Analysis of Genetic Linkage of HIV From Couples Enrolled in the HIV Prevention Trials Network 052 Trial

Susan H. Eshleman,1 Sarah E. Hudelson,1 Andrew D. Redd,2 Lei Wang,2 Rachel Debes,4 Ying Q. Chen,3 Craig A. Martens,5 Stacy M. Ricklefs,6 Ethan J. Selig,6 Stephen F. Porcella,5 Supriya Munshaw,6 Stuart C. Ray,6 Estelle Piwowar-Manning,1 Marybeth McCauley,7 Mina C. Hosseinipour,8 Johnstone Kumwenda,9 James G. Hakim,10 Suwat Chariyalertsak,11 Guy de Bruyn,12 Beatriz Grinsztejn,13 Nagalingeswaran Kumarasamy,14 Joseph Makhema,15 Kenneth H. Mayer,16 Jose Pilotto,17 Breno R. Santos,18 Thomas C. Quinn,2,6 Myron S. Cohen,19 and James P. Hughes3,4

1Department of Pathology, Johns Hopkins University, School of Medicine, Baltimore, Maryland; 2Laboratory of Immunoregulation, Division of Intramural Research, National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Baltimore, Maryland; 3Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington; 4Department of Biostatistics, University of Washington, Seattle; 5Genomics Unit, Research Technologies Section, Rocky Mountain Laboratories, Division of Intramural Research, NIAID, NIH, Hamilton, Montana; 6Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland; 7FHI 360, Health and Development Sciences Division, Arlington, Virginia; 8UNC Project, Kamuzu Central Hospital, Lilongwe, Malawi; 9College of Medicine, Blantyre, Malawi; 10Department of Medicine, University of Zimbabwe, Harare; 11Research Institute for Health Sciences, Chiang Mai University, Thailand; 12Perinatal HIV Research Unit, University of the Witwatersrand, Johannesburg, South Africa; 13Laboratorio de Pesquisa Clinica em DST/AIDS, Instituto de Pesquisa Clinica Evandro Chagas-Fiocruz, Rio de Janeiro, Brazil; 14YRG CARE Medical Ctr., Chennai, India; 15Botsswana–Harvard School of Public Health AIDS Initiative Partnership, Gaborone; 16Fenway Health, Boston, Massachusetts; 17Hospital Geral de Nova Iguacu and Laboratorio de AIDS e Imunologia Molecular-IoC/Fiocruz, Rio de Janeiro, Brazil; 18Hospital Nossa Senhora da Conceicao, Porto Alegre, Brazil; and 19Department of Medicine, University of North Carolina at Chapel Hill

Background. The HIV Prevention Trials Network (HPTN) 052 trial demonstrated that early initiation of antiretroviral therapy (ART) reduces human immunodeficiency virus (HIV) transmission from HIV-infected adults (index participants) to their HIV-uninfected sexual partners. We analyzed HIV from 38 index-partner pairs and 80 unrelated index participants (controls) to assess the linkage of seroconversion events.

Methods. Linkage was assessed using phylogenetic analysis of HIV pol sequences and Bayesian analysis of genetic distances between pol sequences from index-partner pairs and controls. Selected samples were also analyzed using next-generation sequencing (env region).

Results. In 29 of the 38 (76.3%) cases analyzed, the index was the likely source of the partner’s HIV infection (linked). In 7 cases (18.4%), the partner was most likely infected from a source other than the index participant (unlinked). In 2 cases (5.3%), linkage status could not be definitively established.

Conclusions. Nearly one-fifth of the seroconversion events in HPTN 052 were unlinked. The association of early ART and reduced HIV transmission was stronger when the analysis included only linked events. This underscores the importance of assessing the genetic linkage of HIV seroconversion events in HIV prevention studies involving serodiscordant couples.

Human immunodeficiency virus (HIV) from 2 individuals who are linked through transmission shares more genetic homology than does HIV from unrelated individuals. Therefore, genetic analysis of HIV from different individuals can be used to assess the linkage of HIV infections [1]. This approach has been used to identify clusters of HIV infections in clinical trials [2, 3] and was recently used to evaluate linkage of seroconversion events in serodiscordant couples enrolled in an HIV prevention trial [4, 5]. In that trial, 26.5% of the...
HIV seroconversion events were unlinked (ie, the partner likely acquired HIV infection from a source other than the HIV-infected participant) [4]. The frequency with which unlinked events occur in serodiscordant couples is likely to vary between populations and is likely to be influenced by social norms, behavioral factors, host and viral factors that influence infectivity and susceptibility to infection, interventions used for HIV prevention, and other factors.

The HIV Prevention Trials Network (HPTN) 052 is a multicenter, phase III, randomized clinical trial designed to test whether early initiation of antiretroviral therapy (ART) reduces transmission from HIV-infected adults to their HIV-uninfected sexual partners. On 28 April 2011, an independent data safety and monitoring board (DSMB) determined that trial results were adequate to demonstrate prevention, and other factors. This report includes results from work performed both before and after the April 2011 DSMB meeting, as noted.

### METHODS

#### Samples Used for Analysis

HPTN 052 enrolled serodiscordant couples (index-partner pairs; 97% heterosexual). At screening, HIV-infected index participants were ART naive, had a CD4 cell count of 350–550 cells/mm³, and did not require ART for their own health according to local, country-specific guidelines. Partners were HIV uninfected at screening. Enrolled couples were randomized to initiate ART in the index at the time of enrollment (immediate arm) or to initiate ART in the index when 2 consecutive CD4 cell counts were ≥250 cells/mm³ or an AIDS-defining illness developed (delayed arm). From April 2005 through May 2010, 1763 couples with an HIV-infected partner were enrolled at 13 study sites in Africa, Asia, and North and South America. Couples agreed to be followed up for at least 5 years, with HIV testing of the uninfected partner performed at regular intervals [6].

Samples were available for analysis from 38 of 39 seroconversion events that occurred prior to the April 2011 DSMB meeting (Table 1). Samples from randomly selected index participants were analyzed for comparison (control samples). Ten control samples were analyzed for each study site, with 2 exceptions: control samples were not available from the site in the United States because of limited enrollment, and a total of 10 control samples were analyzed from 3 sites in Brazil.

### Table 1. Analysis of Genetic Linkage of HIV From Index-Partner Pairs

<table>
<thead>
<tr>
<th>Region</th>
<th>No. SC Events</th>
<th>Subtype</th>
<th>No. control Samples</th>
<th>Phylogenetic analysis (pol region)b</th>
<th>Bayesian analysis (pol region)c</th>
<th>NGS (env region)</th>
<th>Final linkage assessmentd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>31 C</td>
<td>50</td>
<td>25</td>
<td>22</td>
<td>22</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>Asia</td>
<td>3 C/AE</td>
<td>20</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>America</td>
<td>4 B/F</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Abbreviations: AE, CRF01_AE; HIV, human immunodeficiency virus; ND, not determined; NGS, next-generation sequencing SC, seroconversion events; TBD, to be determined.

a Sequences from subtype-matched local control samples were used for analysis; those samples were obtained from randomly selected index participants at the same study sites. There were no subtype F controls and no controls for the site in the United States.

b Phylogenetic analysis was performed using HIV pol sequences obtained by population sequencing. Sequences from the same individual grouped together in all cases with bootstraps of 100%. Events were considered to be linked if all of the samples from an index-partner pair grouped together on a single, unique branch; in all but 1 case, the cluster of linked sequences had a bootstrap of 100%; in 1 case, the bootstrap was 99%.

c Bayesian analysis was performed by comparing genetic similarity values obtained by comparing pol sequences from seroconversion events (index and partner sequences) and sequences from local control samples (samples from randomly selected index participants from the same study site).

d Results are the same as those available at the time of the April 2011 data safety monitoring board (DSMB) meeting, with 1 exception: linkage of 1 seroconversion event that was provisionally characterized as unlinked by phylogenetic and Bayesian analysis of pol region sequences was subsequently determined to be linked by NGS of the env region (Figure 4).

a “Americas” is defined as the United States and Brazil.

f Thirty-nine seroconversion events occurred prior to the date for data cutoff for the DSMB meeting. Samples from 1 event that occurred close to the time of the April 2011 DSMB meeting were not available for analysis.
Phylogenetic Analysis of HIV pol Sequences

Samples were analyzed using the ViroSeq HIV Genotyping System (Celera). This system provides a consensus sequence of the viral population (pol region) that encodes all 99 amino acids in HIV protease and the first 335 amino acids in HIV reverse transcriptase [7]. The system includes a contamination control system [8] and does not use nested polymerase chain reaction (PCR), which has been associated with sample cross-contamination [9]. Seven samples were amplified using alternate amplification primers.

Pol region sequences (1302 nucleotides) from index-partner pairs and local control samples were aligned with reference sequences recommended for HIV-1 subtype analysis (http://hiv-web.lanl.gov/) with use of MegAlign software, version 5.07 (ClustalW alignment method). Distances between sequences were calculated with DNADIST, and phylogenetic trees with bootstrap support were inferred using neighbor-joining and consense (PHYLIP, version 3.69; http://evolution.genetics.washington.edu/phylip.html). The Kimura 2-parameter model was used with a transition/transversion ratio of 1.5, implemented in DNADIST. Bootstrap values >80% were considered acceptable for subtype assignment. Results obtained with the Kimura-2-parameter model were compared with results obtained using the ML (F84) model [10] implemented in GARLI version 2 [11].

Trees were generated for each geographic region with use of MegAlign, version 5.07; region-specific trees included pol sequences from the relevant index-partner pairs, subtype-matched local control sequences, and subtype-matched reference sequences. A seroconversion event was classified as linked if the index and partner sequences grouped together on a monophyletic branch with a high bootstrap value.

Bayesian Analysis of HIV Linkage

We developed an algorithm based on Bayes’ theorem [4] to estimate the probability of linkage of seroconversion events based on the analysis of sequences from those events and sequences from epidemiologically related and unrelated individuals. Genetic similarity (percentage identity) between paired pol sequences (index, partner, and control sequences) was calculated using MegAlign, version 5.07. Genetic similarity was computed between all possible pairs of sequences, including (i) pairs of sequences from samples collected at different study visits from a single individual, (ii) pairs of sequences from unrelated index participants (control sequences from the same study site or region), and (iii) pairs of sequences from index-partner pairs. We assumed that pairs of sequences from unrelated index participants were unlinked. Therefore, we used the genetic similarities of the control sequences (type ii) to estimate the distribution of percentage identity between known unlinked sequence pairs. Furthermore, we assumed that the distribution of similarities between 2 sequences from the same individual (type i) was similar to the distribution of similarities in linked index-partner pairs. We refer to the type (i) and type (ii) sequence pairs as the training data. We then used Bayes’ theorem to compute the posterior probability of linkage for the unknown sequence pairs (type iii):

$$Pr(Y_i = 1 | s_i) = \frac{[p_0 + f_1(s_i)]}{[p_0 + f_1(s_i) + (1-p_0) + f_0(s_i)]},$$

where $Y_i$ and $s_i$ represent the linkage status (1 = linked; 0 = unlinked) and genetic similarity, respectively, for sequence pair $i$; $f_0(s)$ and $f_1(s)$ give the density for a similarity $s$ among unlinked and linked sequences, respectively (estimated from the training data) and $p_0$ is the prior probability of linkage ($p_0$ was taken as the overall proportion of linked transmissions among the 38 observed transmissions, making the process iterative) (Supplementary File 1).

We used kernel density estimation to estimate $f_0$ and $f_1$ from training data [12]. It is unclear how widely the distributions of similarity vary across sites, HIV subtypes, or other factors. Therefore, we computed 2 probabilities for each sequence pair: unpooled probabilities, which computed $f_0$ and $f_1$ based on training data from the same site as the sequence pair of interest, and pooled probabilities, which computed $f_0$ and $f_1$ based on pooling all subtype C data (for subtype C sequences) or all data (for non–subtype C sequences). Because the pooled estimates of $f_0$ and $f_1$ are more stable (ie, based on larger numbers), we based our conclusions regarding linkage on the pooled results, although the results were qualitatively similar in the 2 analyses. Couples with probability of linkage ≥0.5 for any sequence pair were provisionally categorized as linked; couples with probability of linkage <0.5 for all sequence pairs were provisionally categorized as unlinked.

Analysis of HIV env by Next-Generation Sequencing

Selected samples were analyzed using next-generation sequencing (NGS) [13, 14], as described elsewhere [15]. A combined reverse-transcription polymerase chain reaction (PCR) was used to amplify a region of gp41 (HXB2 coordinates: 7691–8374). A nested PCR reaction was then performed using primer sets that included DNA barcodes for sample identification in NGS. PCR products were analyzed using gel electrophoresis to confirm successful amplification and were purified with the Amplicon Library Preparation Method (Roche). Amplicon pools were prepared by combining 5 μL of each diluted barcoded template to make library pools that contained 14 barcoded amplicons (1 × 10^9 molecules/μL). Templated beads were prepared for NGS using the emPCR Method Manual-Lib-L-MV (Roche). Library pools were diluted to 1 × 10^5 molecules/μL for a target addition of 0.175 copies per bead to the DNA Capture Beads. Enriched DNA Capture Beads were sequenced on the Roche
454 instrument (Roche), according to the manufacturer’s instructions, with use of a 4-region gasket.

Consensus sequences were generated using GS Amplicon Variant Analyzer, version 2.5 (Roche), according to the manufacturer’s recommendations. Consensus sequences that were within 10 bases from both ends of the amplicon and comprised a cluster of ≥10 individual reads were retained in the analysis. Consensus sequences and subtype reference sequences were aligned using ClustalW [16]. Merged phylogenetic trees were generated by the neighbor-joining method, using all available consensus sequences from each index-partner and subtype reference sequences. Statistical support for subtype assignment was obtained by bootstrapping (500 replicates). A seroconversion event was considered linked if all available index and partner samples contained multiple consensus sequences that grouped together with a high bootstrap value.

Analysis of Factors Associated With Linked and Unlinked Seroconversion Events
Fisher exact test was used to assess the association between linkage and categorical variables; Wilcoxon rank-sum test was used to compare the time from enrollment to seroconversion between linked and unlinked transmissions. All P values are 2-sided.

GenBank Accession Numbers
Accession numbers for population sequencing (HIV pol) were JN247047–JN247075 (sequences analyzed in Figure 2) and JN634296–JN634492 (other sequences) and for NGS (HIV env) were JN371773–JN374672.

General Considerations
Personnel who performed the linkage studies were blinded to the study arm assignments of study participants. Laboratory and statistical analyses were performed independently. After the 28 April 2011 DSMB meeting, all of the data were further reviewed by 3 external experts who were also blinded to study arm assignments.

Informed Consent
Human experimental guidelines of the US Department of Health and Human Services and those of the authors’ institutions were followed in the conduct of this research.

Ethical review committees at each local and collaborating organization approved the HPTN 052 trial, and the trial was registered in ClinicalTrials.gov (NCT00074581). Written informed consent was obtained from all study participants.

RESULTS
An overview of the methods used to assess the genetic linkage of seroconversion events in HPTN 052 is shown in Figure 1. HIV pol region sequences were obtained for 2 samples from different study visits for 70 of the 76 participants (92.1%) studied (38 index-partner pairs); 3 index participants and 3 partners had only 1 pol sequence result. The median time between collection of the 2 partner samples was 27 days (range, 1–126 days), the median time between collection of the 2 index samples was 356 days (range, 27–1211 days), and the median time between collection of the earliest index sample and the latest partner sample was 380 days (range, 1–1211 days); in some cases, the second index sample was collected after the seroconversion event. A single phylogenetic tree was generated using all available index and partner sequences, 80 local control sequences, subtype reference sequences, and an outgroup sequence (simian immunodeficiency virus). In all cases, paired sequences from the same individual grouped closely together on monophyletic branches; the median genetic similarity of the paired sequences was 99.5% (range, 94.7%–100%). For all 38 index-partner pairs, the HIV subtype of the index and partner was the same and was consistent with the prevalent subtype(s) in the region (Table 1). Antiretroviral drug resistance mutations were detected in samples from 4 index-partner pairs. In all 4 cases, the pattern of resistance mutations detected was consistent with the final linkage assessment (Supplementary File 2). Phylogenetic trees were also generated for each geographic region (Africa, Asia, and North and South America). The phylogenetic clustering of sequences from index-partner pairs was the same in the large composite tree and the region-specific trees (not shown). A representative tree that includes 5 seroconversion events is shown (Figure 2).
In 30 of the 38 couples analyzed, the sequence(s) obtained for the index grouped with the sequence(s) obtained for the corresponding partner on a unique, monophyletic branch. The bootstrap values for the grouped sequences were 100% in all but 1 case (for 1 event, the bootstrap value was 99%). Those 30 seroconversion events were provisionally characterized as linked. In the remaining 8 couples, sequences from the index and partner did not group together. Those 8 seroconversion events were provisionally characterized as unlinked. There was no difference in the linkage assessments using other models for determining genetic distance (see Materials and Methods).

The genetic linkage of HIV from index-partner pairs was also assessed by comparing genetic similarity of paired HIV pol sequences using Bayesian analysis. A posterior probability of linkage was computed for each index-partner sequence pair (up to 4 sequences per couple) using pooled and unpooled training data (Figure 3). Pooled results were used for the primary determination of linkage. Using this approach, 27 of the 38 (71%) seroconversion events were classified as linked and 11 as unlinked.

In 35 of the 38 (92.1%) seroconversion events analyzed, the linkage assessments based on phylogenetic and Bayesian analyses were concordant (Table 1). In 27 cases, both were linked; in 8 cases, both were unlinked. In the remaining 3 cases, phylogenetic analysis suggested that the seroconversion events were linked, whereas statistical analysis suggested that they were not linked; those events were provisionally characterized as “to be determined” (TBD). Examples of results from phylogenetic and Bayesian analyses are shown in Table 2.

Samples from 12 of the 38 (31.6%) seroconversion events were further analyzed using NGS (env region [gp41]). This included 1 event provisionally characterized as linked, 8 events
provisionally characterized as unlinked, and 3 events provisionally characterized as TBD. NGS analysis confirmed that 7 of the 8 provisionally unlinked events were unlinked. However, for 1 event that was provisionally classified as unlinked, NGS analysis revealed that the event was linked (event 052-2989; Figure 4A). NGS results for 1 event provisionally characterized as TBD indicated that the event was linked. NGS results for the other 2 TBD events did not sufficiently meet the criteria established for linkage; the status for those events was not changed. One sample provisionally characterized as linked was confirmed to be linked by NGS.

Final linkage status was determined for 36 of the 38 events analyzed (Table 1). In all but 1 case, the final linkage assessment was the same as the assessment made based on data available at the time of the April 2011 DSMB meeting. In 1 case, a seroconversion event in the delayed ART study arm that was previously characterized as TBD was subsequently characterized as linked (event 052-1168; Figure 4B).

The median minimum time between collection of index and partner samples was 28 days (interquartile range [IQR], 0–84 days; range, 0–1083 days) and was longer for the 7 unlinked events (median, 266 days; IQR, 180–537 days; range, 77–696 days) than for the 29 linked events (median, 4.0 days; IQR, 0–66 days; range, 0–1083 days). Nonetheless, the range of time between paired index specimens (median, 355 days; IQR, 178–503 days), for which 29 of 30 similarities were 97%, was similar to the range of time between unlinked index-partner samples, all of which had similarities <95%. This suggests that the timing of specimen collection did not significantly influence the final linkage assessment.

We analyzed the association of demographic, behavioral, and clinical factors with the linkage status of seroconversion events (Table 3). Linked HIV transmission was more frequent when the index participant was in the delayed ART study arm and was not receiving ART at the time of the partner's seroconversion. There was also a significant association between unlinked transmission

---

**Table 2. Representative Results Obtained From Linkage Analysis**

<table>
<thead>
<tr>
<th>Seroconversion event identification number</th>
<th>No. of available samples</th>
<th>Phylogenetic clustering</th>
<th>Bootstrap value</th>
<th>Sequence similarity</th>
<th>Bayesian posterior probability, pooled</th>
<th>Bayesian posterior probability, unpooled</th>
<th>Final linkage assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>052-1692</td>
<td>4</td>
<td>Grouped</td>
<td>100%</td>
<td>98.2–98.5</td>
<td>1.00</td>
<td>1.00</td>
<td>Linked</td>
</tr>
<tr>
<td>052-2899</td>
<td>3</td>
<td>Grouped</td>
<td>100%</td>
<td>97.9–98.0</td>
<td>0.99–1.00</td>
<td>0.91–1.00</td>
<td>Linked</td>
</tr>
<tr>
<td>052-0621</td>
<td>4</td>
<td>Grouped</td>
<td>100%</td>
<td>96.9–98.5</td>
<td>0.94–1.00</td>
<td>0.99–1.00</td>
<td>Linked</td>
</tr>
<tr>
<td>052-3172</td>
<td>4</td>
<td>Not grouped</td>
<td>ND</td>
<td>92.2</td>
<td>0.00</td>
<td>0.00</td>
<td>Unlinked</td>
</tr>
<tr>
<td>052-2377</td>
<td>4</td>
<td>Grouped</td>
<td>100%</td>
<td>98.5</td>
<td>1.00</td>
<td>1.00</td>
<td>Linked</td>
</tr>
</tbody>
</table>

---

The median minimum time between collection of index and partner samples was 28 days (interquartile range [IQR], 0–84 days; range, 0–1083 days) and was longer for the 7 unlinked events (median, 266 days; IQR, 180–537 days; range, 77–696 days) than for the 29 linked events (median, 4.0 days; IQR, 0–66 days; range, 0–1083 days). Nonetheless, the range of time between paired index specimens (median, 355 days; IQR, 178–503 days), for which 29 of 30 similarities were >97%, was similar to the range of time between unlinked index-partner samples, all of which had similarities <95%. This suggests that the timing of specimen collection did not significantly influence the final linkage assessment.

We analyzed the association of demographic, behavioral, and clinical factors with the linkage status of seroconversion events (Table 3). Linked HIV transmission was more frequent when the index participant was in the delayed ART study arm and was not receiving ART at the time of the partner’s seroconversion. There was also a significant association between unlinked transmission

---

**Table 2. Representative Results Obtained From Linkage Analysis**

<table>
<thead>
<tr>
<th>Seroconversion event identification number</th>
<th>No. of available samples</th>
<th>Phylogenetic clustering</th>
<th>Bootstrap value</th>
<th>Sequence similarity</th>
<th>Bayesian posterior probability, pooled</th>
<th>Bayesian posterior probability, unpooled</th>
<th>Final linkage assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>052-1692</td>
<td>4</td>
<td>Grouped</td>
<td>100%</td>
<td>98.2–98.5</td>
<td>1.00</td>
<td>1.00</td>
<td>Linked</td>
</tr>
<tr>
<td>052-2899</td>
<td>3</td>
<td>Grouped</td>
<td>100%</td>
<td>97.9–98.0</td>
<td>0.99–1.00</td>
<td>0.91–1.00</td>
<td>Linked</td>
</tr>
<tr>
<td>052-0621</td>
<td>4</td>
<td>Grouped</td>
<td>100%</td>
<td>96.9–98.5</td>
<td>0.94–1.00</td>
<td>0.99–1.00</td>
<td>Linked</td>
</tr>
<tr>
<td>052-3172</td>
<td>4</td>
<td>Not grouped</td>
<td>ND</td>
<td>92.2</td>
<td>0.00</td>
<td>0.00</td>
<td>Unlinked</td>
</tr>
<tr>
<td>052-2377</td>
<td>4</td>
<td>Grouped</td>
<td>100%</td>
<td>98.5</td>
<td>1.00</td>
<td>1.00</td>
<td>Linked</td>
</tr>
</tbody>
</table>

---

The median minimum time between collection of index and partner samples was 28 days (interquartile range [IQR], 0–84 days; range, 0–1083 days) and was longer for the 7 unlinked events (median, 266 days; IQR, 180–537 days; range, 77–696 days) than for the 29 linked events (median, 4.0 days; IQR, 0–66 days; range, 0–1083 days). Nonetheless, the range of time between paired index specimens (median, 355 days; IQR, 178–503 days), for which 29 of 30 similarities were >97%, was similar to the range of time between unlinked index-partner samples, all of which had similarities <95%. This suggests that the timing of specimen collection did not significantly influence the final linkage assessment.

We analyzed the association of demographic, behavioral, and clinical factors with the linkage status of seroconversion events (Table 3). Linked HIV transmission was more frequent when the index participant was in the delayed ART study arm and was not receiving ART at the time of the partner’s seroconversion. There was also a significant association between unlinked transmission
and the partner’s report of a higher number of sexual partners during the 3 months before seroconversion. We did not find an association between linked transmission and geographic region (Africa, Asia, and America), index sex, index CD4 cell count at the time of the partner’s seroconversion, or the length of time between enrollment and seroconversion. Linked transmission was less frequent in male–male couples. However, there were only 2 male–male couples in the analysis.

DISCUSSION

At the time of the April 2011 DSMB meeting, 39 seroconversion events had been documented in the HPTN 052 trial. Linkage status was determined for 36 of 39 events (linkage status of 2 events could not be definitively determined; samples from 1 event were not available for analysis). For 7 of the 36 (19.4%) seroconversion events for which linkage was established, the partner likely acquired HIV infection from a source other than the index participant. This frequency of unlinked events is lower than that observed in the Partners in Prevention trial (26.5%), which enrolled serodiscordant couples in 7 sub-Saharan African countries [4], and is higher than that observed in a cohort of serodiscordant couples from Zambia (13%) [17]; however, none of these differences is statistically significant.

The primary linkage analysis in this study was based on HIV pol sequences that have been used previously for HIV linkage studies [14–17]. Data from this study indicate that there was sufficient genetic diversity in the pol region among participants in HPTN 052 to discriminate between linked and unlinked HIV infections. This is apparent from the posterior probability values, most of which were very close to 0 (definitely unlinked) or 1 (definitely linked), and from the very limited overlap in the distribution of genetic similarities (Figure 3). The pol region generally has a low rate of genetic diversification during HIV infection. This was important because the collection dates of samples used for the analysis of some index-partner pairs differed by >1 year. Data from this study indicate that genetic distance increased in the region analyzed by only 0.65% per year. Very few drug resistance mutations were detected in the sequences analyzed (Supplementary File 2). Analysis of 2 independent samples from each study participant and inclusion of local control samples obtained from randomly selected index
participants enrolled in HPTN 052 provided an additional level of quality control for the linkage assessment.

In the Partners in Prevention study, unlinked events were more frequent in couples in which the seroconverting partner was male [4]. We did not find an association between linkage and sex in HPTN 052. However, our ability to detect statistically significant associations with sex and other factors may have been limited by the small number of events examined (36 classifiable).

Of interest, ART use (in either the immediate or delayed ART study arm) and viral suppression in the index case were associated with a higher frequency of unlinked seroconversion events. All 7 of the unlinked events occurred in couples in which the index participant’s viral load was very low at the time of the partner’s seroconversion (>400 copies/mL in 6 cases, 434 copies/mL in 1 case). This was somewhat surprising; the probability of observing 6 of 7 or 7 of 7 suppressed index participants among the unlinked seroconversion events by chance is 8.4%. As one might expect, there was a significant association between unlinked transmission and a higher number of sexual partners reported by the partner in the 3 months before HIV seroconversion. The analysis of factors associated with linkage of seroconversion events was based on a small number of seroconversion events (those that occurred before the DSMB meeting that led to early release of the study results). This may have limited the power to detect some associations. We cannot determine whether any of the associations that we observed would also be observed with longer follow-up or whether new associations would be identified with longer follow-up of study participants.

The impact of early ART on HIV transmission in HPTN 052 was analyzed for all 39 seroconversion events that occurred before the DSMB meeting and for the subset of 28 linked events identified at the time of the DSMB meeting [6]. When all 39 events were included in the analysis (4 in the immediate ART arm, 35 in the delayed ART arm), the difference in HIV incidence between the 2 study arms was highly statistically significant (rate ratio, 0.114; exact 95% CI, .041–.321; \( P < .001 \)). When the analysis included only the 28 linked events (1 in the immediate arm, 27 in the delayed arm), the association of the study intervention (early ART) and transmission was even

### Table 3. Factors Associated With Linkage of Seroconversion Events

<table>
<thead>
<tr>
<th></th>
<th>Linked</th>
<th>Unlinked</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
<td>.60a</td>
</tr>
<tr>
<td>Africa</td>
<td>24 (83%)</td>
<td>5 (71%)</td>
<td></td>
</tr>
<tr>
<td>Asia/America</td>
<td>5 (17%)</td>
<td>2 (29%)</td>
<td></td>
</tr>
<tr>
<td><strong>Index sex</strong></td>
<td></td>
<td></td>
<td>.39a</td>
</tr>
<tr>
<td>Male</td>
<td>10 (34%)</td>
<td>4 (57%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>19 (66%)</td>
<td>3 (43%)</td>
<td></td>
</tr>
<tr>
<td><strong>Couple type</strong></td>
<td></td>
<td></td>
<td>.033a</td>
</tr>
<tr>
<td>Male–male</td>
<td>0 (0%)</td>
<td>2 (29%)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>29 (100%)</td>
<td>5 (71%)</td>
<td></td>
</tr>
<tr>
<td><strong>Index study arm</strong></td>
<td></td>
<td></td>
<td>.018a</td>
</tr>
<tr>
<td>Early ART</td>
<td>1 (3%)</td>
<td>3 (43%)</td>
<td></td>
</tr>
<tr>
<td>Deferred ART</td>
<td>28 (97%)</td>
<td>4 (57%)</td>
<td></td>
</tr>
<tr>
<td><strong>Index on ART at time of SC</strong></td>
<td></td>
<td></td>
<td>.0076a</td>
</tr>
<tr>
<td>Yes</td>
<td>20 (7%)</td>
<td>4 (57%)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>27 (93%)</td>
<td>3 (43%)</td>
<td></td>
</tr>
<tr>
<td><strong>Index CD4 count at time of SC, median (IQR)</strong></td>
<td>352 (283–468)</td>
<td>404 (301–540)</td>
<td>.41c</td>
</tr>
<tr>
<td><strong>Years between enrollment and seroconversion, median (IQR)</strong></td>
<td>1.0 (0.5–2.0)</td>
<td>2.0 (0.7–2.7)</td>
<td>.25c</td>
</tr>
<tr>
<td><strong>No. of sex partners in the 3 months before SC</strong></td>
<td>1 0 (0%)</td>
<td>4 (57%)</td>
<td>(&lt; .0001^a)</td>
</tr>
<tr>
<td>&gt;1</td>
<td>26 (90%)</td>
<td>3 (43%)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2 (7%)</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** ART, antiretroviral therapy; IQR, interquartile range; SC, seroconversion; VL, viral load.

* \( P \) value from Fisher exact test.

* Two index participants in linked index-partner pairs were on ART at the time of that human immunodeficiency virus infection was documented in the partner (time of seroconversion). One of these participants was in the immediate ART arm of the HIV Prevention Trials Network (HPTN) 052 trial and initiated ART at the time of enrollment into the trial, approximately 3 months before the seroconversion event. The other index participant was in the delayed ART arm of the HPTN 052 trial and initiated ART approximately 4 weeks before the seroconversion event.

* \( P \) value from Wilcoxon rank–sum test.

* One participant was on ART due to pregnancy. She was considered not to be taking ART at the time of her partner’s seroconversion.

* One participant in a linked seroconversion couple did not provide information about the number of sex partners in the 3 months before seroconversion.
stronger (rate ratio, 0.04; exact 95% CI, 0.01–0.27; \( P < 0.001 \)) [6].

Data obtained after the DSMB meeting resulted in only 1 additional linkage assignment: 1 seroconversion event in the delayed ART arm was characterized as linked. These additional data further strengthen the association between early ART initiation and risk reduction of HIV transmission. Data from this study and other recent studies [4, 17] indicate that a significant proportion of seroconversion events in serodiscordant couples may be unlinked. These findings underscore the importance of assessing the genetic linkage of HIV in seroconversion events in HIV prevention studies involving serodiscordant couples.

**Supplementary Data**

Supplementary materials are available at The Journal of Infectious Diseases online (http://www.oxfordjournals.org/our_journals/jid/).

Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyrighted. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

**Notes**

**Acknowledgments.** The authors thank the entire HPTN 052 team for their dedication, commitment, and efforts; the participants in the HPTN 052 study for invaluable contributions; and the laboratory staff at Johns Hopkins University, at the Rocky Mountain Laboratories, and at the HPTN 052 study sites for assistance with sample and data management. The authors thank Suzanna Lamers (BioInfo Experts) for help submitting NGS data to GenBank, and thank Ronald Swanson and Susan Fiscus (Univ. of North Carolina at Chapel Hill) for helpful discussions.

**Disclaimer.** The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the National Institutes of Health (NIH). Use of trade names is for identification purposes only and does not constitute endorsement by the NIH.

**Financial support.** This work was supported by the HIV Prevention Trials Network (HPTN) sponsored by the National Institute of Allergy and Infectious Diseases (NIAID), the National Institute on Drug Abuse (NIDA), the National Institute of Mental Health (NIMH), and the Office of AIDS Research, of NIH, Department of Health and Human Services (U01AI068613/UM1AI068613 to HPTN Network Laboratory; U01AI068617 to HPTN Statistical and Data Management Center; and U01AI068619 to HPTN Core and Operations Center); the Division of Intramural Research, NIAID, NIH; NIAID, NIH (R01AI029168); and NIDA, NIH (R01DA024565).

**Potential conflicts of interest.** S. H. E. has given presentations at meetings sponsored by Abbott Diagnostics (distributor of the ViroSeq HIV Genotyping System) and has collaborated with Celera (manufacturer of the ViroSeq HIV Genotyping System) and Abbott Diagnostics on evaluation of HIV-related assays. All authors report no potential conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

**References**


