MAJOR ARTICLE

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,[3] Rachel Debes,[4] Ying Q. Chen,[3] Craig A. Martens,[5] Stacy M. Ricklefs,[5] Ethan J. Selig,[5] 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. Hughes[3,4]

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

Received 1 June 2011; accepted 6 July 2011.

Presented in part: 6th IAS Conference on HIV Pathogenesis, Treatment and Prevention, Rome, Italy, 17–20 July 2011.

Correspondence: Susan H. Eshleman, MD/PhD, Department of Pathology, Johns Hopkins University, School of Medicine, Room 646, 720 Rutland Ave, Baltimore, MD 21205 (seshlem@jhmi.edu).

The Journal of Infectious Diseases 2011;204:1918–26

0022-1899 (print)/1537-6613 (online)/2011/20412-0015$14.00
DOI: 10.1093/infdis/jir651

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

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

Region	No. SC Events	Subtype	No. control Samples[a]	Phylogenetic analysis (*pol* region)[b]		Bayesian analysis (*pol* region)[c]		NGS (*env* region)			Final linkage assessment[d]		
				Linked	Unlinked	Linked	Unlinked	Linked	Unlinked	ND	Linked	Unlinked	TBD
Africa	31	C	50	25	6	22	9	3	5	2	24	5	2
Asia	3	C/AE	20	2	1	2	1		1		2	1	
America[e]	4	B/F	10	3	1	3	1		1		3	1	
	38[f]		80	30	8	27	11	3	7	2	29	7	2

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).

[e] ''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.

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 from an interim analysis adequately demonstrated a prevention benefit of early ART initiation and a treatment benefit of early ART to HIV-infected index participants. The DSMB recommended release of the primary study results, which are described elsewhere [6]. Thirty-nine seroconversion events occurred during follow-up in HPTN 052 by the April 2011 DSMB meeting; 1 of those events occurred shortly before the DSMB, and samples were not available for analysis. Analysis was performed to assess the genetic linkage of HIV in the remaining 38 seroconversion events. On the basis of data available at the time of the DSMB meeting, the difference in HIV transmission in the immediate study arm versus the delayed ART study arm based on linked events was highly statistically significant (hazard ratio, 0.04; 95% confidence interval [CI], <.01 to .27; $P < .001$). This report presents the methods and results used to assess the linkage of HIV seroconversion events in HPTN 052 and the association of demographic, behavioral, and clinical factors associated with linked versus unlinked HIV infection. 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/mm^3, 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/mm^3 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.

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 DNADIST. Phylogenetic trees were also generated using maximum likelihood criterion 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):

$$\Pr(Y_i=1|s_i)=\left[p_0 * f_1(s_i)\right]/\left[p_0 * f_1(s_i)+(1-p_0) * f_0(s_i)\right],$$

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

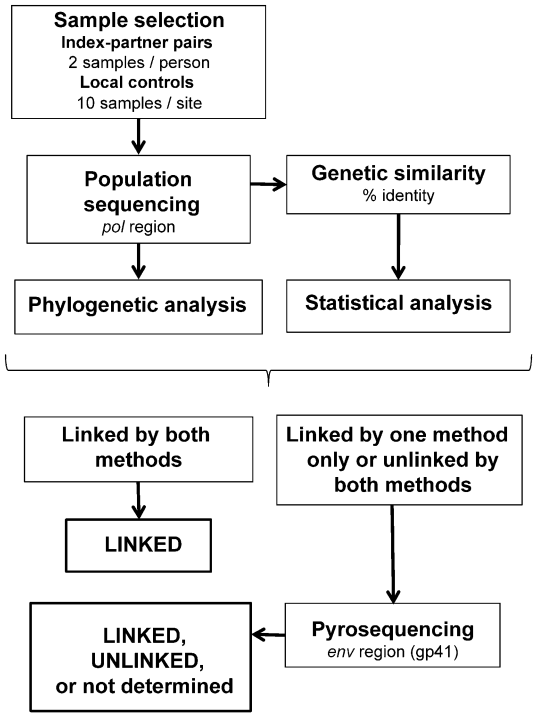

Figure 1. Outline of approach used for linkage analysis (see text).

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).

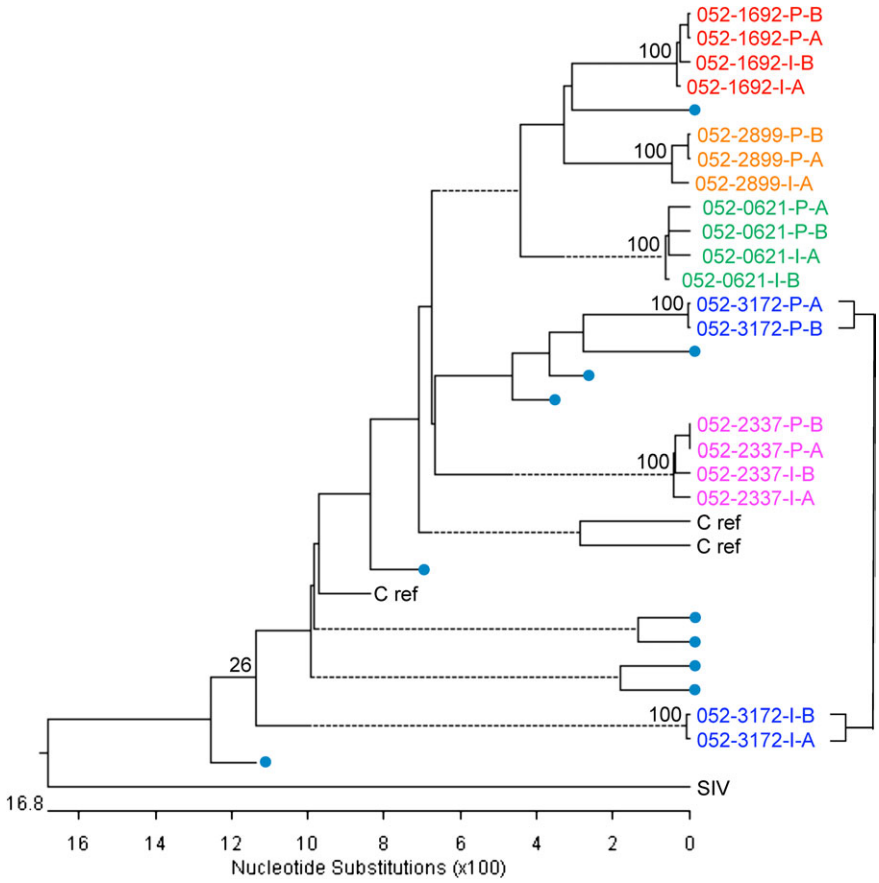

Figure 2. Example of phylogenetic analysis of *pol* sequences from index-partner pairs. A phylogenetic tree was generated for 5 seroconversion events from Africa. The tree includes sequences from index participants and their partners (*colored text*: I, index; *P*, partner). Sequences from 2 different study visits were available for 9 of the 10 participants analyzed (*A*, sequence from the index or partner collected at the earlier study visit; *B*, sequence from the index or partner collected at the later study visit). The tree also includes subtype-matched reference sequences (*C*, ref), 10 local control sequences (randomly selected index participants from the same study site *[blue dots]*) and an outgroup sequence (CPZ.GA.CPZGAB, SIV). Bootstrap values are shown for the grouping of index and partner sequences for each event based on 500 bootstrap replications. In all cases, paired sequences from a given participant grouped together. In 4 cases, all of the sequences from a given event (all index and partner sequences from an index-partner pair) grouped on a single branch with a bootstrap value of 100% (052-1692, 052-2899, 052-0621, and 052-2337), indicating that the events were linked. In 1 case, the 2 sequences from the index did not group with the 2 sequences from the partner (052–3172), indicating that the event was unlinked.

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

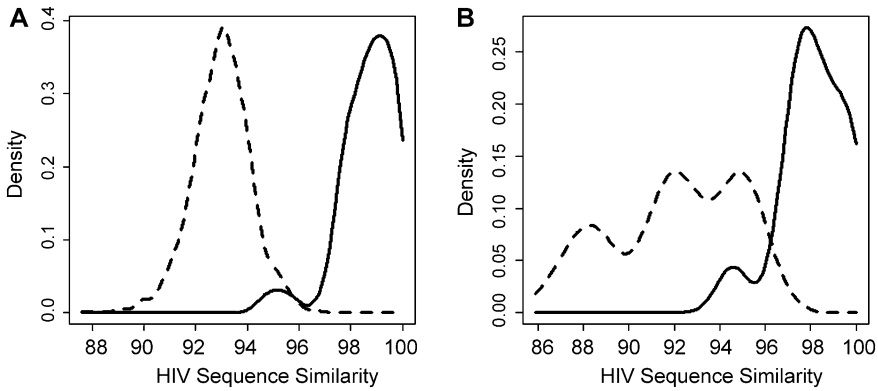

Figure 3. Densities of similarities from linked and unlinked sequence pairs. *A*, Densities of similarities from linked and unlinked sequence pairs from 32 subtype C couples, along with subtype C training data. *B*, Densities of similarities from linked and unlinked sequence pairs from 6 couples with non–subtype C human immunodeficiency virus (HIV) infection, along with non-subtype C unlinked training data and all linked training data (there were insufficient data to estimate a non-subtype C linked density).

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 4*A*). 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 4*B*).

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[a]

	Seroconversion event identification number				
Linkage criterion[b]	052-1692	052-2899	052-0621	052-3172	052-2377
No. of available samples	4	3	4	4	4
Phylogenetic clustering	Grouped	Grouped	Grouped	Not grouped	Grouped
Bootstrap value	100%	100%	100%	ND	100%
Sequence similarity	98.2–98.5	97.9–98.0	96.9–98.5	92.2	98.5
Bayesian posterior probability, pooled	1.00	0.99–1.00	0.94–1.00	0.00	1.00
Bayesian posterior probability, unpooled	1.00	1.00	0.99–1.00	0.00	1.00
Final linkage assessment	Linked	Linked	Linked	Unlinked	Linked

[a] Results are shown from analysis of 5 seroconversion events from Africa (the same events included in Figure 2).

[b] No. available sequences indicates the total number of index and partner samples available for the analysis of each seroconversion event; the number 4 indicates that samples were available for 2 different study visits for both the index and the partner. Phylogenetic clustering: ''Grouped'' indicates that all available index and partner sequences were grouped together on a single, unique branch. ''Not grouped'' indicates that the sequences from the partner did not group with the sequences from the index participant. Bootstrap values are for the grouping of index and partner sequences. A bootstrap value was not determined for the unlinked event (052-3172, not determined [ND]); that event was confirmed to be unlinked based on results from next-generation sequencing.

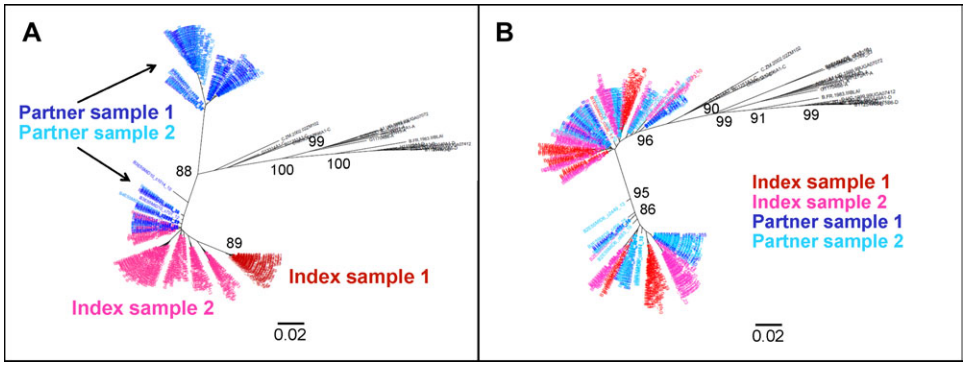

Figure 4. Next-generation sequencing results for the *env* region (gp41). The figure shows phylogenetic trees generated using data from next-generation sequencing (NGS, *env* region [gp41]). Each panel shows results obtained for 1 seroconversion event where linkage of the index and partner was established based on NGS data. Merged neighbor-joining trees were constructed using subtype reference sequences (*black text*, subtypes A, B, C, and D) and gp41 consensus NGS sequences from the relevant index participant (2 samples) and partner (2 samples). Note that only consensus sequences with ≥10 primary NGS reads were used. Color-coding indicates sequences from individual samples (*red*, index sample 1; *pink*, index sample 2; *dark blue*, partner sample 1; *teal*, partner sample 2). Bootstrap values are shown for selected branches (percentage of trees where the sequences group together, among 500 trees generated). Genetic distance is shown (*scale line*). *A*, Data for the seroconversion event from couple 052-2989. For this event, the second index sample was collected 1211 days after the first index sample; the partner samples were collected 1169 and 1211 days after the first index sample. The number of primary sequence reads per sample for these 4 samples ranged from 6555 to 13 696, and the number of consensus sequences per sample for these 4 samples ranged from 35 to 108. This event was provisionally characterized as unlinked based on data obtained by phylogenetic and Bayesian analysis of *pol* region sequences. The linkage status was changed to linked based on the results from NGS. *B*, Data for the seroconversion event from couple 052-1168. For this event, the second index sample was collected 466 days after the first index sample; the partner samples were collected 539 and 556 days after the first index sample. The number of primary sequence reads per sample for these 4 samples ranged from 6891 to 11 700, and the number of consensus sequences per sample for these 4 samples ranged from 38 to 80. The linkage status of this event could not be determined based on data obtained by phylogenetic and Bayesian analysis of *pol* region sequences. The linkage status was changed to linked based on the results from NGS. This analysis was repeated using local control sequences in addition to subtype reference sequences. Both of these events (052-2989 and 052-1168) were from Malawi. The local control sequences used in the analysis of each event included the most prominent consensus sequence from each of 22 other individuals enrolled in the HIV Prevention Trials Network (HPTN) 052. Twenty of the 22 individuals were from Africa and had subtype C human immunodeficiency virus (HIV), and 12 of those individuals were from Malawi. The results obtained using subtype reference sequences only (*A* and *B*) or subtype reference sequences plus local control sequences (not shown) did not differ in any meaningful way.

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

Table 3. Factors Associated With Linkage of Seroconversion Events

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

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

[a] *P* value from Fisher exact test.

[b] 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.

[c] *P* value from Wilcoxon rank–sum test.

[d] One participant was on ART due to pregnancy. She was considered not to be taking ART at the time of her partner's seroconversion.

[e] One participant in a linked seroconversion couple did not provide information about the number of sex partners in the 3 months before seroconversion.

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

stronger (rate ratio, 0.04; exact 95% CI, .001–.27; $P < .001$) [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 copyedited. 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 Sussana Lamers (BioInfo Experts) for help submitting NGS data to GenBank, and thank Ronald Swanstrom 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 other 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

1. Lam TT, Hon CC. Tang JW. Use of phylogenetics in the molecular epidemiology and evolutionary studies of viral infections. Crit Rev Clin Lab Sci **2010**; 47:5–49.
2. Eshleman SH, Husnik M, Hudelson S, et al. Antiretroviral drug resistance, HIV-1 tropism, and HIV-1 subtype among men who have sex with men with recent HIV-1 infection. AIDS **2007**; 21:1165–74.
3. Perez-Losada M, Jobes DV, Sinangil F, et al. Phylodynamics of HIV-1 from a phase III AIDS vaccine trial in Bangkok, Thailand. PLoS One **2011**; 6:e16902.
4. Campbell MS, Mullins JI, Hughes JP, et al. Viral linkage in HIV-1 seroconverters and their partners in an HIV-1 prevention clinical trial. PLoS One **2011**; 6:e16986.
5. Celum C, Wald A, Lingappa JR, et al. Acyclovir and transmission of HIV-1 from persons infected with HIV-1 and HSV-2. N Engl J Med **2010**; 362:427–39.
6. Cohen MS, Chen YQ, McCauley M, et al. Treatment and prevention effects of early antiretroviral therapy for HIV-1 infection: results from the multinational HPTN052 randomized controlled trial. New Engl J Med **2011**; 365:493–505.
7. Eshleman SH, Hackett J Jr, Swanson P, et al. Performance of the Celera Diagnostics ViroSeq HIV-1 Genotyping System for sequence-based analysis of diverse human immunodeficiency virus type 1 strains. J Clin Microbiol **2004**; 42:2711–7.
8. Longo MC, Berninger MS, Hartley JL. Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. Gene **1990**; 93:125–8.
9. Apfalter P, Reischl U, Hammerschlag MR. In-house nucleic acid amplification assays in research: how much quality control is needed before one can rely upon the results? J Clin Microbiol **2005**; 43: 5835–41.
10. Kishino H, Hasegawa M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol **1989**; 29: 170–9.
11. Zwickl DJ. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD thesis. Austin: The University of Texas, **2006**.
12. Scott DW. Multivariate density estimation. Theory, practice and visualization. New York, NY: John Wiley & Sons, Inc., **1992**.
13. Pacold M, Smith D, Little S, et al. Comparison of methods to detect HIV dual infection. AIDS Res Hum Retroviruses **2010**; 26:1291–8.
14. Craig DW, Pearson JV, Szelinger S, et al. Identification of genetic variants using bar-coded multiplexed sequencing. Nat Methods **2008**; 5:887–93.
15. Redd AD, Collinson-Streng A, Martens C, et al. Identification of HIV superinfection in seroconcordant couples in Rakai, Uganda using next generation deep sequencing. J Clin Microbiol **2011**; 49:2859–67.
16. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **1994**; 22:4673–80.
17. Trask SA, Derdeyn CA, Fideli U, et al. Molecular epidemiology of human immunodeficiency virus type 1 transmission in a heterosexual cohort of discordant couples in Zambia. J Virol **2002**; 76:397–405.