Title:

HIV-1 transmission linkage in an HIV-1 prevention clinical trial

Author(s):

Thomas K. Leitner, Z#: 120084, T-6/T Division
Mary S. Campbell, University of Washington-Dept. Medicine
James I. Mullins, University of Washington-Microbiology
James P. Hughes, University of Washington-Biostatistics
Kim G. Wong, University of Washington-Microbiology
Dana N. Raugi, University of Washington-Microbiology
Stefanie Scrensen, University of Washington-Microbiology

Intended for:

Journal: PLOS Medicine

HIV-1 transmission linkage in an HIV-1 prevention clinical trial

Mary S. Campbell[1], James I. Mullins[1,2*], James P. Hughes[3], Kim G. Wong[2], Dana N. Raugi[2], Stefanie Sorensen[2], Julia N. Stoddard[2], Hong Zhao[2], Wenjie Deng[2], Erin Kahle[4], Dana Panteleeff[4], Jared M. Baeten[1,4], Francine E. McCutchan[5], Jan Albert[6], Thomas Leitner[7], Connie Celum[1,4], Anna Wald[1,,9,11], Lawrence Corey[1,11], Jairam R. Lingappa[1,4,10]

Departments of Medicine[1], Microbiology[2], Biostatistics[3], Global Health[4], Laboratory Medicine[8], Epidemiology[9], and Pediatrics[10], University of Washington School of Medicine, Seattle, WA, USA

[5] Bill & Melinda Gates Foundation, Seattle, WA, USA
[6] Karolinska Institute and Swedish Institute for Infectious Disease Control, Solna, Sweden
[7] Los Alamos National Laboratory, Los Alamos, NM, USA
[11] Vaccine and Infectious Disease Institute, Fred Hutchinson Cancer Institute, Seattle WA, USA

* Corresponding author, University of Washington School of Medicine, Box 358070, Seattle, WA, USA. Tel: 1-206-732-6163, Fax: 1-206-732-6167, email: jmullins@uw.edu

Funding:

This work was part of the Partners in Prevention HSV/HIV Transmission Study funded by the Bill and Melinda Gates Foundation (grant ID #26469). Some computational analyses were also supported by a grant through the University of Washington Center for AIDS Research (UW CFAR, AI-27757) Computational Biology Core and a National Institutes of Health grant to JPH (AI029168).

Abstract

Background: HIV-1 sequencing has been used extensively in epidemiologic and forensic studies to investigate patterns of HIV-1 transmission. However, the criteria for establishing genetic linkage between HIV-1 strains in HIV-1 prevention trials have not been formalized. The Partners in Prevention HSV/HIV Transmission Study (ClinicalTrials.gov NCT00194519) enrolled 3408 HIV-1 serodiscordant heterosexual African couples to determine the efficacy of genital herpes suppression with acyclovir in reducing HIV-1 transmission. The trial analysis required laboratory confirmation of HIV-1 linkage between enrolled partners in couples in which seroconversion occurred. Here we describe the process and results from HIV-1 sequencing studies used to perform transmission linkage determination in this clinical trial.

Methods and Findings: Consensus Sanger sequencing of *env* (C2-V3-C3) and *gag* (p17-p24) genes was performed on plasma HIV-1 RNA from both partners within 3 months of seroconversion; *env* single molecule or pyrosequencing was also performed in some cases. For linkage, we required monophyletic clustering between HIV-1 sequences in the transmitting and seroconverting partners, and developed a Bayesian algorithm using genetic distances to evaluate the posterior probability of linkage of participants' sequences. Adjudicators classified transmissions as linked, unlinked, or indeterminate. Among 151 seroconversion events, we found 108 (71.5%) linked, 40 (26.5%) unlinked, and 3 (2.0%) to have indeterminate transmissions. Nine (8.3%) were linked by consensus *gag* sequencing only and 8 (7.4%) required deep sequencing of *env*.

Conclusions: In this first use of HIV-1 sequencing to establish endpoints in a large clinical trial, more than one-fourth of transmissions were unlinked to the enrolled partner, illustrating the relevance of these methods in the design of future HIV-1 prevention trials in serodiscordant couples. A hierarchy of sequencing techniques, analysis methods, and expert adjudication contributed to the linkage determination process.

Introduction

HIV-1 sequencing and phylogenetic analysis have been used for more than 15 years to study intra-host viral evolution and patterns of transmission between individuals and groups at local, regional, and global levels. Several properties of HIV-1, including rapid replication, inaccuracy in reverse transcription, and extensive recombination contribute to viral diversification. This rapid rate of change in HIV-1 has allowed researchers to trace the pathways of viral transmission between individuals with greater accuracy than has been possible with other pathogens. The surface envelope (*env*) glycoprotein (gp120) coding sequence has been used frequently in such studies, as it displays the greatest variability in the HIV-1 genome, evolving within individuals at a rate of ~1% per year [1,2]. The group-specific antigen (*gag*) and polymerase (*pol*) genes have also been used [3][4].

Forensic investigations of HIV-1 transmission also have relied upon phylogenetic analysis of HIV-1 sequence data from suspects and victims [4-7] (reviewed in [8]), and in that context required the highest burden of proof ('beyond a reasonable doubt') to establish linkage. In contrast, the use of viral sequence-based linkage determination in HIV-1 prevention trials has not been described, as most prevention trials involve interventions to reduce HIV-1 acquisition in uninfected persons (e.g., vaccines, microbicides or pre-exposure prophylaxis) rather than interventions provided to HIV-1 infected persons to reduce HIV-1 infectiousness. In prevention trials focusing on HIV-1 infected persons, linking each incident seroconversion to the enrolled HIV-infected source partner minimizes misclassification, thereby maximizing the ability to assess the intervention's efficacy to reduce HIV-1 transmission.

An example of such a trial is the Partners in Prevention HSV/HIV Transmission Study. This phase III, randomized, placebo-controlled trial enrolled African HIV-1 serodiscordant heterosexual couples to evaluate the efficacy of herpes simplex virus type 2 (HSV-2) suppression with acyclovir in reducing HIV-1 transmission from HIV-1/HSV-2 dually-infected participants to their HIV-1-uninfected heterosexual partners [9]. Although the enrollment of stable HIV-1 serodiscordant couples for this study increased the likelihood that HIV-1 strains within seroconverting partner-pairs would be linked, our ability to detect the efficacy of the intervention was improved by linkage confirmation. Here we describe the laboratory and analytic methods used to evaluate the genetic linkage of HIV-1 sequences from couples with partners who seroconverted during the trial and discuss the results and implications of this work.

Methods

The Partners in Prevention HSV/HIV Transmission Study. The study design, recruitment, baseline characteristics and primary study findings of the Partners in Prevention HSV-2/HIV-1 Transmission Study are detailed elsewhere [9-11]. Briefly, 3408 HIV-1 serodiscordant heterosexual couples were enrolled in 7 sub-Saharan African countries. HIV-1/HSV-2 dually-infected partners were randomized to either acyclovir (400 mg orally twice-daily) or placebo and followed for up to 24 months. HIV-1 seroconversions were detected using rapid and enzyme-linked serologic assays [10] and confirmed by HIV-1 Western blot [9]. The primary trial endpoint was defined as incident HIV-1 infection in a previously HIV-1 uninfected partner ('seroconverting partner') confirmed to be genetically linked to his/her putative transmitting partner ('HIV-1 infected partner') by viral sequence analysis.

Overview of transmission linkage methods. Blood plasma collected within 3 months of seroconversion from both partners in each putative HIV-1 transmitting pair was used to perform consensus sequencing of a population of partial HIV-1 *env* and *gag* genes. For those pairs whose sequences did not show clear evidence of linkage by consensus sequencing, multiple single molecule (SM) C2-V3-C3 *env* sequences were obtained following endpoint dilution of cDNA from the HIV-1 infected partner's plasma to identify linked variants present at lower frequency. Furthermore, for a subset of pairs that remained unlinked after both consensus and

SM sequencing, we performed *env* amplicon pyrosequencing of the HIV-1 infected partner's virus population to detect rarer variants that may have been transmitted. Finally, to provide phylogenetic context for the partners' sequences at sites with <10 study-related seroconversions, we sequenced *env* and *gag* from 2-14 HIV-1 infected individuals enrolled at those sites who were epidemiologically unlinked to the putative transmission pairs. An adjudication committee of 3 experts reviewed sequence data to assign linkage classification as described below. Figure 1 shows an overview of these laboratory and analysis methods.

Laboratory methods for HIV-1 sequencing. Technicians were blinded to specimen identification and partnerships. To minimize the risk of specimen mix-up and contamination, laboratory work on HIV-1 infected and seroconverting participants was physically and temporally separated, with pre-PCR steps performed in PCR clean rooms. Viral RNA was extracted from blood plasma using the Qiagen RNA Blood Mini Kit (Qiagen, Valencia, CA). cDNA was synthesized with Superscript III reverse transcriptase (Invitrogen, Carlsbad, CA) and primers RT2 (HXB2 coordinates 3301-3321) and Nef3 (9015-9038). This was followed by nested PCR targeting *gag* (p17-p24 region) and envelope (*env*, C2-V3-C3 region). We used Expand High Fidelity polymerase (Roche Applied Science, Indianapolis, IN) and primers gag1 (772-793), RT2 (3301-3321), ED3 (5957-5986), and Nef3 (9015-9038) for the first round and Taq polymerase (Invitrogen) and primers gag2 (793-818), gag5 (1826-1847), ED31 (6817-6845), and ED33 (7360-7381) for the second round. Final sequence lengths were thus ~1,009bp for *gag* and ~516 for *env*. For SM sequencing on HIV-1 infected partners' plasma, we used endpoint serially diluted cDNA to ensure that PCR amplification of the targeted regions would originate from a single amplifiable template. Clonal sequencing of the C2-V5 region of *env* was performed on some cases using ED31 and BH2 (7697-7725) for the first round and DR7 (6990-7021) and DR8 (7638-7668) for the second round. These amplicons were cloned into TOPO TA vector (Invitrogen). Plasmid DNA and PCR products were purified with the FastPlasmid Mini kit

(Eppendorf, Westbury, NY) and the QIAQuick PCR or gel extraction kit (Qiagen), respectively. Standard dideoxy terminator sequencing was performed at a local facility. Sequence chromatograms were manually edited with Sequencher 4.5 (Gene Codes, Ann Arbor, MI). Amplicon pyrosequencing of C2-V3-C3 *env* (two, ~220bp reads from the 5' and 3' ends of the ED31/ED33 *env* amplicon were obtained and analyzed separately) was performed on HIV-1 infected partners' plasma using the Roche 454 Genome Sequencer 20 (Roche Diagnostics, Branford, CT). The number of templates sequenced ranged from 60-2000 per specimen, estimated by endpoint dilution PCR [12].

Phylogenetic and genetic distance analysis. We screened study sequences against our local laboratory database and with the Los Alamos National Laboratory (LANL) database using ViroBLAST [13] (http://indra.mullins.microbiol.washington.edu/blast/viroblast.php) to identify specimens mixup or laboratory contamination. Viral subtypes were determined using REGA 2.0 (http://dbpartners.stanford.edu/RegaSubtyping) or the NCBI subtyping tool (http://www.ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi). We collected all high quality *env* and *gag* sequences from the LANL database of subtypes A, C, and D from the corresponding gene regions we sequenced, one per subject, and created separate alignments for subtypes A, C, and D for *env* (N = 172, 250, and 97, respectively by subtype) and *gag* (N = 142, 304, and 90, respectively) using CLUSTALW [14] or MUSCLE [15], followed by manual adjustment to optimize codon alignments in Seaview v3 [16] or MacClade 4.08 [17]. We added each study sequence to the appropriate alignment, in some cases along with the 5 most closely related sequences found in the LANL database. Maximum likelihood phylogenetic trees and pairwise distances were determined using the DIVER web server [18] (http://indra.mullins.microbiol.washington.edu/cgi-bin/DIVER/diver.cgi) using a generalized time reversible (GTR) model of evolution.

For cases in which pyrosequencing was performed, reads were initially aligned to an HXB2 reference sequence using Mosaik [19]. We removed reads containing ambiguous bases and of read lengths <100 nucleotides, separated those derived from + and – strands, and manually trimmed trailing ends to remove poor quality data. Local realignments were performed using MUSCLE [15] implemented within the Seaview v3 alignment program [16], followed by further manual refinement in Seaview. Perl scripts were written for Mosaik alignment, conversion of .ace files to .fasta alignments, removal of short reads and those containing N's, sorting alignments at their 5' and 3' ends, and determining pairwise distances to the HIV-1 infected and seroconverter consensus sequences (scripts available upon request).

Reference datasets. We created two reference sequence datasets from individuals with known linkage status to establish the distributions of linked and unlinked *env* and *gag* sequence pairs. The "linked" dataset was derived using sequences from acutely infected individuals and known transmission pairs. This included sequences from the Multicenter AIDS Cohort Study (MACS) [20,21] (acute infections) as well as from heterosexual [22,23], male-to-male (unpublished data and [1]), and mother-to-infant [24,25] transmission cases. Newly available sequences from adjudicator-confirmed linked partner-pairs from this clinical trial were added to the dataset following each of the interim adjudications. In total, sequences from 35/0, 90/57, 117/104, and 147/148 pairs in *env* and *gag* for the first through final adjudications, respectively.

The "unlinked" reference dataset was composed of epidemiologically unlinked sequences using a dataset composed of sequences from individuals with no known epidemiologic linkage, including sequences from the LANL database and from this study cohort. This included, in the final analysis, 362/309, 485/474, 186/133 sequences in *gag* and *env* from subtype A, C and D, respectively.

Bayesian analysis of genetic distances. We developed a Bayesian algorithm to derive an estimate of the probability of linkage based on the genetic distance datasets described above.

According to Bayes' theorem, the posterior probability that two sequences are linked (i.e., the probability of linkage, given existing data) is a function of the prior probability that they are linked and the distributions of genetic distances from known linked and unlinked sequences described above, as shown in the following equation:

$$\mathrm{P(linked}\,|\,\mathrm{X)} = \frac{f(\mathrm{X}\,|\,\mathrm{linked})\mathrm{P(linked)}}{f(\mathrm{X}\,|\,\mathrm{linked})\mathrm{P(linked)} + f(\mathrm{X}\,|\,\mathrm{unlinked})\mathrm{P(unlinked)}}$$

X denotes the objective data obtained during sequence analysis, the pairwise genetic distance in this case. P(linked) and P(unlinked) are the prior probabilities of linkage and lack of linkage for pairs of sequences from HIV-1 infected partner participants in our dataset and f(X | linked) and f(X | not linked) are conditional densities of the genetic distances for linked or unlinked sequences based on the distribution of genetic distances in the reference datasets. As opposed to a 'pure Bayes' approach in which an acceptable value or range of values for P(linked) and P(unlinked) are specified, this approach uses an 'empirical Bayes' approach. Here, an initial value of P(linked) is chosen (P(linked) = 0.5), the posterior probabilities of linkage for each couple is computed, and P(linked) is updated as the proportion of partners who are classified as linked in the Partners in Prevention HSV/HIV Transmission Study. This procedure is then iterated until convergence.

Criteria for assignment of linkage. For each enrolled pair and each level of sequence analysis (consensus, SM and pyrosequencing), HIV-1 linkage was assigned by first requiring that partner-pair derived HIV-1 *env* and/or *gag* sequences form monophyletic clusters (i.e., originating from the same terminal node) in maximum-likelihood phylogenetic trees that included sequences from unrelated individuals ('local controls'). Second, the pairwise genetic distances were required to be associated with a Bayesian posterior probability ≥50%. Partner-pair sequences that met these two requirements were tentatively classified as linked.

An adjudication committee consisting of three independent experts in HIV-1 viral genetics (J.I.M., F.E.M., and either J.A. or T.L) who had not participated in the clinical trial protocol design and were blinded to participants' treatment assignments evaluated phylogenetic and Bayesian linkage probability for each seroconverter pair. If at least two adjudicators concordantly assigned linkage status, the pair was tentatively classified by that assignment. Pairs with linkage status that could not be determined definitively received indeterminate classifications. Interim adjudication occurred before each meeting of the Partners in Prevention HSV-2/HIV Transmission Study Data Safety Monitoring Board and a comprehensive review of the dataset to finalize linkage assignments by consensus was performed before the final clinical trial analysis.

Statistical analysis. Selected epidemiologic and biological variables were compared in linked and unlinked pairs, using two-sided Fisher's exact tests to evaluate for statistical significance.

Ethical review. The University of Washington Human Subjects Review Committee and ethical review committees at each local and collaborating organization approved the Partners in Prevention HSV-2/HIV Transmission Study protocol; the trial was registered in ClinicalTrials.gov (NCT00194519).

Results

During follow-up of the 3408 enrolled couples at 14 sites in 4 East African and 3 southern African countries, HIV-1 serology performed at the study site identified 155 incident infections (Table 1). Of these, 19 seroconversions with negative HIV-1 Western blot and detectable HIV-1 RNA at the time of enrollment (identified by an 'SC' in the partner-pair identifier) were not included as clinical trial endpoints due to HIV infection at randomization, but are included in this linkage analysis of all observed seroconversions. Fifty-six individuals with prevalent HIV-1 infections were evaluated as controls for locally circulating HIV-1 variants. A linkage flow diagram is shown in Figure 2. Four pairs were not confirmed as transmission events, as seroconverting partners were found during confirmatory testing to have negative HIV-1 Western blot and undetectable plasma HIV-1 RNA.

Of the 151 confirmed infections, 108 (71.5%) transmissions were classified as linked to the HIV-1 infected partner. Linkage determination was based on consensus HIV-1 *env* sequence data for 91 (84.3%), consensus *gag* for 9 (8.3%), and SM (from sequencing multiple clones or single molecule-derived amplicons) *env* for 8 (7.4%). Forty transmissions (26.5%) were found to be unlinked and 3 (2.0%) had indeterminate linkage. Among the 151 with confirmed HIV-1 infection in the seroconverting partner, 20 linked and 3 unlinked pairs had successful PCR amplification from only one gene. Table 1 summarizes final linkage status by site, with sequence data for each pair summarized in Supplementary Table S1. Phylogenetic trees are available at (http://www.mullinslab.microbiol.washington.edu/publications_campbell_2009_2).

Phylogenetic analysis. Among the 151 pairs, monophyly, the sharing of the most recent common ancestor (MRCA) on the tree, was found for 16 pairs (10.6%) in *env* only, 9 (6.0%) in *gag* only, and 84 (55.6%) in both *env* and *gag* (Table 1). However, phylogenetic discordance was found in only 3 pairs, as only one gene was successfully amplified in the other 20 cases. Including the 'local controls', we obtained sequences from a median of 16 individuals (range 10-96) at each study site. No local control sequence was found to split a monophyletic linkage between enrolled partners. Hence, linkages were not erroneously assigned due to geographic proximity. Figure 3 shows examples of monophyletic and polyphyletic partner pairs and the corresponding distance and Bayesian posterior probability data used for linkage adjudication.

Genetic distance and Bayesian analysis. The median pairwise genetic distance for linked pairs was 2.8% (range 0.0-13.0%) in *env* and 1.3% (range 0.0-9.2%) in *gag* (Table 2). In unlinked pairs, median distances were 17.2% (range 11.2-34.6%) and 11.3% (range 6.0-21.6%)

in *env* and *gag,* respectively. Distance ranges for linked and unlinked pairs overlapped due to pairs in which linkage was found in only one gene. Two of the three indeterminate pairs had genetic distances within the range of pairs that were linked in *env* (PP92) and *gag* (PP4 and 92). However, only one indeterminate pair (PP92) exceeded the $\geq$50% posterior Bayesian probability cutoff, observed in *env* only. Figure 4A shows the distribution of *env* genetic distances for linked and unlinked study pairs superimposed on the genetic distance distribution for intrasubject and intersubject linked and unlinked reference data (analogous data for *gag* sequences are shown in Supplementary Figure S1). Median Bayesian posterior probabilities for linked and unlinked pairs were 99.8% and 1.0% in *env* and 99.7% and 0.0% in *gag*, respectively. As pairwise distance between couples' sequences increased, the Bayesian posterior probability of linkage decreased rapidly, with the majority of couples' pairwise distances associated with posterior probabilities approaching 1 (100% probability of linkage) or 0 (0% probability of linkage) (Figure 4B).

In 2 instances adjudicated as linked (PP47 and SC1), sequence pairs were monophyletic in *env* and *gag*, but with posterior probabilities <50% for *env* (45.8% and 33.0%) but high for *gag* (99.4% and 99.9%). In *gag*, no monophyletic pairs had posterior probabilities in an intermediate range (Figure 4B and Table S1). One pair (SC6) had sequences that were monophyletic in *gag* with pairwise genetic distance of 5.8% between sequences, but with a Bayesian posterior probability of ~0%. However, since SC6's *env* sequences met all criteria for linkage, it was classified as linked. In only one instance (SC2) did Bayesian analysis suggest linkage (posterior probability of 51.5%) in the absence of monophyly in the same gene (*env*), but since *gag* analysis met both phylogenetic and Bayesian criteria for linkage, this pair was also classified as linked.

Deep sequencing (SM and pyrosequencing) for linkage determination. We evaluated clonal or single molecule (SM) *env* sequences in 43 pairs that were unlinked or indeterminate by consensus sequencing with a median of 19 sequences evaluated per HIV-1 infected participant (range 3-62). Linkage was found in 8 (18.6%), with linked variants constituting 25-50% of the sequences evaluated for each linked pair. An example of the use of SM sequencing to establish linkage for a case (PP17) in which consensus sequences from the HIV-1 infected and seroconverting participants were unlinked is shown in Figure 5. In this case, 3 sequences from the HIV-1 infected partner had distances and Bayesian posterior probabilities that were categorized as linked to the seroconverter, whereas 9 other sequences did not meet this criterion. No relationship was found between classification of a pair as linked and the number of SM *env* sequences obtained.

When sufficient numbers of amplifiable viral templates (N=~50) were available for study, deep resequencing by pyrosequencing was used to probe for low-level variants. In 11 of 12 cases evaluated, involving a median of 119 templates per pair, we failed to detect sequences closely related to that in the seroconverter (Supplemental data, figure 2). In the remaining case (PP92), 3.8% of the sequence reads from ~61 viral templates from the HIV-1 infected case were closely related to viruses found in the seroconverter, as described below.

Viral subtype. HIV-1 subtype was determined for both *env* and *gag* sequences (Tables 1 and S1) from each partner pair. In both genes, participants' viruses were predominantly subtype A or C (43% each in *env*, 44% and 36% in *gag*, respectively), with 13% of the *env* sequences and 10% of the *gag* sequences found to be subtype D. One subtype G infected pair was detected, and 2% of the *env* and 10% of the *gag* sequences were detectably intersubtype recombinants. Among the 128 partner pairs with sequences determined for both *env* and *gag*, 13 pairs (10.1%) had different subtypes in each gene suggesting the presence of additional intersubtype recombinant viruses. In an additional 13 couples (4 linked, 6 unlinked, and 3 indeterminate pairs) discordant subtypes were noted between *env* and *gag* sequences in one partner, without such a discrepancy in the other partner.

When stratified by subtype, no difference in the frequency linked and unlinked pairs was found: among linked partners, 69.8%/68.9%, 66.7%/64%, and 73.7%/73.3% of *env*/*gag* sequences, were subtype A, C, and D, respectively. Among intersubtype recombinants, 84.6% (11/13) were classified as linked.

Discordant findings. Eighty-four (95.5%) of linked pairs having *env* and *gag* sequences met criteria for linkage in both genes. Among those classified as linked, two pairs met criteria for linkage in *gag* only (PP135 and SC2); one pair met criteria for linkage in *env* only (PP133). Participants in these 3 pairs may have been infected by more than one HIV-1 strain. Eleven (7%) pairs had *env* and *gag* sequences of different subtypes, 2 from Kitwe, Zambia, 5 from Kisumu, Kenya and 4 from Kampala, Uganda. Of these, 4 were classified as linked, with concordant *env* and *gag* subtypes between partners, suggesting that a virus with a recombinant subtype may have been transmitted from the HIV-1 infected partner to the seroconverter. In the remaining 7 pairs, 1 indeterminate and 6 unlinked, each partner's virus had a different mosaic subtype pattern.

James I Mullins 9/29/09 11:09 AM
Comment: Mary – please indicate if results were reproduced with additional samples from these people and those mentioned immediately below.

Mullins Lab 9/30/09 4:15 PM
Comment: Reproduced in PP133 & PP135

Adjudicator agreement and indeterminate pairs. At the end of the study, complete agreement was reached between adjudicators' classification of all linked and unlinked pairs. Six (3.9%) pairs required discussion before all three adjudicators determined they were linked at the final adjudication meeting.

Adjudicators were unable to determine the linkage status of 3 pairs in which sequencing was completed. Two pairs' data (PP4 and PP9) were suggestive of linkage in *env* only. PP4's consensus *env* and *gag* sequences were polyphyletic, with distances and Bayesian posterior probabilities outside the expected range for linked transmissions (Table S1). The viral subtype in *env* was C for the female HIV-1 infected participant and A for her male seroconverting partner. After SM *env* sequencing, 1 of 17 sequences from the HIV-1 infected participant was found to be of subtype A and fell in a monophyletic cluster with the partner's sequences.

However, *env* pairwise genetic distances and Bayesian posterior probability were inconsistent with linkage, as was *gag* data, so Pair 4 was categorized as indeterminate.

Similarly, consensus *env* sequences from the female HIV-1 infected and male seroconverting partners of PP9 were of different subtypes (C and A, respectively). Both consensus *env* and *gag* sequences were polyphyletic and with large distances (25.5 and 16.1%, respectively). SM *env* sequencing from both participants (N = 16 and 29, respectively) did not reveal any more closely related sequences. However, approximately 61 *env* templates from the HIV-1 infected participant were pyrosequenced, which did reveal a variant that was closely related to the seroconverting partner's virus, comprising 3.8% of the viral population on the 3' ends of the amplicon, with no close relatives above the 100nt cutoff read length from the 5' end reads (4 short reads, corresponding to 0.2% of the total sequences were found to be related to the seroconverter consensus but were discarded due to poor quality). The adjudication team concluded that the small fraction of related sequences found by a sequencing technique that is still in development for applications related to HIV-1 evolution did not provide sufficient evidence to categorize this pair as linked.

Finally, PP92's consensus *env* sequences were monophyletic, but with a large pairwise distance of 8.9%. After consensus *gag* sequences were found to be polyphyletic and relatively distant (6.8%) and 17 SM *env* sequences from the HIV-1 infected partner did not reveal a sequence with a smaller genetic distance to the serooconverter's virus, this pair was also classified as indeterminate.

Epidemiologic support for linkage assignments. We evaluated epidemiologic support for our linkage assignments by comparing demographic and clinical characteristics of linked and unlinked partnerships. The seroconverting partner was male in 88 (58.3%) and female in 63 (41.7%) of the 151 couples, reflecting in part the study enrollment gender distribution of 67% of enrolled HIV-1 infected partners being female. However, seroconverters were female in a larger proportion of linked relative to unlinked pairs (46.3% versus 27.5%, $p = 0.04$). The timing of seroconversion also was associated with linkage, with linked pairs having a shorter average time to seroconversion than unlinked pairs (6 versus 12 months, $p = 0.001$). Similarly, there was a trend toward the proportion of linked transmissions being greater among seroconversions identified at the first 3-month study visit compared to seroconversions identified after 3 months (89.5% versus 66.9%, $p = 0.06$). Sexual activity with non-enrolled partners was reported more commonly by unlinked than linked seroconverters (30% versus 1.9%, $p < 0.001$). Finally, baseline plasma HIV-1 RNA levels for the HIV-1 infected partner were higher among linked pairs than unlinked pairs (4.7 versus 4.0 $\log_{10}$ copies/ml, $p < 0.001$).

Discussion

We used phylogenetic and genetic distance data to confirm HIV-1 transmission linkage in an HIV-1 prevention clinical trial involving HIV-1 serodiscordant couples from East and southern Africa predominately infected with clade A and C HIV-1 strains, and found that over one quarter (26.5%) of putative transmission events were not linked to the enrolled partner.

Our analysis represents the first use of viral sequencing for transmission linkage as an integral component in the primary analysis of a large randomized HIV-1 prevention trial. As with previous linkage assessments in observational studies [22], we evaluated sequence data from both *env* and *gag* for monophyly in maximum likelihood trees to determine linkage. However, to provide additional statistical support for our linkage determinations, we developed a Bayesian algorithm incorporating prior probability of linkage and genetic distance data and increased our sensitivity for rare sequence variants using deep sequencing techniques. While consensus *env* sequencing identified 85% of linked pairs, *gag* and deep *env* sequencing permitted classification of 9 (8.3%) and 8 (7.4%) linked pairs, respectively, that would not have been linked if only consensus *env* were used to define linkage. Overall, linked and unlinked pairs were clearly separated by phylogenetic relationships, genetic distance, and Bayesian posterior probability estimates.

Our methods were robust, with individual adjudicators reaching identical independent assessments in 96% of cases, followed by full concurrence after discussion. In only 3 (2.0%) of cases were adjudicators unable to determine the linkage status conclusively, possibly due to HIV-1 dual or superinfection followed by recombination of viral strains. Additional deep or whole genome sequencing could resolve such indeterminate classifications, but was beyond the scope of the clinical trial.

The high fraction of unlinked infections we found (26.5%), differs from a cohort study of HIV-1 serodiscordant couples in Zambia from 1994-2000, in which 13% of prospectively identified seroconverters had viruses not linked to their stable partner based on consensus *env* or *gag* sequences [22]. While our study cannot be directly compared to the Zambian study due to differences in design, locations, and periods of conduct, it is notable that male partners were HIV-1 infected at baseline in a greater proportion of the couples in the Zambian cohort compared to our cohort (52% versus 33%). The strong association we found between unlinked transmission and reported sexual activity with non-enrolled partners, and the higher likelihood we found for female seroconverters to be infected from their stated partners, corroborates our linkage assignments and suggests that behavioral rather than biological factors underlie the higher rate of non-linkage in our cohort. The design of future clinical trials in HIV-1 serodiscordant couples should take the risk of HIV-1 transmission from non-enrolled sexual partners into account.

In summary, we determined the linkage status of HIV-1 strains in 151 sub-Saharan African couples enrolled in a trial to evaluate an intervention provided to HIV-1 infected individuals to prevent transmission to their HIV-1 uninfected partners. Our methods were efficient and accurate and our linkage classifications were supported by epidemiologic characteristics of these subgroups. Given that >25% of transmissions may arise from sexual activity with non-enrolled partners, linkage analyses such as ours should be considered for use in future HIV-1 prevention trials, particularly in serodiscordant couples in which the intervention is focused on the HIV-1 infected partner. These methods will also be essential for characterizing putative transmission pairs in advance of studies of virology, immunology, and host genetics associated with HIV-1 transmission.

References

1. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, et al. (1999) Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol 73: 10489-10502.
2. Wolinsky SM, Korber BT, Neumann AU, Daniels M, Kunstman KJ, et al. (1996) Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. Science 272: 537-542.
3. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, et al. (2001) Evolutionary and immunological implications of contemporary HIV-1 variation. Br Med Bull 58: 19-42.
4. Albert J, Wahlberg J, Uhlen M (1993) Forensic evidence by DNA sequencing. Nature 361: 595-596.
5. Holmes EC, Zhang LQ, Simmonds P, Rogers AS, Brown AJ (1993) Molecular investigation of human immunodeficiency virus (HIV) infection in a patient of an HIV-infected surgeon. J Infect Dis 167: 1411-1414.
6. Holmes EC, Brown AJ, Simmonds P (1993) Sequence data as evidence. Nature 364: 766.
7. Ou CY, Ciesielski CA, Myers G, Bandea CI, Luo CC, et al. (1992) Molecular epidemiology of HIV transmission in a dental practice. Science 256: 1165-1171.
8. Learn GH, Mullins, James I. (2003) The Microbial Forensic Use of HIV Sequences. In: Leitner T. FB, Hahn B., Marx P., McCutchan F., Mellors J., Wolinsky S., Korber B, editor. HIV Sequence Compendium. Los Alamos, NM: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory. pp. 22-37.
9. Celum C, Wald A, Lingappa R, Magaret A, Wang RS, Mugo N, Mujugira A, Baeten JM, Mullins JI, Hughes J, Bukusi EA, Cohen CR, Katabira E, Ronald A, Kiarie J, Farquhar C, John Stewart G, Makhema J, Essex M, Were E, Fife KH, de Bruyn G, Gray GE, McIntyre J, Manongi R, Kapiga S, Coetzee D, Allen S, Inambao M, Kayitenkore K, Karita E, Kanweka W, Delany S, Rees H, Vwalika B, Stevens W, Campbell MS, Thomas K, Coombs RW, Morrow R, Whittington WLH, McElrath MJ, Barnes L, Ridzon R, and Corey L for the Partners in Prevention HSV/HIV Transmission Study Team (Submitted) Twice-daily acyclovir to reduce HIV-1 transmission from HIV-1 / HSV-2 co-infected persons within HIV-1 serodiscordant couples: a randomized, double-blind, placebo-controlled trial.
10. Lingappa JR, Kahle E, Mugo N, Mujugira A, Magaret A, et al. (2009) Characteristics of HIV-1 discordant couples enrolled in a trial of HSV-2 suppression to reduce HIV-1 transmission: the partners study. PLoS One 4: e5272.

11. Lingappa JR, Lambdin B, Bukusi EA, Ngure K, Kavuma L, et al. (2008) Regional differences in prevalence of HIV-1 discordance in Africa and enrollment of HIV-1 discordant couples into an HIV-1 prevention trial. PLoS One 3: e1411.

12. Rodrigo AG, Goracke PC, Rowhanian K, Mullins JI (1997) Quantitation of target molecules from polymerase chain reaction-based limiting dilution assays. AIDS Res and Hum Retrovir 13: 737-742.

13. Deng W, Nickle DC, Learn GH, Maust B, Mullins JI (2007) ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. Bioinformatics 23: 2334-2336.

14. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.

15. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.

16. Galtier N, Gouy M, Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 12: 543-548.

17. Maddison WP, Maddison DR (2001) MacClade - Analysis of Phylogeny and Character Evolution - Version 4. Sunderland, MA: Sinauer Associates, Inc. 503 p.

18. Deng W, Maust BS, Nickle DC, Learn GH, Liu Y, et al. (Submitted) DIVER: a Web Server to Perform, Summarize and Visualize Molecular Sequence Divergence, Diversity and Phylogenetic Analyses.

19. Strömberg M (2007) Mosaik is a suite comprising of three modular programs: MosaikBuild, MosaikAligner, and MosaikAssembler.

20. Gottlieb GS, Heath L, Nickle DC, Wong KG, Leach SE, et al. (2008) HIV-1 variation before seroconversion in men who have sex with men: analysis of acute/early HIV infection in the multicenter AIDS cohort study. J Infect Dis 197: 1011-1015.

21. Shankarappa R, Gupta P, Learn GH, Jr., Rodrigo AG, Rinaldo CR, Jr., et al. (1998) Evolution of human immunodeficiency virus type 1 envelope sequences in infected individuals with differing disease progression profiles. Virology 241: 251-259.

22. Trask SA, Derdeyn CA, Fideli U, Chen Y, Meleth S, et al. (2002) Molecular epidemiology of human immunodeficiency virus type 1 transmission in a heterosexual cohort of discordant couples in Zambia. J Virol 76: 397-405.

23. Leitner T, Escanilla D, Franzen C, Uhlen M, Albert J (1996) Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. Proc Natl Acad Sci U S A 93: 10864-10869.

24. Wu X, Parast AB, Richardson BA, Nduati R, John-Stewart G, et al. (2006) Neutralization escape variants of human immunodeficiency virus type 1 are transmitted from mother to infant. J Virol 80: 835-844.

25. Hahn T, Matala E, Chappey C, Ahmad N (1999) Characterization of mother-infant HIV type 1 gag p17 sequences associated with perinatal transmission. AIDS Res Hum Retroviruses 15: 875-888.

Table 1. Transmission pairs and local prevalent infections evaluated and linkage findings by study site and region. († other subtype)

Region	Country	Site	# Putative Transmission Pairs	# with Monophyly in *env* or *gag* & Post. Prob. ≥50%	*env* Subtype				*gag* Subtype				# Linked	# Unlinked	# Ind.	# Local Controls
					A	C	D	†	A	C	D	†				
East Africa (EA)	Kenya	Eldoret	6	6	3	0	2	0	4	0	1	1	6	0	0	2
		Kisumu	42	26	28	5	6	1	28	2	4	6	26	14	0	12
		Nairobi	11	8	8	1	2	0	7	0	1	3	8	2	1	2
		Thika	5	5	5	0	0	0	5	0	0	0	5	0	0	0
	Tanzania	Moshi	2	2	2	0	0	0	1	0	0	1	2	0	0	6
	Uganda	Kampala	30	21	16	3	16	1	16	0	9	2	21	7	0	14
EA Subtotal			96	68	62	9	26	2	61	2	15	13	68	23	1	36
Southern Africa (SA)	Botswana	Gaborone	7	4	0	7	0	0	0	7	0	0	4	3	0	0
	South Africa	Cape Town	15	10	0	16	0	0	0	17	0	0	9	5	1	2
		Orange Farm	5	3	0	11	0	0	0	11	0	0	3	2	0	6
		Soweto	6	5	0	9	0	0	0	5	0	0	5	1	0	3
	Zambia	Kitwe	10	6	1	9	0	0	0	5	0	0	6	3	1	0
		Lusaka	3	2	0	10	0	0	0	9	0	0	2	1	0	7
		Ndola	13	11	0	15	0	0	0	13	0	0	11	2	0	2
SA Subtotal			59	41	1	77	0	0	0	67	0	0	40	17	2	20
Total			151	109	63	86	26	2	61	69	15	13	108	40	3	56

Table 2. Comparison of monophyly, genetic distance, and Bayesian posterior probability results by linkage status and gene.

	Linked (N = 108)		Unlinked (N = 40)		Indeterminate (N = 3)	
Criterion	*env*	*gag*	*env*	*gag*	*env*	*gag*
Proportion with Monophyly	0.935	0.880	0	0	0.667	0
Median Pairwise Genetic Distance (range)	0.028 (0.000-0.130)	0.013 (0.000-0.092)	0.172 (0.112-0.346)	0.113 (0.060-0.216)	0.188 (0.089-0.255)	0.079 (0.068-0.161)
Median Bayesian Posterior Probability (range)	0.998 (0.038-1.000)	0.997 (0.000-1.000)	<0.001 (0.000-0.271)	0.000 (0.000-0.000)	0.000 (0.000-0.743)	0.000 (0.000-0.000)

Table 3. Association of demographic and clinical factors with linkage.

	All Pairs (N = 155)	Linked Pairs (N = 108)	Unlinked Pairs (N = 40)	p-value
Gender				
SC partner female	64 (41.3%)	50 (46.3%)	11 (27.5%)	0.0412
Median age at enrollment (range)				
SC partner	30 (26-38)	30 (25-38)	32 (26-36)	0.425
HIV-infected partner	31 (25-37)	30.5 (26-36)	31 (25-39.5)	0.935
Time to seroconversion				
Identified at 3 month visit	19 (12.3%)	17 (15.7%)	2 (5.0%)	0.101
Months of follow-up before seroconversion	9 (3-15)	6 (3-15)	12 (9-17)	0.001
Study site location				
East Africa	96 (61.9%)	68 (63.0%)	23 (57.5%)	0.572
Southern Africa	59 (38.1%)	40 (37.0%)	17 (42.4%)	0.572
Behavioral characteristics of SC partner prior to seroconversion				
Reported unprotected sex with HIV-infected partner	46 (29.7%)	36 (33.3%)	9 (22.5%)	0.23
Reported relationship with a non-enrolled partner	14 (9.0%)	2 (1.9%)	12 (30.0%)	<0.001
Reported unprotected sex with a non-enrolled partner	0 (0%)	0 (0%)	0 (0%)	-
Herpes and STI in SC partner				
SC partner HSV-2+ at enrollment	127 (81.9%)	86 (79.6%)	35 (87.5%)	0.342
Any STI at enrollment	32 (22.1%)	21 (20.8%)	9 (24.3%)	0.648
Characteristics of HIV-infected partner				
Enrollment plasma HIV-1 RNA ($\log_{10}$)	4.6 (4.0-5.1)	4.7 (4.3-5.1)	4.0 (3.5-4.8)	<0.001
CD4 count at visit closest to seroconversion	379 (281-506)	364 (255-495)	369 (307-502)	0.323
ARV at visit prior to seroconversion	3 (1.9%)	1 (0.9%)	2 (5.0%)	0.178

* Comparison of linked and unlinked transmission pairs

SC = seroconverting

Figure Legends

Figure 1. Overview of laboratory and analysis methods. (A) Overview of laboratory methods. RNA was extracted from blood plasma, cDNA synthesized, and multiplex PCR targeting *env* and *gag* was performed. Sequences were aligned and analyzed in the context of reference and 'local control' sequences of the same subtype. Phylogenetic relationships, pairwise genetic distances, and Bayesian posterior probabilities were obtained. **(B)** Process by which posterior probabilities of linkage were obtained. The linked dataset corresponded to sequences derived from the Los Alamos National Laboratory HIV database (HIVDB) and trimmed to match the amplicons sequenced in the current study in *env* and *gag*. The linked dataset was composed of intrasubject sequences from <2 years of infection from the MACS, from available linked partner pairs from the literature and intermediate adjudications in this study, and from mother-infant transmission pairs. Three unlinked datasets were initially derived, from HIV-1 subtypes A, B and C, one sequence per subject and from individuals with no known epidemiologic linkage. After each set of sequences were aligned, pairwise distances were determined and the each dataset combined to create one “linked' and one “unlinked” pairwise distance dataset. Alignments are available at (http://www.mullinslab.microbiol.washington.edu/publications_campbell_2009_2). These datasets were used to estimate prior probabilities of linkage using the Bayesian approach described in Methods. **(C)** Schematic of the process by which linkage was assessed. For each pair, adjudicators evaluated monophyly (yes/no), genetic distance, and Bayesian posterior probability (≥0.5 or <0.5) and classified the pair as 'linked', 'unlinked', or 'indeterminate'. Further evaluation of 'unlinked' or 'indeterminate' pairs

involved gathering additional data, including sequencing of consensus *gag* and/or clonal, single molecule or pyrosequencing of *env*, as well as obtaining sequences from non-transmitting HIV-1 infected participants from the same study site. New trees, distance distributions and Bayesian priors were generated and each pair was re-adjudicated to make final linkage assignments.

Figure 2. Flow chart of sequences obtained and linkage results for all pairs evaluated. *Consensus *gag* sequence analysis contributed 5 linkages in eligible pairs and 4 linkages in 3-month seroconverters (circles) over consensus *env* sequencing alone. Deep sequencing by clonal or single molecule (SM) and amplicon pyrosequencing (pyro) of *env* revealed 8 additional linked pairs. Deep sequencing was not performed in 3-month seroconverter pairs, as they were not included in the modified intention to treat analysis.

Figure 3. Examples of linked monophyletic (PP73 and PP82) and unlinked polyphyletic (PP45) pairs and the adjudication criteria for each.

Figure 4. **(A)** Distributions of pairwise genetic distances for *env* reference datasets, within acutely infected individuals from the Multicenter AIDS Cohort Study at different intervals post infection, between epidemiologically-unlinked pairs of sequences from the Los Alamos National Laboratory database of subtypes A, C, and D (lines) and between enrolled partner-pairs from the Partners in Prevention cohort that were adjudicated as linked (red bars) and unlinked (blue bars) through sequencing of *env*, *gag*, or both. To improve visibility of the data, the y-axis scale ranges from 0 to 0.25 for bars representing the Partners in Prevention cohort. **(B)** Plot showing relationship between Bayesian posterior probabilities and genetic distance between partner pairs from the Partners in Prevention cohort in *env* and *gag*.

Figure 5. Example of a pair (Pair 17) whose consensus *env* sequences were unlinked, with linkage subsequently determined by single molecule *env* sequences. The linkage criteria used during adjudication are displayed in the table. Three linked sequences from the HIV-1 infected partner, PP17A variant 1, along with the sequences from the seroconverting partner PP17B are bounded by the solid rectangle. Unlinked sequences from the HIV-1 infected partner, PP17A variant 2 are delineated by the dotted rectangle.

Supplementary Material

Supplementary Table S1. Demographic, sequence, and linkage data for each pair. Pairwise nucleotide distances shown are the smallest pairwise distances obtained, from either consensus or single molecule sequencing in *env* or consensus sequencing in *gag*, with Bayesian posterior probabilities corresponding to the distances shown.

Supplementary Table S2. Pyrosequencing analysis of the enrolled partner's *env* sequences in pairs of individuals without prior evidence of linkage. The approximate number of templates evaluated in each pyrosequencing reaction are shown, along with the number of raw and final reads used in the evaluation. 400 bp amplicons were sequenced using primers from the 5' and 3' ends. The ~220 bp reads from each end were analyzed separately. A variable number of sequences were removed from the final alignments as described in the Methods.

Supplementary Figure 1. Distributions of pairwise genetic distances for *gag* reference datasets and between enrolled partner-pairs from the Partners in Prevention cohort that were adjudicated as linked (red bars) and unlinked (blue bars) through sequencing of *env*, *gag*, or both.

Supplementary Figure 2. Pyrosequencing analysis. Each panel shows the distribution of pairwise genetic distances between a reference sequence (the consensus of *env* sequences from each seroconverting partner) and pyrosequences derived from the index partner. See Table S2 for details. The graph on the left side of each panel shows the analysis of the 5' and 3' reads, respectively. Distributions marked in blue indicate the relationship of the enrolled partners' sequences to the consensus of the seroconverting partners' sequence. Distributions marked in red indicate the relationship of enrolled partners' sequences to the consensus of the index partners' sequence.

Figure 1a, Campbell et al

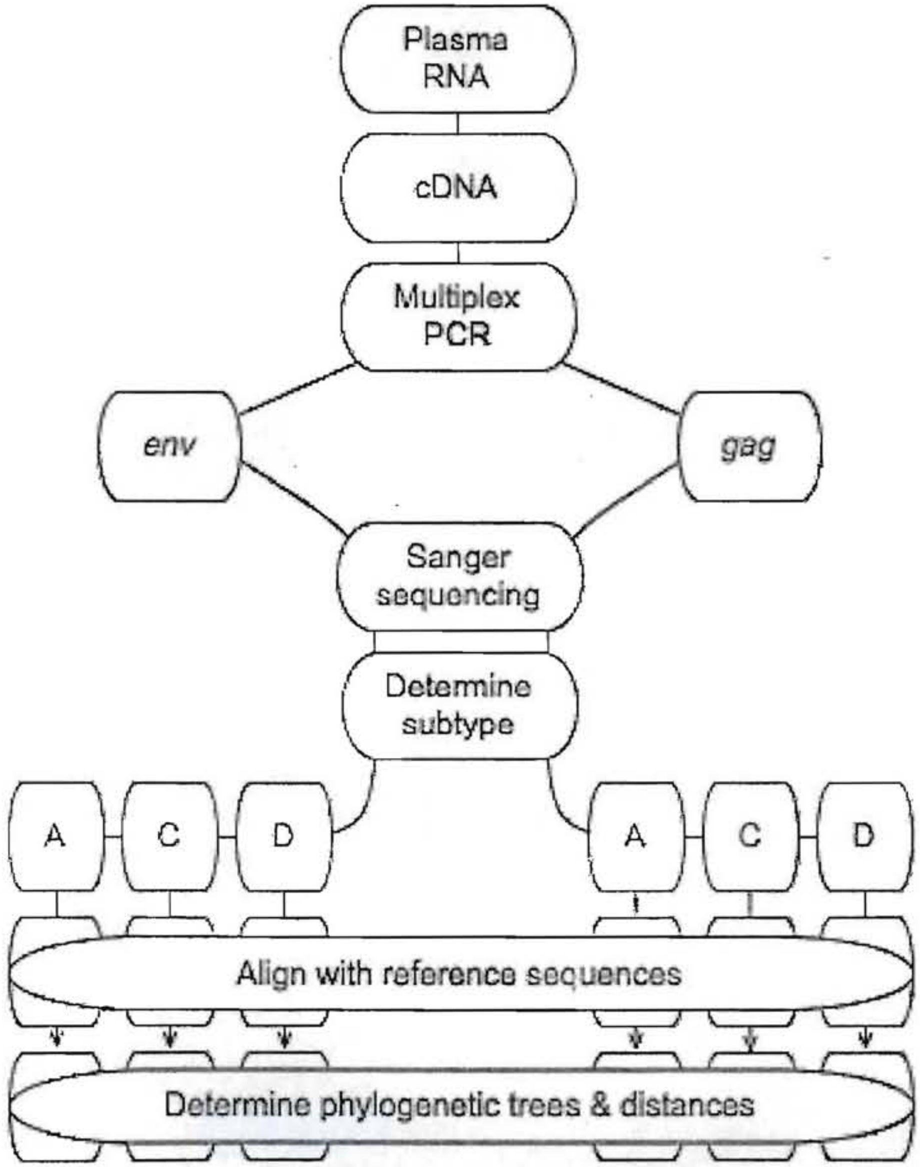


Figure 1b, Campbell et al

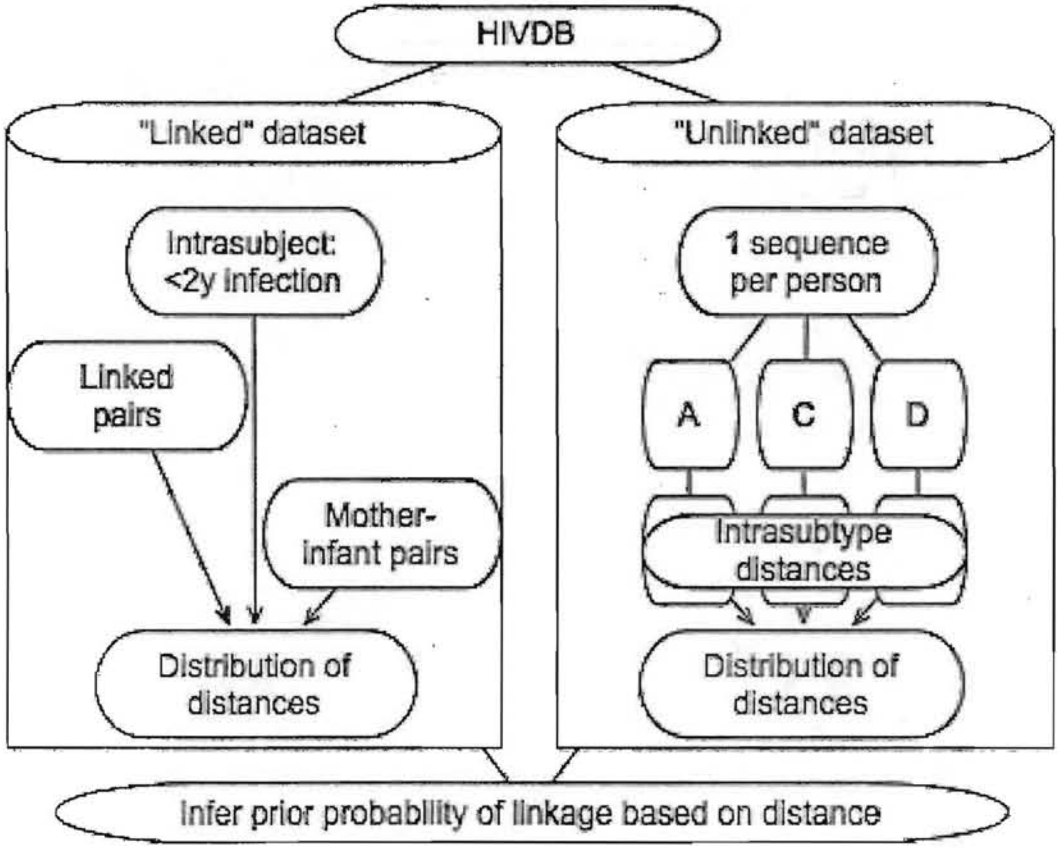


Figure 1c, Campbell et al

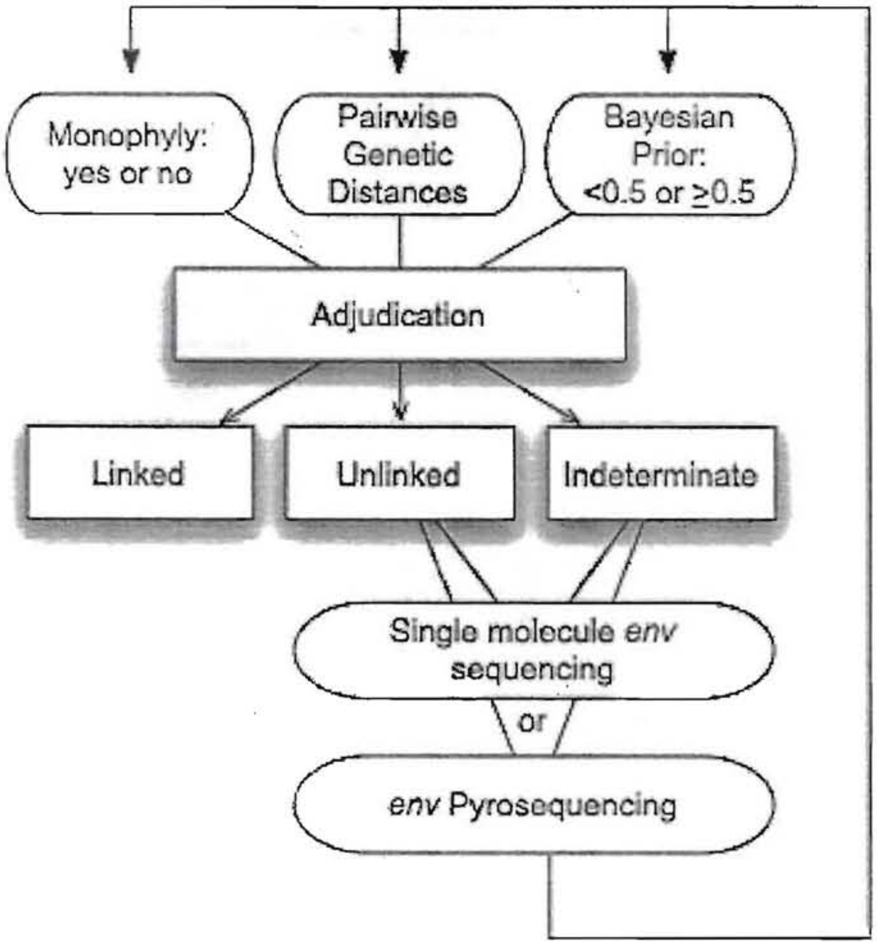

Monophyly: yes or no
Pairwise Genetic Distances
Bayesian Prior: <0.5 or ≥0.5
Adjudication
Linked
Unlinked
Indeterminate
Single molecule *env* sequencing
or
env Pyrosequencing

Figure 2, Campbell et al

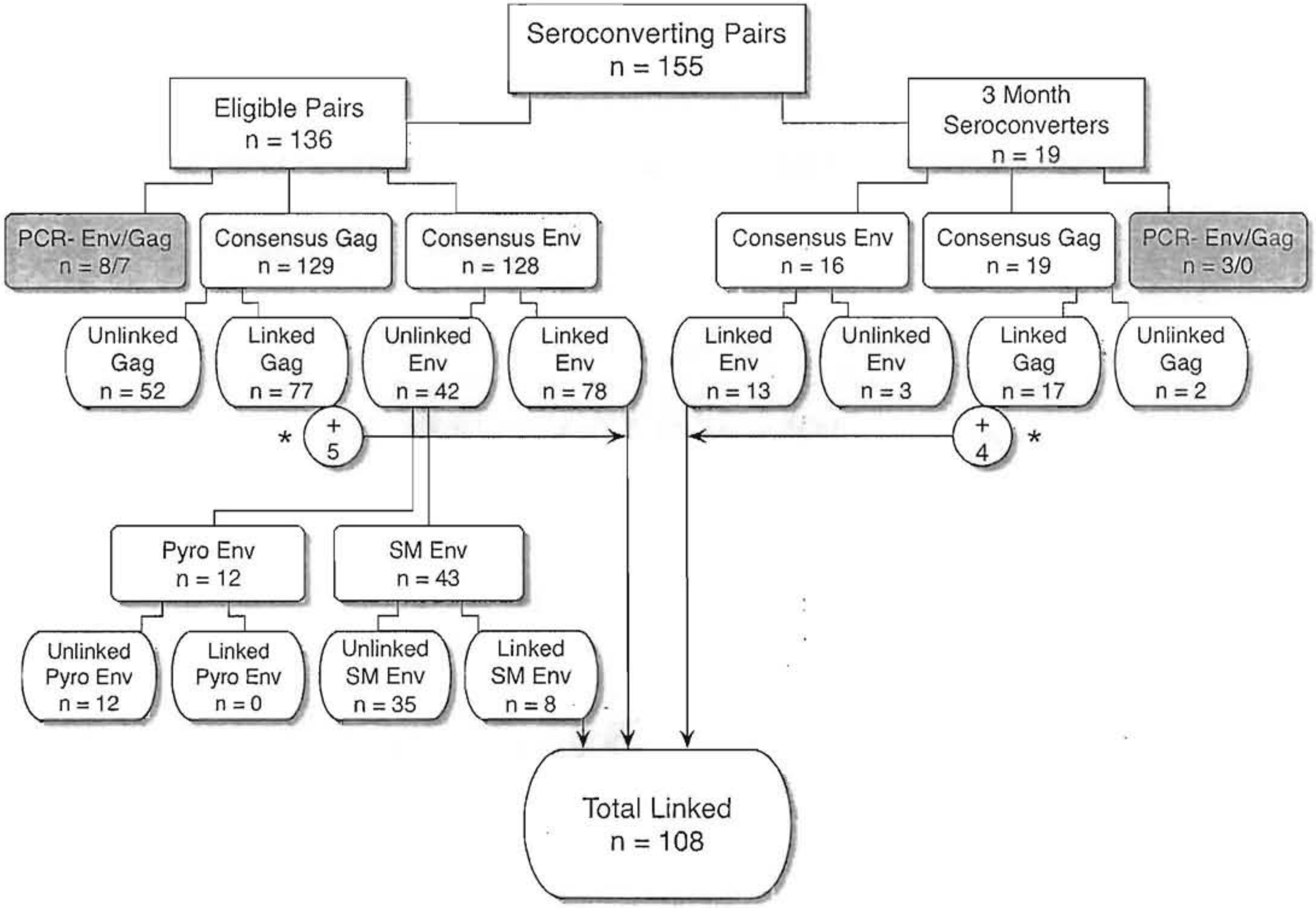


Figure 3, Campbell et al

Linkage Criterion	PP45	PP82	PP73
Monophyly	No	Yes	Yes
Genetic Distance	0.2021	0.0079	0.0203
Bayesian Posterior Probability	0.0000	0.9996	0.9985
Decision	No	Yes	Yes

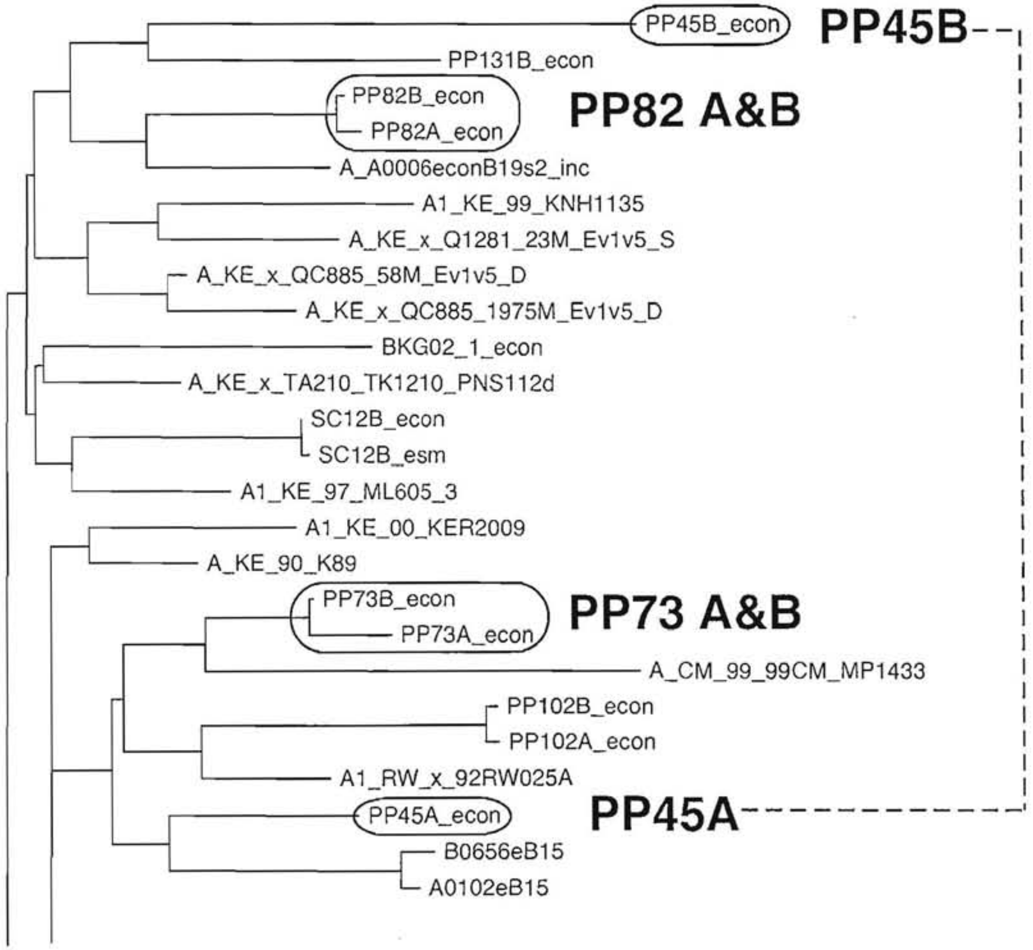


Figure 4A, Campbell et al

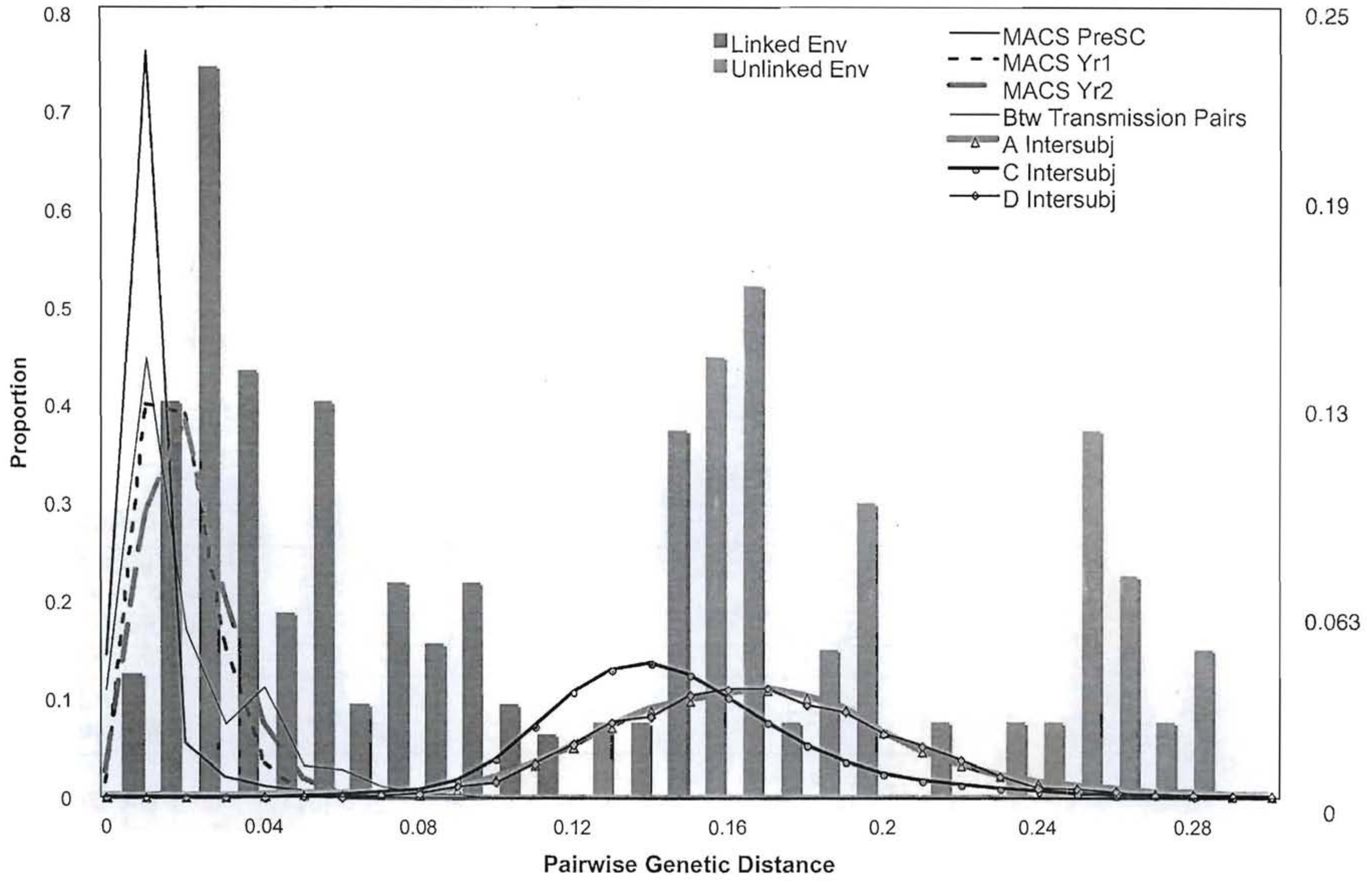


Figure 4B, Campbell et al

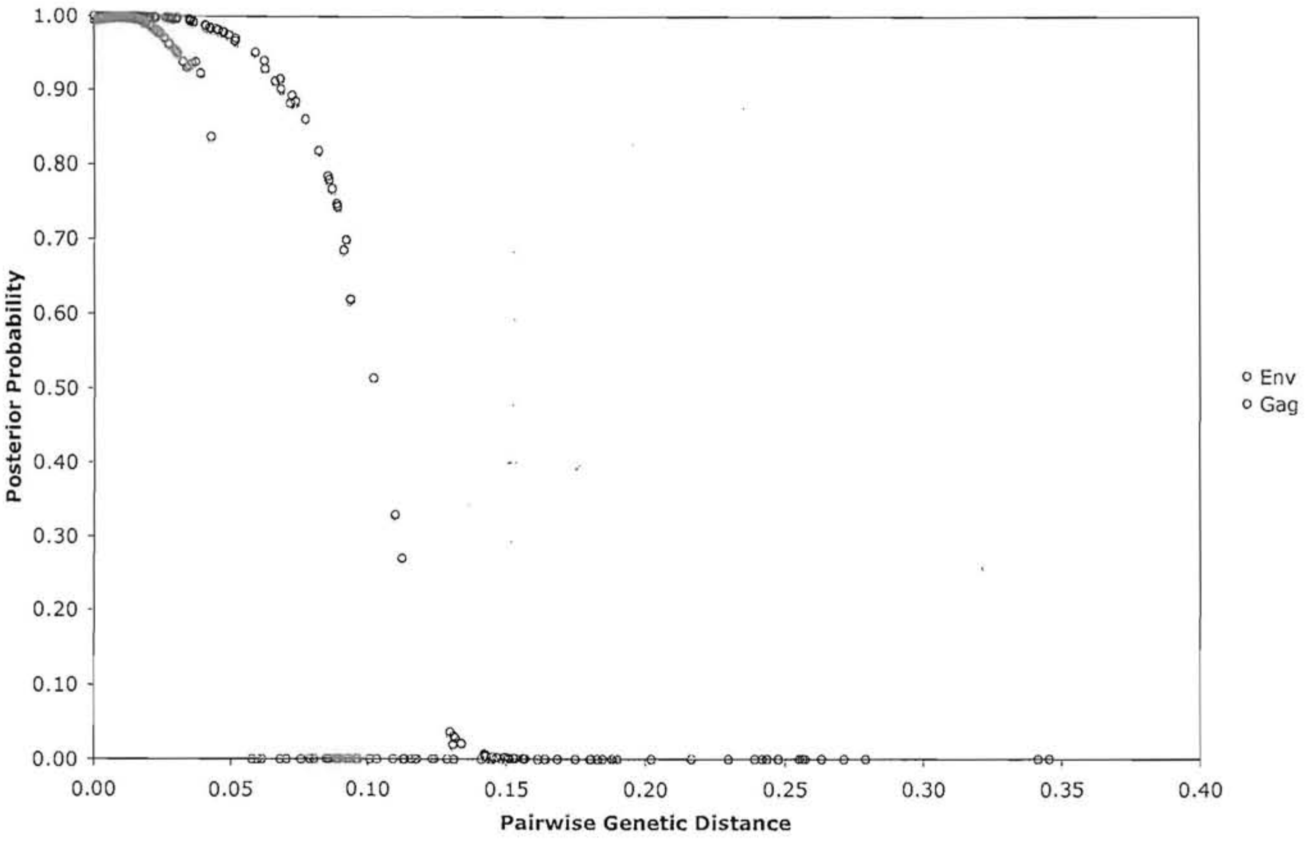


Figure 5, Campbell et al

Linkage Criterion	PP17A variant 1	PP17A variant 2
Monophyly	Yes	Yes
Genetic Distance	0.0159	0.1126
Bayesian Posterior Probability	0.9991	0.0000
Decision	Yes	Indeterminate

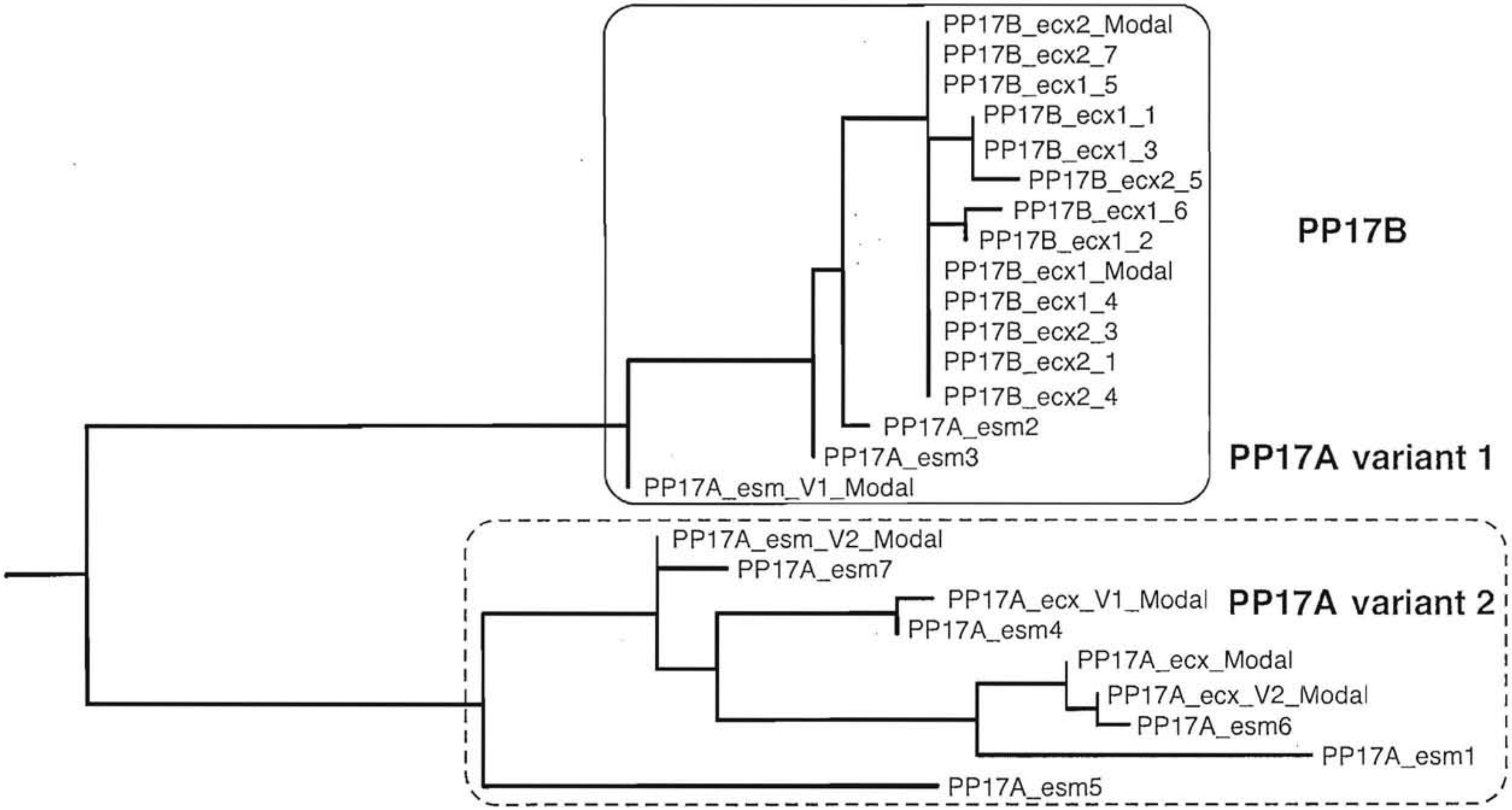


Figure S1, Campbell et al

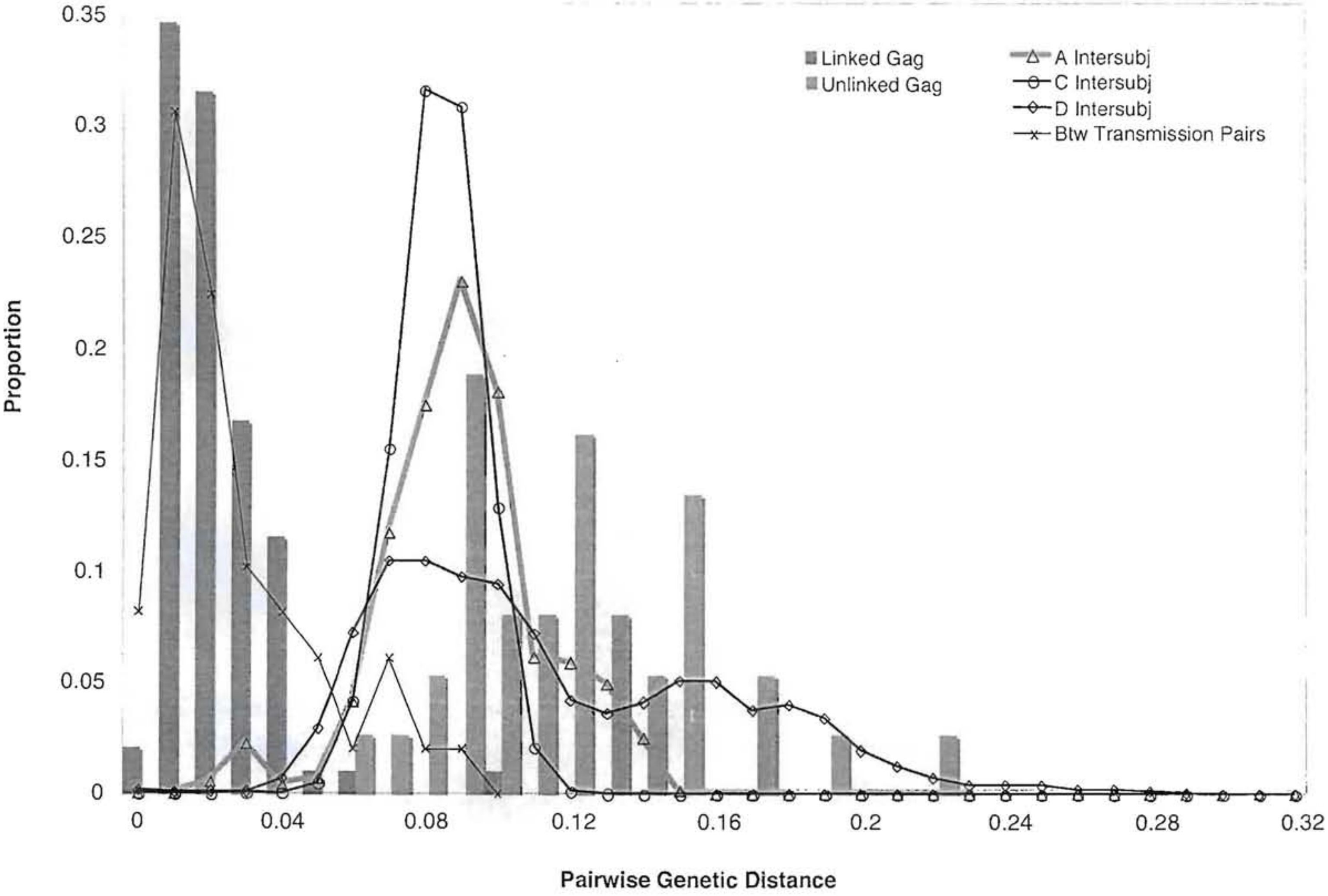


Table S2, Campbell et al

Partner Pair	Approx # templates	Reads	Read avg L	Reads in final alignment (Left side)	Reads in final alignment (Right side)	Total % of reads in final alignments
9	61	4035	218	1727	1531	80.7%
14	61	2729	215	1860	316	79.7%
43	197	588	212	289	88	64.1%
74	59	6571	247	2051	3996	92.0%
79	78+	14769	230	7592	5116	86.0%
81	151	6365	223	3895	1091	78.3%
93	306	8469	238	4744	1325	71.7%
98	87	11423	242	6622	2769	82.2%
110	68	6583	220	2162	1438	54.7%
111	2017	4490	225	1341	2821	92.7%
116	1313	5668	133	495	728	21.6%
117	167	11045	143	1520	1296	25.5%
118	161	3967	70	0	0	0.0%

Figure S2 (page 1), Campbell et al

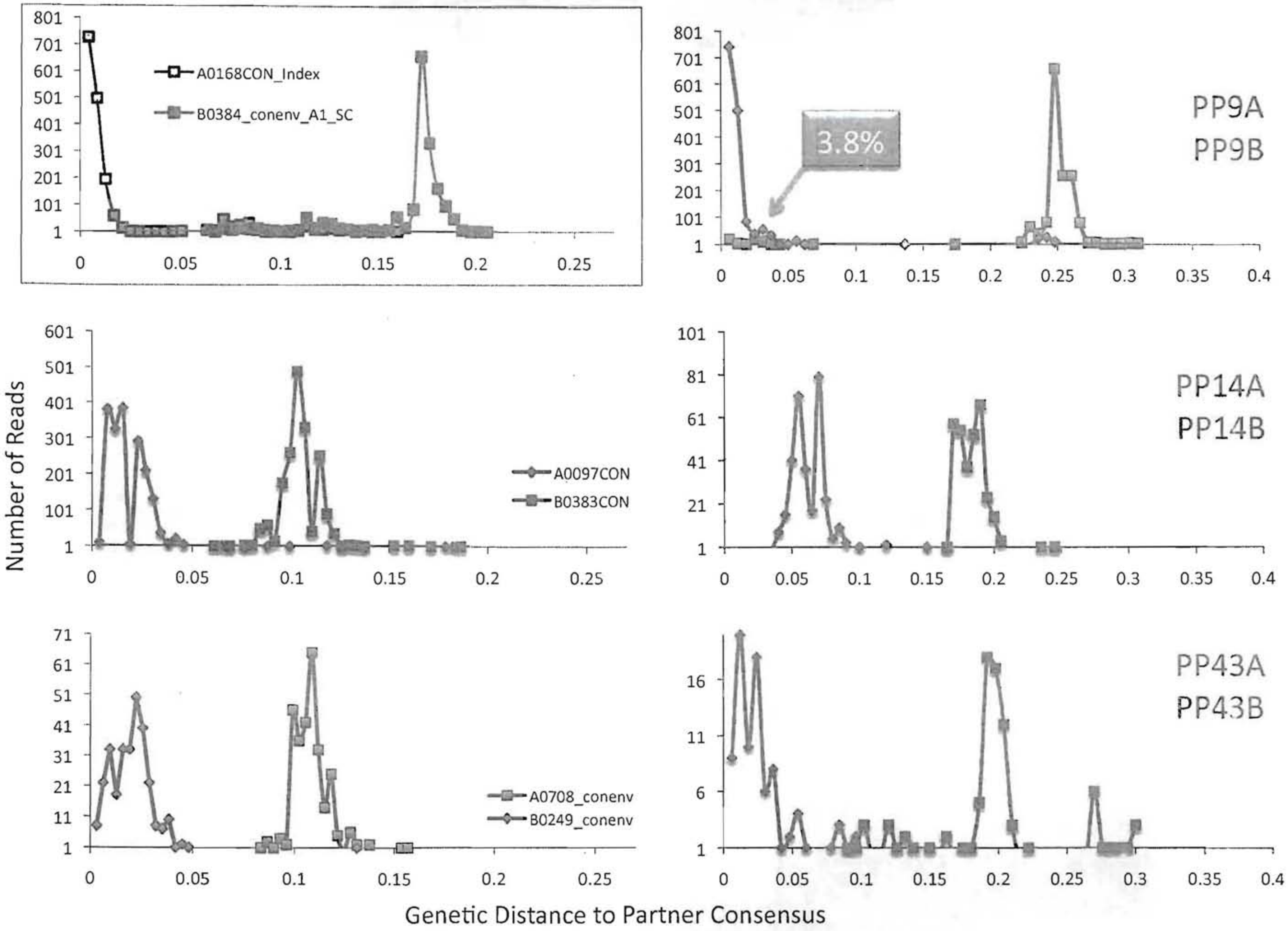


Figure S2 (page 2), Campbell et al

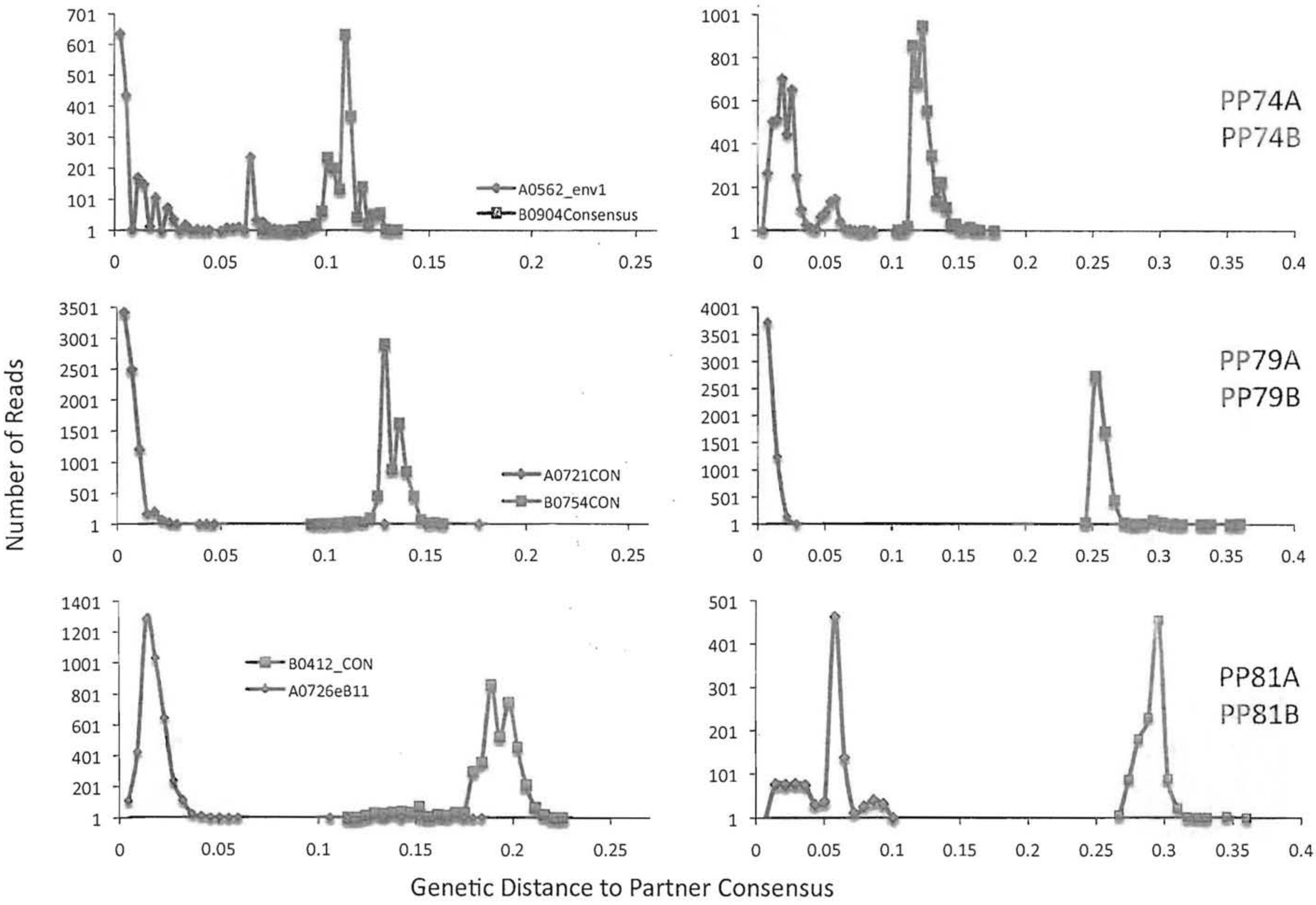


Figure S2 (page 3), Campbell et al

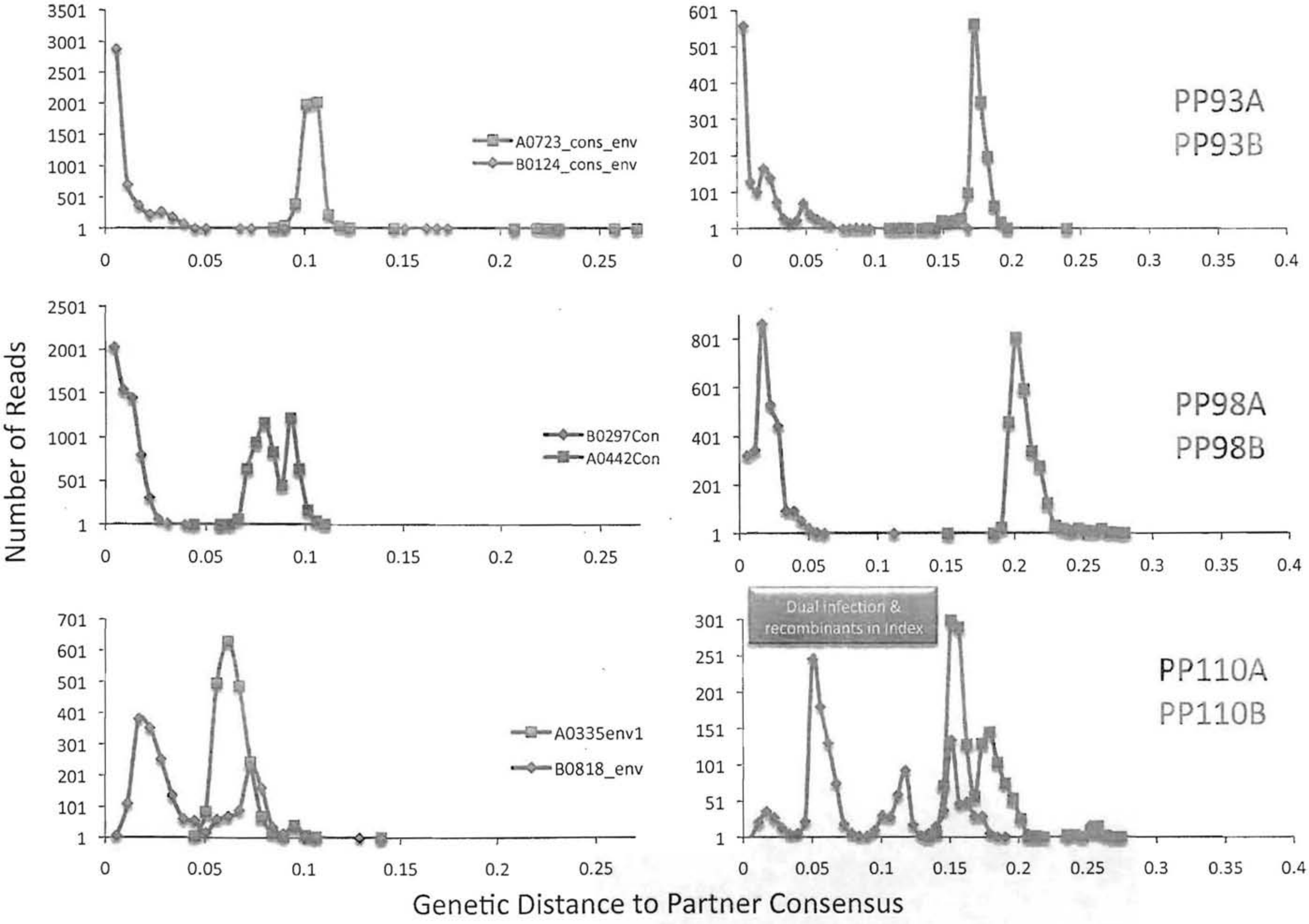
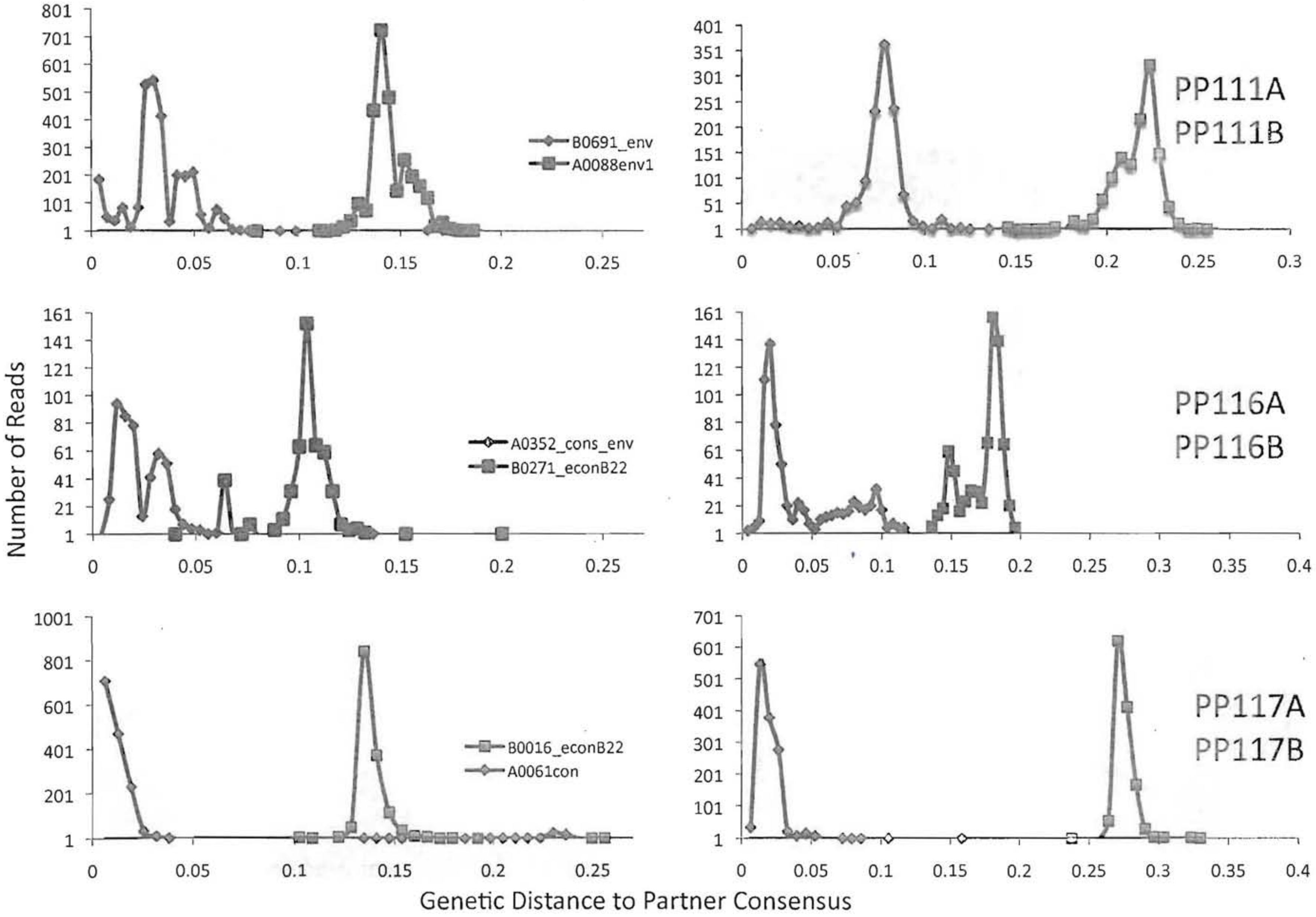


Figure S2 (page 4), Campbell et al


B0691_env
A0088env1
PP111A
PP111B
Number of Reads
A0352_cons_env
B0271_econB22
PP116A
PP116B
B0016_econB22
A0061con
PP117A
PP117B
Genetic Distance to Partner Consensus