A., Swanson, R., 2001. Using HIV-1 sequence variability to explore virus biology. *Virus Res.* 76, 137–160.

McNeely, T.B., Dealy, M., Dripps, D.J., Orenstein, J.M., Eisenberg, S.P., Wahl, S.M., 1995. Secretory leukocyte protease inhibitor: a human saliva protein exhibiting anti-human immunodeficiency virus 1 activity in vitro. *J. Clin. Invest.* 96, 456–464.

Nduati, R., John, G., Mbori-Ngacha, D., Richardson, B., Overbaugh, J., Mwatha, A., Ndinya-Achola, J., Bwayo, J., Onyango, F.E., Hughes, J., Kreiss, J., 2000. Effect of breastfeeding and formula feeding on transmission of HIV-1: A randomized clinical trial. *JAMA* 283, 1167–1174.

Nei, M., 1972. Genetic distance between populations. *Am. Nat.* 106, 283–291.

Nelson, J.A., Fiscus, S.A., Swanson, R., 1997. Evolutionary variants of the human immunodeficiency virus type 1 V3 region characterized by using a heteroduplex tracking assay. *J. Virol.* 71, 8750–8758.

Nelson, J.A., Baribaud, F., Edwards, T., Swanson, R., 2000. Patterns of changes in human immunodeficiency virus type 1 V3 sequence populations late in infection. *J. Virol.* 74, 8494–8501.

Oxtoby, M.J., 1988. Human immunodeficiency virus and other viruses in human milk: placing the issues in broader perspective. *Pediatr. Infect. Dis. J.* 7, 825–835.

Semba, R.D., Kumwenda, N., Hoover, D.R., Taha, T.E., Quinn, T.C., Mtivalye, L., Biggar, R.J., Broadhead, R., Miotti, P.G., Sokoll, L.J., van der Hoeven, L., Chipangwi, J.D., 1999. Human immunodeficiency virus load in breast milk, mastitis, and mother-to-child trans-



- mission of human immunodeficiency virus type 1. *J. Infect. Dis.* 180, 93–98.
- Shepard, R.N., Schock, J., Robertson, K., Shugars, D.C., Dyer, J., Vernazza, P., Hall, C., Cohen, M.S., Fiscus, S.A., 2000. Quantitation of human immunodeficiency virus type 1 RNA in different biological compartments. *J. Clin. Microbiol.* 38, 1414–1418.
- Starcich, B.R., Hahn, B.H., Shaw, G.M., McNeely, P.D., Modrow, S., Wolf, H., Parks, E.S., Parks, W.P., Josephs, S.F., Gallo, R.C., et al., 1986. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* 45, 637–648.
- Thorne, C., Newell, M.L., 2000. Epidemiology of HIV infection in the newborn. *Early Hum. Dev.* 58, 1–16.
- Van de Perre, P., Simonon, A., Hitimana, D.G., Dabis, F., Msellati, P., Mukamabano, B., Butera, J.B., Van Goethem, C., Karita, E., Lepage, P., 1993. Infective and anti-infective properties of breastmilk from HIV-1-infected women. *Lancet* 341, 914–918.
- Williamson, C., Morris, L., Maughan, M.F., Ping, L.H., Dryga, S.A., Thomas, R., Reap, E.A., Cilliers, T., van Harmelen, J., Pascual, A., Ramjee, G., Gray, G., Johnston, R., Karim, S.A., Swanstrom, R., 2003. Characterization and selection of HIV-1 subtype C isolates for use in vaccine development. *AIDS Res. Hum. Retroviruses* 19, 133–144.