Longitudinal Population Analysis of Dual Infection with Recombination in Two Strains of HIV Type 1 Subtype B in an Individual from a Phase 3 HIV Vaccine Efficacy Trial

DAVID V. JOBES,1* MELISSA DAOUST,1 VIVIAN T. NGUYEN,1 ALLAN PADUA,1 FARUK SINANGIL,1 MARCOS PÉREZ-LOSADA,2 KEITH A. CRANDALL,2,3 THEODORE OLIPHANT,3 DAVID POSADA,2,4 ANDREW RAMBAUT,2,5 JONATHAN FUCHS,6 and PHILLIP W. BERMAN7

ABSTRACT

This study documents a case of coinfection (simultaneous infection of an individual with two or more strains) of two HIV-1 subtype B strains in an individual from a Phase 3 HIV-1 vaccine efficacy trial, conducted in North America and the Netherlands. We examined 86 full-length gp120 (env) gene sequences from this individual collected from nine different time points over a 20-month period. We estimated evolutionary relationships using maximum likelihood and Bayesian methods and inferred recombination breakpoints and recombinant sequences using phylogenetic and substitutional methods. These analyses identified two strongly supported monophyletic clades (clades A and B) of 14 and 69 sequences each and a small paraphyletic recombinant clade of three sequences. We then studied the genetic characteristics of these lineages by comparing estimates of genetic diversity generated by mutation and recombination and adaptive selection within a coalescent and maximum likelihood framework. Our results suggest significant differences on the evolutionary dynamics of these strains. We then discuss the implications of these results for vaccine development.

INTRODUCTION

Occurrence of infection with more than one strain of HIV-1 (dual infection) has important implications for understanding HIV transmission and for AIDS vaccine development1,2 and has led to many instances of recombinant viral strains of global epidemiologic relevance.3 Dual HIV-1 infection can be the result of either superinfection (infection with a second strain during some period after the initial established infection) or coinfection (establishment of simultaneous or near simultaneous infection with two or more heterologous HIV-1 strains within a brief period of time). The current literature defines, arbitrarily, HIV coinfection as occurring within the first month of infection.2 Coinfection has been reported in many cases of HIV-1-infected individuals. Some of these cases have been observations of infected individuals with two subtypes (intersubtype)4–10 but others involved HIV-1 variants from the same subtype (intrasubtype).11–14 Less common are the studies where recombination between different coinfecting HIV strains has been documented.11,12 Although recombination is a frequent phenomenon in vivo15–17 and constitutes a significant factor in assisting HIV-1 diversification and escape from host immunity and antiviral therapies, the role that recombination and dual infection play in contributing to virus variation is not well characterized.18 For example, the generation of new strains should be a more common phenomenon in cases of dual infection. Additionally, an overall greater amount of genetic diversity in a coinfected population with recombination and therefore a greater ability to defeat an immune system, drug treatment, or vaccine treatment, potentially resulting in a faster progression to AIDS, should be expected.
Here we report a case of coinfection between two strains of HIV-1 subtype B in a volunteer from the world’s first Phase 3 HIV vaccine efficacy trial (VAX004). Using phylogenetic and population genetic methods, we present congruent and well-supported evidence of recombination between two coinfecting strains of HIV. Additionally, we genetically characterized the evolution of these strains separately and combined using statistical approaches of population parameter inference. Population genetics is indeed important in understanding the evolutionary history, epidemiology, and population dynamics of pathogens, the potential for and mode of evolution of drug and vaccine resistance, and ultimately for public health control strategies and vaccine design.\textsuperscript{19,20} Within this framework, we examine the population dynamics of the coinfecting HIV-1 strains by estimating and comparing their coalescent and likelihood estimates of genetic diversity generated by mutation and recombination, growth, and selection.

**MATERIALS AND METHODS**

**Clinical trial and volunteer characteristics**

VaxGen enrolled 5417 individuals to participate in a Phase 3 vaccine clinical trial (VAX004) testing the efficacy of a recombinant subunit vaccine candidate (AIDSVAX B/B). The vaccine was composed of two HIV-1 gp120 envelope protein antigens derived from the CCR5-dependent GNE8 and CXCR4-dependent MN strains. Trial participants received up to seven injections of vaccine or placebo over a 36-month period and were seen every 6 months for regular HIV testing and counseling. Details of the study design and efficacy results are reported by the rgp120 HIV Vaccine Study Group.\textsuperscript{21}

From these studies, one trial participant was identified as being potentially dually infected with at least two different HIV strains, so an additional molecular study was undertaken to analyze the complete set of longitudinal samples from this person. The clinical progression of HIV-1 viremia and CD4 cell counts within this trial participant, a white male who received three immunizations prior to infection, is shown in Table 1. Screening of sera samples with the Procleix HIV-1/HCV Assay (Chiron) indicated that the first HIV-1-positive timepoint for this volunteer was September 21, 2000 (data not shown). However, no viral load or CD4 count information was available for that collection time. Between November 28, 2000 (T3) and December 18, 2000 (T4) the volunteer started antiretroviral therapy. After that the viral loads were always lower than $7 \times 10^3$ copies/ml. CD4 cell counts were higher than 550 cells/mm$^3$ from November 20, 2000.

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Nb</th>
<th>Collection date</th>
<th>Time point</th>
<th>ART therapy</th>
<th>VL (RNA copies/mL)</th>
<th>CD4 count (cells/mm$^3$)</th>
<th>Sampled tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td>101600P</td>
<td>21</td>
<td>16-Oct-00</td>
<td>T1</td>
<td>—</td>
<td>456,987</td>
<td>277</td>
<td>Plasma</td>
</tr>
<tr>
<td>112000P</td>
<td>20</td>
<td>20-Nov-00</td>
<td>T2</td>
<td>—</td>
<td>801,458</td>
<td>619</td>
<td>Plasma</td>
</tr>
<tr>
<td>112800P</td>
<td>5</td>
<td>28-Nov-00</td>
<td>T3</td>
<td>—</td>
<td>367,984</td>
<td>551</td>
<td>Plasma</td>
</tr>
<tr>
<td>112800C</td>
<td>5</td>
<td>28-Nov-00</td>
<td>T3</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Plasma</td>
</tr>
<tr>
<td>121800P</td>
<td>4</td>
<td>18-Dec-00</td>
<td>T4</td>
<td>Zidovudine and lamivudine/nelfinavir</td>
<td>3,966</td>
<td>585</td>
<td>Plasma</td>
</tr>
<tr>
<td>121800C</td>
<td>3</td>
<td>18-Dec-00</td>
<td>T4</td>
<td>Zidovudine and lamivudine/nelfinavir</td>
<td>—</td>
<td>—</td>
<td>Plasma</td>
</tr>
<tr>
<td>022201M</td>
<td>3</td>
<td>22-Feb-01</td>
<td>T5</td>
<td>Zidovudine and lamivudine/nelfinavir</td>
<td>7,011</td>
<td>638</td>
<td>PBMC</td>
</tr>
<tr>
<td>070201P</td>
<td>4</td>
<td>02-Jul-01</td>
<td>T6</td>
<td>Zidovudine and lamivudine/efavirenz</td>
<td>746</td>
<td>379</td>
<td>Plasma</td>
</tr>
<tr>
<td>070201C</td>
<td>4</td>
<td>02-Jul-01</td>
<td>T6</td>
<td>Zidovudine and lamivudine/efavirenz</td>
<td>—</td>
<td>—</td>
<td>Plasma</td>
</tr>
<tr>
<td>100901P</td>
<td>3</td>
<td>09-Oct-01</td>
<td>T7</td>
<td>Zidovudine and lamivudine/efavirenz</td>
<td>&lt;400</td>
<td>633</td>
<td>Plasma</td>
</tr>
<tr>
<td>100901M</td>
<td>3</td>
<td>09-Oct-01</td>
<td>T7</td>
<td>Zidovudine and lamivudine/efavirenz</td>
<td>—</td>
<td>—</td>
<td>PBMC</td>
</tr>
<tr>
<td>021202P</td>
<td>3</td>
<td>12-Feb-02</td>
<td>T8</td>
<td>Zidovudine and lamivudine/efavirenz</td>
<td>&lt;400</td>
<td>743</td>
<td>Plasma</td>
</tr>
<tr>
<td>021202M</td>
<td>3</td>
<td>12-Feb-02</td>
<td>T8</td>
<td>Zidovudine and lamivudine/efavirenz</td>
<td>—</td>
<td>—</td>
<td>PBMC</td>
</tr>
<tr>
<td>060302P</td>
<td>3</td>
<td>03-Jun-02</td>
<td>T9</td>
<td>Zidovudine and lamivudine/efavirenz</td>
<td>775</td>
<td>615</td>
<td>Plasma</td>
</tr>
<tr>
<td>060302C</td>
<td>2</td>
<td>03-Jun-02</td>
<td>T9</td>
<td>Zidovudine and lamivudine/efavirenz</td>
<td>—</td>
<td>—</td>
<td>Plasma</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Samples are designated by blood collection date and a letter indicating whether the sequence was derived from PPT plasma (P), CPT plasma (C), or PBMC (M), followed by a clone number. For example, 101600P1 is a sequence from October 16, 2000 isolated from PPT plasma and is clone number 1. Infection date for this trial participant was estimated as September 21, 2000. PPT and CPT are different plasma preparation tubes. PBMC, peripheral blood mononuclear cells.

\textsuperscript{b}N, the number of sequences for each time point; the number in parentheses indicates the total number of sequences.
and (T2) to the end, except for July 2001 (T6), when they decreased to 380 cells/mm$^3$.

We present here a comprehensive analysis of phylogenetic relationships, recombination, and population dynamics of HIV strains collected from this volunteer.

**Molecular biology methods**

**HIV-1 gp120 viral RNA preparation and RT-PCR.** Viral RNA and DNA was isolated from 0.5–1.0 ml of frozen plasma or peripheral blood mononuclear cells (PBMCs) using the ViralSeq Sample Preparation Kit (Applied Biosystems, Foster City, CA) and following the manufacturer’s instructions. Full-length gp120 sequences were amplified from samples using reverse-transcription polymerase chain reaction (RT-PCR). Random hexamers were utilized for the RT reaction followed by nested PCR transcription polymerase chain reaction (RT-PCR). Random hexamers were utilized for the RT reaction followed by nested PCR.

Reactions with the first round primers ED3/ED14 (ED3 5$'$mers were utilized for the RT reaction followed by nested PCR transcription polymerase chain reaction (RT-PCR). Random hexamers were amplified from samples using reverse-tran

and following the manufacturer’s instructions. Full-length gp120 sequences leaving a final aligned data set of 86 gp120 sequences.

Phylogenetic analyses

A total of 89 sequences were collected from nine different time points from a single individual from the VAX004 Phase 3 AIDSVAX vaccine trial (Table 1). These sequences were aligned with other sequences from the laboratory (including the vaccine antigens), known laboratory strains, and other common subtypes to guard against contamination, mislabeling, and misidentification. The data set showed no signs of contamination or mislabeling and all gp120 sequences were classified as HIV-1 subtype B (data not shown).

Population dynamics

Identifying recombinants, especially among viruses of the same subtype, can be problematic. Based on both simulation studies$^{33-35}$ and empirical results from HIV-1, we$^{36}$ methods to detect recombinants perform differentially well based on the overall genetic diversity in the data set with the typically used (in HIV-1 studies) phylogenetic methods tending to perform more poorly than substitution pattern approaches. Therefore, we chose to test for recombination using a variety of methods shown to be useful in detecting recombinants in these earlier studies, including MaxChiGlobal,$^{34}$ PhyloPro,$^{37}$ Geneconv$^{37}$, and several recombination detection tools implemented in the software TOvAP$^{39}$ and RDP2.$^{40}$

We estimated genetic diversity ($\theta_W$) and the population recombination rate ($\rho$) using the standard likelihood coalescent approach implemented in the LDhat2.0 package.$^{41}$ Within this framework $\rho$ can be expressed as $2N_e r$, where $N_e$ is again the effective population size and $r$ is the recombination rate per gene per generation. This method relaxes the infinite-sites assumption (typically violated by many empirical data sets $^{36}$) and accommodates different models of molecular evolution. LDhat2.0 also includes a powerful likelihood permutation test (LPT) to test the hypothesis of no recombination ($\rho = 0$).

LDhat implements a corrected version (i.e., based on a finite-sites model) of the algorithm of Watterson$^{42}$ for estimating $\theta_W$. This estimator relies on the number of segregating sites in the sequences and it has been shown that although less efficient than coalescent ML, $\theta_W$ is still remarkably good.$^{43,44}$
Adaptive selection

Adaptive selection at the molecular level is usually assessed by comparing the fixation rates of nonsynonymous (amino acid-altering) and synonymous (silent) substitutions within an ML phylogenetic framework estimating the nonsynonymous/synonymous substitution rate ratio or $\omega = dN/dS$. $\omega$ measures the selective pressure at the protein level, with $\omega = 1$ suggesting neutral evolution, $\omega < 1$ negative selection, and $\omega > 1$ positive selection. We inferred the extent of selection by estimating $\omega$ and the proportion of sites ($p$) with $\omega > 1$ under the models M0 (one ratio), M1 (neutral), M2 (selection), M3 (discrete), M7 (beta), and M8 (beta and $\omega$) of Goldman and Yang and Yang et al. Likelihood scores for the following nested models were compared using likelihood ratio tests (LRT): M0–M3, M1–M2, and M7–M8. We also applied the empirical Bayesian approaches implemented by Nielsen and Yang and Yang et al. to identify the potential sites under diversifying selection as in-

FIG. 1. Maximum likelihood estimate of phylogenetic relationships. A mix model Bayesian analysis generated almost the same relationships. Clades supported by bootstrap values $\geq 70\%$ and/or posterior probabilities $\geq 0.95$ are indicated with an asterisk (*) on the branches. Branch lengths are shown proportional to the amount of evolutionary change. The sequence labels are described in Table 1 and sequence labels are coded by collection date.
dicated by a posterior probability (pP) > 0.95. All these analyses were carried out in PAML3.14 and were performed under initial \( \omega \) values > and <1, as recommended by the author. Here we report the ML estimates. ML trees were estimated for each clade under the best-fit substitution models.

**RESULTS**

*Phylogenetic analyses*

The Akaike Information Criterion (AIC), as implemented in ModelTest3.724 and outlined in Posada and Buckley,25 was used to select the most appropriate model of evolution from 56 different models, compared within an ML framework. The chosen model of evolution was then implemented into PAUP* v4b1026 to estimate an ML tree27 for the full data set. Under the AIC criterion the optimized model of evolution chosen for the entire data set of 86 sequences resulted in the TVM + I + \( \Gamma \) model, a model often found for HIV-1 sequences.50 The ML parameters associated with this model were as follows: base frequencies A = 0.3622, C = 0.1811, G = 0.2233, T = 0.2334, rate parameters of \( R_1 \) = 2.0370, \( R_2 \) = 7.0780, \( R_3 \) = 0.6577, \( R_4 \) = 0.4955, \( R_5 \) = 7.0780, and \( R_6 \) = 1.0, an alpha value for the gamma shape distribution of \( a \) = 0.9108, and a proportion of invariant sites of 0.4104. Using this model, we estimated an ML tree of evolutionary relationships (Fig. 1). Model testing under the AIC criterion for first, second, and third positions of the alignment resulted in the following models: GTR + I + \( \Gamma \), GTR + \( \Gamma \), and TVM + I + \( \Gamma \), respectively. These three models were incorporated under the mixed-model Bayesian estimation. The resulting BMCMC tree was almost identical to that depicted in Fig. 1, but with greater clade support. The resulting phylogeny showed two distinct monophyletic groups of sequences, labeled as clades A and B, and a small group of three sequences clearly distinct connecting them (recombinant A&B). This topology is suggestive of recombination between clades A and B. The long branches separating these clades were supported by high bootstrap values (100%) and high posterior probabilities (1.0). The earliest time point (16OCT00 or 101600 as labeled on the phylogeny) was found in all lineages. Lineage B includes only members of this time point (T1) and the next time point in sampling (20NOV00 or 112000; T2). All subsequent time points are found exclusively in lineage A. Within this clade, longitudinal separation of time points T1–5, T6, and T7–9 can be observed. The recA&B group included the following three sequences: 101600C5, 101600C14, and 101600C12, all from T1.

*Genetic divergence* (GD) between and within strains based on corrected distances under the TVM + I + \( \Gamma \) model is shown in Table 2. Clade A (0.000–0.049) showed greater GD between clones than clade B (0.001–0.013), but at least two times smaller than the GD observed between strains (0.097–0.131). The recA&B showed intermediate values of GD within clones (0.0090.022) and, as reflected in our trees, lower GD with clade B (0.032–0.059) than with clade A (0.051–0.095).

*Detection of recombinants*

As suggested in the initial phylogenetic analyses (Fig. 1), the presence of recombination in the full data set was confirmed by the MaxChiGlobal test (\( p < 0.001 \)). The phylogenetic pro-

![FIG. 2. Phylogenetic profile analysis of the whole alignment.37 Negative correlation values (x-axis) indicate recombination breakpoints. Sequences 060302P83 and 060302P85 were excluded from the analysis because they contained many gaps. Low correlation values at the beginning and at the end of the alignment are noise.](image-url)
file analysis of the whole alignment (Fig. 2) and the distance plots obtained in RDP2 (Fig. 3) indicate that two or three recombination events seemed to have occurred at the beginning of the infection, giving rise to the recA&B group, with breakpoints around nucleotide positions 830, 966, and 1224, respectively. We used breakpoints 830 and 1224 to divide the whole alignment into three subregions and ML trees were constructed for each of them separately under the best-fit models (Fig. 4). Clear clustering by phylogenetic analysis of subregions of gp120 sequences highlights the regional contribution of the two parent populations to the construction of the new recombinant gp120 molecules (Fig. 4).

**FIG. 3.** Distance plots obtained in RDP2 for the three putative recombinant sequences against two representatives of the two subtype B lineages.
**Evolutionary dynamics of intrasubtype B strains**

We characterized the population dynamics of this volunteer's infecting virus with estimates of genetic diversity generated by mutation ($\theta_W$) and recombination ($\rho$) for each strain and time point(s) (Table 3). $\theta_W$ and $\rho$ were higher for clade A ($\theta_W = 0.039, \rho = 81\%$) compared to rates estimated for clade B ($\theta_W = 0.011, \rho = 2\%$). Sample size differences between strains are a concern when interpreting these estimates. These sample differences by themselves illustrate different evolutionary dynamics: clade A was present in all time points sampled (20-month period), but clade B did not appear in plasma or blood after the second time point (2 months). Nevertheless, genetic differences between clades A and B were clear since the first time point, as indicated by the $\theta_W$ and $\rho$ estimates: $\theta_W$ and $\rho$ were considerably higher for clade A at T1 and T2 or T1–2 (0.021–0.028 and 20–67, respectively) than they were for clade B (0.004–0.011 and 0–2, respectively) at the same time points (Table 3). Within clade A, $\theta_W$ diminishes after T4 probably as a result of a dramatic reduction in viral load (i.e., effective population size) due to the beginning of ART therapy application. We tested for positive selection by obtaining likelihood values.
Bayesian approach appears to be less affected. \(^5^1\) However, mistakes recombination as evidence of positive selection. Iden-

M1–M2 but much less the LRT of models M7–M8, and often matically the accuracy of the LRT test of models M0–M3 and

0.01), as those reported here for clade A, seem to affect dra-

factions would be expected to be even more rarely reported. In

under positive selection (\(pP\)). Previous work has also demonstrated convincing evidence for both adaptive selection and recombination.\(^4^5,^5^2,^5^3\)

and parameter estimates for different codon-substitution mod-

els of selection (Table 4). Model comparisons M0–M3, M1–M2, and M7–M8 for lineage B were not significant as in-

dicated by the LRTs, hence no evidence of positive selection was observed. Consistent and significant strong levels of selection (\(\omega > 4\)), however, for more than 5% of the sites were detected under site-specific models for lineage A. Under these models the Bayesian approaches were able to detect 8–16 sites under positive selection \((P > 0.95)\). Simulations by Anisi-

move et al.\(^5^1\) indicate that high levels of recombination (\(C = 0.01\)), as those reported here for clade A, seem to affect dra-

matically the accuracy of the LRT test of models M0–M3 and M1–M2 but much less the LRT of models M7–M8, and often mistakes recombination as evidence of positive selection. Identifi-

cation of sites under positive selection, however, using the Bayesian approach appears to be less affected.\(^5^1\) However, when data were simulated at very high levels (\(\omega = 6\)) of posi-

tive selection (similar to those observed in our study) and the population was expanding, the LRT and Bayesian test become more accurate and powerful (concrete values are not reported). Therefore, although we think that recombination is probably in-

flating our PAML estimates of positive selection, the observed values are so high that it is clear that natural selection is a sig-

nificant driving force in the evolution of these sequences as well. Previous work has also demonstrated convincing evidence for both adaptive selection and recombination.\(^4^5,^5^2,^5^3\)

### DISCUSSION

Reports in the literature of dual infection are rare; for in-

stance, superinfection has been documented in only 16 cases so far and most of those involved infection with different HIV sub-

types.\(^2\) Due to the difficulty of detection, intrasubtype dual in-

fections would be expected to be even more rarely reported. In

the absence of substantive clinical indications of a dual infec-

tion (e.g., increases in viral load or accelerated disease pro-

gression while on drug therapy), the ability to detect multiple infection events is generally limited to molecular or serologi-

cal analyses. However, even using molecular techniques, the detection of dual infection and recombination can be problem-

atic for several reasons, especially if infection occurs with HIV isolates of the same subtype. For example, inadequate or im-

proper use of phylogenetic and recombination detection meth-

ods, insufficient longitudinal sampling, and the use of partial gene sequences instead of complete genes or genomes make it likely that estimates of dual infection and recombination are un-
derestimated in many at-risk populations.

In this study we reported on an individual from a Phase 3 HIV vaccine efficacy trial who was infected with multiple HIV strains and a recombinant population, despite receiving repeated vaccinations with a recombinant gp120-based vaccine. This study clearly demonstrates the importance of sampling close to the infection event in order to detect dual infection. Indeed, if the initial sampling took place after the second time point, clade B and the recA&B clade would never have been sampled.

In the current study, we also thoroughly document recombi-

nation using multiple phylogenetic and pattern matching meth-

ods. We expect that the recA&B clade is indeed a true recom-

binant group (i.e., and not generated by PCR) for the following reasons: (1) the recombinant breakpoints, while closely associ-

ated by region, are distinct, suggesting that more than one re-

combination event generated these new mutants and (2) despite at least three breakpoints at the approximate positions 830, 966, and 1224, all three gp120 sequences recombined to form an uninterrupted open reading frame. For this study, in fact, we only used sequences with full open reading frames (i.e., presumably functional molecules), to reduce the risk of analyzing non-

functional sequences.

While phylogenetic methods are not always reliable predictors of recombination, in this case, both phylogenetic and pattern matching methods substantiated the recombinants and clearly dif-

ferentiated the three populations (Figs. 1 and 4). While it is widely known that recombination occurs frequently in vitro,\(^1^7,^1^8,^5^4\) it is unclear how common recombination is between viruses from dif-

ferent infection events and in various geographic regions. For ex-

ample, in Africa circulating recombinant forms are widespread and can make up a large percentage of the virus population in many countries,\(^5^5-^5^7\) whereas recombination in North American viruses is seldom observed. Indeed this is partly a consequence of the fact that recombination between subtypes is much more easily detected than within subtypes and that there are far fewer subtypes circulating in North America (predominantly subtype B). This study thus complements prior studies reporting evidence of recombination between different strains of a single HIV-1 sub-

type,\(^1^1,^1^2\) which taken together indicate that recombination of vi-

ral genomes within the different strains present in an infected in-

dividual may constitute a significant factor in the generation of HIV-1 diversity.\(^1^1\)

Differences in sample size per time point (Table 1) suggest im-

portant differences in population dynamics between invasive strains since we see no representation of lineage B or the recom-

binant recA&B after time point two (before treatment with ART therapy). Population dynamics can be efficiently studied using explicit statistical models of evolution, such as the coalescent\(^5^8\) and

![Table 3. Estimates of genetic diversity (\(\theta_w\)) and population recombination rate (\(\rho\))](image-url)

<table>
<thead>
<tr>
<th>Strain</th>
<th>(\theta_w)</th>
<th>(\rho)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strain A_T1_9</td>
<td>0.039</td>
<td>81***</td>
</tr>
<tr>
<td>Strain AT1</td>
<td>0.021</td>
<td>20***</td>
</tr>
<tr>
<td>Strain AT2</td>
<td>0.025</td>
<td>58***</td>
</tr>
<tr>
<td>Strain AT3</td>
<td>0.021</td>
<td>75***</td>
</tr>
<tr>
<td>Strain AT4</td>
<td>0.019</td>
<td>3**</td>
</tr>
<tr>
<td>Strain AT5</td>
<td>0.009</td>
<td>0</td>
</tr>
<tr>
<td>Strain AT6</td>
<td>0.008</td>
<td>0</td>
</tr>
<tr>
<td>Strain AT7</td>
<td>0.003</td>
<td>0</td>
</tr>
<tr>
<td>Strain AT8</td>
<td>0.007</td>
<td>0</td>
</tr>
<tr>
<td>Strain AT9</td>
<td>0.009</td>
<td>0</td>
</tr>
<tr>
<td>Strain AT1_2</td>
<td>0.028</td>
<td>67***</td>
</tr>
<tr>
<td>Strain AT1_2</td>
<td>0.011</td>
<td>2**</td>
</tr>
<tr>
<td>Strain BT1</td>
<td>0.009</td>
<td>1**</td>
</tr>
<tr>
<td>Strain BT2</td>
<td>0.004</td>
<td>0</td>
</tr>
</tbody>
</table>

\(\theta_w\) and parameter estimates for different codon-substitution mod-

els of selection (Table 4). Model comparisons M0–M3, M1–M2, and M7–M8 for lineage B were not significant as in-

dicated by the LRTs, hence no evidence of positive selection was observed. Consistent and significant strong levels of selection (\(\omega > 4\)), however, for more than 5% of the sites were detected under site-specific models for lineage A. Under these models the Bayesian approaches were able to detect 8–16 sites under positive selection \((P > 0.95)\). Simulations by Anisi-

move et al.\(^5^1\) indicate that high levels of recombination (\(C = 0.01\)), as those reported here for clade A, seem to affect dra-

matically the accuracy of the LRT test of models M0–M3 and M1–M2 but much less the LRT of models M7–M8, and often mistakes recombination as evidence of positive selection. Identifi-

cation of sites under positive selection, however, using the Bayesian approach appears to be less affected.\(^5^1\) However, when data were simulated at very high levels (\(\omega = 6\)) of posi-

tive selection (similar to those observed in our study) and the population was expanding, the LRT and Bayesian test become more accurate and powerful (concrete values are not reported). Therefore, although we think that recombination is probably in-

flating our PAML estimates of positive selection, the observed values are so high that it is clear that natural selection is a sig-

nificant driving force in the evolution of these sequences as well. Previous work has also demonstrated convincing evidence for both adaptive selection and recombination.\(^4^5,^5^2,^5^3\)
Table 4. Likelihood Values, \( \omega \) Estimates (\( \omega_{M2}, \omega_{M3}, \) and \( \omega_{M8} \)), Proportion of Sites with \( \omega > 1 \) (\( p_{M2}, p_{M3}, \) and \( p_{M8} \)) and Number of Positively Selected Sites with a Posterior Probability > 0.95 (\( n_{M2}, n_{M3}, \) and \( n_{M8} \))^a

|          | \(-ln\) | \( L_{M0} \) | \(-ln\) | \( L_{M1} \) | \(-ln\) | \( L_{M2} \) | \( \omega_{M2} \) | \( p_{M2} \) | \( n_{M2} \) | \(-ln\) | \( L_{M3} \) | \( \omega_{M3} \) | \( p_{M3} \) | \( n_{M3} \) | \(-ln\) | \( L_{M7} \) | \(-ln\) | \( L_{M8} \) | \( \omega_{M8} \) | \( p_{M8} \) | \( n_{M8} \) |
|----------|---------|-------|-------|-------|-------|-------|--------|-------|-------|-------|-------|--------|-------|-------|-------|-------|-------|-------|--------|-------|-------|-------|-------|-------|
| Clade A  | 60      | 10.4  | 59    | 53.3  | 59    | 11.0  | 5.56   | 0.053 | 8     | 59    | 13.6  | 4.46   | 0.082 | 16     | 59    | 59.2  | 11.2  | 5.08  | 0.064 | 9     |

^aThe comparisons between models M0 and M3, M1 and M2, and M7 and M8 were only significant (\( p < 0.001 \)) for clade A. \( \omega \) and \( p \) estimates under purifying selection (\( \omega < 1 \)) or neutral selection (\( \omega = 0 \)) are not reported for any model.
DUAL INFECTION IN A PHASE 3 VACCINE TRIAL

the ML\textsuperscript{27} approaches that describe their effect on gene sequences. Here we used different implementations of these approaches to compare key population parameters such as $\theta$, $\rho$, and $\omega$ between two HIV-1 strains. Observed significant differences in these parameters confirm sampling observations. Clade B is much more homogeneous than clade A, as indicated by their lower values of genetic diversity ($\theta_{00}$) and recombination ($\rho$) (Table 3) and genetic divergence (Table 2). We found no evidence of adaptive selection in clade B (Table 4); hence, based on the gpl200 gene region analyzed, it can be postulated that clade B viruses were not able to escape the immune system and ultimately failed to establish. Nevertheless, other factors such as different growth kinetics or stochastic loss due to severe viral drop (only for the later time points) cannot be ruled out to explain the failure of this strain. After the second month of infection, clade A began to dominate the viral population and eventually took over completely (Table 1). High levels of selection pressure on this clade were observed (Table 4), as expected for populations subjected to significant immunological pressure. Finally, the phylogeny itself (Fig. 1) also reflects this difference in growth rate among the strains and also the relative increase in fitness due to adaptation experienced by clade A as evidenced by the different clustering of the samples: early samples (T1–5) are depicted at the bottom of clade A and late samples (T7–9) at shallower nodes, with sample T6 in an intermediate position.

In summary, this study reports a case of coinfecion and recombination from an HIV vaccine efficacy trial. As part of this presumably single infection event, this trial participant harbors two major virus populations and a small recombinant group, all of subtype B. Under the same host and treatment conditions it has been demonstrated that these three virus populations responded with different evolutionary dynamics over a period of about 2 years probably because of their different genetic backgrounds. This study underscores the need for a much larger survey to document the actual prevalence of multiple infection and recombination events in both vaccine and placebo groups from trial populations to better understand the relative population dynamics of HIV evolution in high-risk populations and in response to vaccine treatment.

ACKNOWLEDGMENTS

The authors are indebted to the thousands of clinical trial volunteers who participated in the first Phase 3 HIV vaccine efficacy trial and who, as a result, have contributed a great deal to our current understanding of HIV. K.C. and D.P. were supported in part by NIH Grant GM 12 13 66276.

SEQUENCE DATA

All DNA sequences are deposited in GenBank under accession numbers P2924864–DQ924949.

REFERENCES


Address reprint requests to:
David Jobes
Presidio Pharmaceuticals, Inc.
One Post Street, Suite 875
San Francisco, California 94104

E-mail: david@presidiopharma.com