Selection in Context: Patterns of Natural Selection in the Glycoprotein 120 Region of Human Immunodeficiency Virus 1 Within Infected Individuals

**Alan R. Templeton,*,1 Rebecca A. Reichert,*,2 Anton E. Weisstein,*,3 Xiao-Fang Yu†
and Richard B. Markham†**

*Department of Biology, Washington University, St. Louis, Missouri 63130-4899 and †Department of Molecular Microbiology
and Immunology, Johns Hopkins University School of Hygiene and Public Health, Baltimore, Maryland 21205

ABSTRACT

Evolution of the HIV-1 V3 loop was monitored in 15 subjects over a period of 5 years at ∼6-month intervals. Putative recombination was detected in many of the sequences. Evolutionary trees were estimated from the nonrecombinant viral sequences found in each individual. Selection and altered demographic regimes were detected with logit and other contingency analyses in a highly context-dependent fashion. Mutations leading to amino acid substitutions are subject to positive selection over a broad range of clinical conditions in the nonsyncytium-inducing (NSI) form, and the growth rates of the NSI strains and their level of genetic subdivision change little in going from a healthy immune system to a severely compromised immune system. In contrast, the SI form has a significant increase in growth rate as the immune system goes from healthy to compromised, particularly in those subjects who did not receive any antiviral drug therapy. This increase in SI growth rate results in a significant growth advantage of SI over NSI when the immune system is compromised. The SI strains also show more demographic subdivision when the immune system is healthy than when the immune system is compromised, and the SI form has greater demographic subdivision than NSI in subjects with healthy immune systems who also are not receiving antiviral drug therapy. Positive selection on amino-acid-changing mutations weakens and then intensifies again in the SI strains in going from healthy to compromised immune systems. These patterns are consistent with other studies that suggest that NSI strains inhibit replication of SI strains, that the V3 loop is more hidden from the immune system in the NSI form, that evolution in the V3 loop influences cell tropism and coreceptor usage, that substrate for replication of SI forms increases as the disease progresses, and that death of CD8 cells is influenced by the type of coreceptor usage typically found in SI but not in NSI strains. Finally, the transition between NSI and SI forms is associated with a burst of evolutionary change due to strong positive selection at sites other than those that define the NSI/SI phenotypes.

THERE are three stages to human immunodeficiency virus (HIV) infection (Leigh Brown and Holmes 1994; Coffin 1999). The first stage, termed seroconversion, lasts for 7–8 weeks and is characterized by high viral titers within the body, the initiation of a host immune reaction to the virus, and conversion from testing negative to testing positive for antibodies to HIV. The second stage, termed latency, can last for a variable period of years and is characterized by relatively constant viral titers within the body and by decreases in host CD4 T-cell counts in most individuals. The third stage is termed acquired immune deficiency syndrome (AIDS) and is characterized by a final collapse of the host's immune system, high HIV viral titers, and a series of opportunistic infections due to the host's severely suppressed immune system; these infections often become lethal. Although these three stages sequentially occur within any individual who dies of AIDS-related illnesses, the time it takes a subject to progress through the last two stages is highly variable. Some individuals have a rapidly progressing infection and can move through all three stages in <2 years. Others remain in the latency period for over a decade without CD4 counts significantly decreasing and/or any indication of AIDS-related illnesses.

Despite its name, the latency period is highly dynamic for both the host and the virus, involving continuous rapid HIV-1 turnover, mutation, and evolution (Leigh Brown and Holmes 1994; Bonhoeffer et al. 1995; Wei et al. 1995; Coffin 1999) along with a constant interaction between the host's immune system and HIV-1 (Nowak et al. 1995; Goudsmit 1997). The potential for evolution during the latency period is high because the

viral reverse transcriptase has a large error rate (LEIGH BROWN and HOLMES 1994; COFFIN 1999), HIV-1 has a short generation time of 1.2–2.6 days, and population sizes are 10^9–10^{10} viruses within the body at any one time in patients not treated with HIV antiretroviral medications (PERELSON *et al.* 1996). Hence, during the latency period, the HIV-1 population within one individual can come to express high levels of variation and evolutionary divergence (COFFIN 1995).

This study focuses on the glycoprotein coat that affects HIV-1's interaction with the host's immune system (GOUDSMIT 1997) and the target of host cell-type preference (SHIINO *et al.* 2000). The third variable domain on gp120, termed the V3 loop, contains an antigenic site to which HIV-1 antibodies respond (SEIBERT *et al.* 1995; GOUDSMIT 1997). The V3 domain also specifies the coreceptor usage of HIV-1, which in turn determines target cell preference or cell tropism (DITTMAR *et al.* 1997). Specific amino acid changes in the V3 loop region alter coreceptor usage patterns and are therefore associated with a major phenotype switch in HIV-1: the transition between nonsyncytium-inducing (NSI) to syncytium-inducing (SI) forms that cause infected cells to form large multinucleate bodies (syncytia). The NSI phenotype, associated with viral use of the CCR5 coreceptor, dominates immediately after primary infection and during the latent phase. Although SI mutants, which employ the CXCR4 coreceptor for viral entry, can arise from NSI forms by just one or two amino acid changes in the V3 loop (DITTMAR *et al.* 1997), SI strains generally do not emerge as a dominant form until late in infection and are correlated with worsened prognosis and accelerated disease progression. The fact that the SI forms can evolve from the NSI form by only one or two mutations yet do not dominate until late in infection implies that there is some sort of selective force modulating this phenotypic switch (CALLAWAY *et al.* 1999; SHIINO *et al.* 2000). Hence, the V3 loop is a candidate region for selective effects exerted by the host's immune system, cell tropism, and the transition from NSI to SI forms.

In light of the many important functions coded for by the V3 loop of the *envelope* (*env*) gp120 gene, it is not surprising that several studies have found evidence for selection in this gene region, although the type and pattern of inferred selection has varied among studies (BONHOEFFER and NOWAK 1994; KELLY 1994; NOWAK *et al.* 1995; SEIBERT *et al.* 1995; ENDO *et al.* 1996; WOLINSKY *et al.* 1996; GANESHAN *et al.* 1997; LIU *et al.* 1997; McDONALD *et al.* 1997; YAMAGUCHI and GOJOBORI 1997; MARKHAM *et al.* 1998; SHANKARAPPA *et al.* 1998). Given the diversity of selective forces that can potentially operate on this gene, this heterogeneity among studies is not surprising because many studies involved only a small cohort of infected subjects, the subjects chosen did not reflect the range of disease progressions observed in the overall HIV positive population, and/or a limited

number of time points were analyzed for each subject (MARKHAM *et al.* 1998). Indeed, this selective heterogeneity was documented in our previous study on one of the larger cohorts to be studied: 15 subjects, followed from seroconversion at ∼6-month intervals, who display a broad range of overall disease progression trajectories (MARKHAM *et al.* 1998). MARKHAM *et al.* (1998) used analyses of synonymous and nonsynonymous differences and rates of divergence to reveal selective heterogeneity as a function of disease progression category.

The purpose of this article is to examine selection in the same 15 subjects as in the MARKHAM *et al.* (1998) study but to investigate a broader range of potential selective contexts. The contexts that we examine include overall disease progression categories; the CD4 and CD8 T-cell counts observed at each visit; evolution within NSI, SI, and transitional viral forms; and interactions among these factors.

SUBJECTS AND METHODS

The study population: The 15 HIV-1$^+$ subjects were selected from a cohort of injection drug users participating in the AIDS Linked to Intravenous Experience (ALIVE) study in Baltimore. This ongoing cohort study follows infected or at-risk injection drug users at 6-month intervals (visits), at which times blood was obtained for virologic and immunologic studies. All 15 subjects are thought to have contracted HIV from IV drug use after enrollment in the study. (The only possible exception involves two subjects who were sexually involved. While both used IV drugs and hence could have contracted HIV from IV drug use, there is the possibility that one contracted HIV via drug use and then sexually transmitted it to the other.) The earliest samples used in this study were taken in 1989 and the last in 1993. The subset of individuals selected for this study was followed from the point of HIV-1 seroconversion and reflects different trajectories of CD4 T-cell counts. Rapid progressors were defined as having attained a level of <200 CD4 T cells within 2 years of seroconversion; moderate progressors had CD4 T-cell levels decline to 200–650 during the period of observation, and nonprogressors maintained CD4 T-cell levels >650 throughout the observation period.

Only 5 of the 15 subjects received as monotherapy the weak antiviral drug AZT, which targets the product of the *pol* gene and not *env*, the focus of this study. The compliance of these 5 subjects in taking their medication is not known, although compliance in injection drug users is often poor (CARRIERI *et al.* 2003; PALEPU *et al.* 2003; WOOD *et al.* 2003). With the exception of 1 subject and 1 visit for another, AZT was not prescribed until the CD4 counts were <500. Given these facts and the observation that resistance emerges rapidly to single-drug therapy (the norm for the time period of these studies), antiretroviral drug therapy should not have

had much impact in this group on the selective forces on the *env* gene. However, these drugs might have had an impact on viral demographic patterns, so our analyses of viral demography will be done both with and without stratification on whether or not the subject received antiviral drugs.

Sequencing of HIV-1 *env* genes: Nested PCR was used to amplify a 285-bp region of the *envelope* gene, including the V3 region of gp120 from peripheral blood mononuclear cells (PBMC). Full details are given in MARKHAM *et al.* (1998). A random sample of genes was sequenced from each subject visit, yielding a database of 6–21 gp120 genomic sequences for 80 subject visits, with individual subjects being sampled over periods ranging from 18 months to 5 years. The sequences analyzed for this study may be obtained through GenBank (accession nos. AF01670–AF016825 and AF089109–AF089708).

Determination of CD4 and CD8 cell counts and the presence of SI forms: CD4 and CD8 T-cell counts were made on each subject at each visit. These counts were converted into three categories for the statistical analysis. The CD4 T-cell brackets used were (1) $\geq$650 cells/μl, (2) between 200 and 650 cells/μl, and (3) <200 cells/μl. These categories were consistent with those selected for previous analyses of this cohort (MARKHAM *et al.* 1998). The CD8 T-cell brackets used were (1) $\geq$1050 cells/μl, (2) between 650 and 1050 cells/μl, and (3) <650 cells/μl. These categories were chosen because they result in roughly equal numbers of observed mutations in all three categories.

SI haplotypes were identified as those haplotypes coding for an arginine or a lysine at amino acid position 306 in gp120 or a glycine at position 306 coupled with an arginine or lysine at position 320 (DE JONG *et al.* 1992). The viral population sampled at each visit was tested directly for the presence of SI by examination of the ability of virus to grow on MT-2 cells, which will support only the growth of SI viruses. Cell-free viruses recovered from infected PBMC were passed twice through MT-2 cells to enrich for any SI variants that might be present. Virus replication was monitored by the presence of p24 in the culture supernatants. Syncytium formation was examined daily by use of a light microscope. In all cases, the inferences made from the amino acid sequence were concordant with the experimental determination of the presence of SI forms.

Estimation of phylogenetic trees, recombinants, and mutational counts: Trees were estimated separately for the sequences obtained from each of the 15 subjects using the procedure given in TEMPLETON *et al.* (1992) that is now commonly called statistical parsimony (SP; CRANDALL 1994; CRANDALL and TEMPLETON 1996) as implemented with the computer program TCS (available at http://darwin.uvigo.es/). Statistical parsimony utilizes the fact that multiple mutational hits at the same nucleotide are more likely to have occurred between

haplotype pairs that differ at many other nucleotide sites than to haplotype pairs that differ at only one or a few nucleotide sites. We estimated a SP haplotype tree for each of the 15 individuals. In some cases, there were multiple solutions, all of which were statistically parsimonious. Hence, we sometimes had an SP set rather than a single, unique tree.

The initial SP haplotype tree or trees for each subject is estimated under the assumption of no recombination/gene conversion. This initial tree was used to test for the presence of recombination/gene conversion (CRANDALL and TEMPLETON 1999; TEMPLETON *et al.* 2000). This test is based on the property that programs for estimating haplotype trees will generate homoplasies (multiple events at the same site) when recombination or gene conversion occurs. Such homoplasies could also be due to multiple mutational hits, but in that case the homoplasies should be scattered over the haplotype tree and scattered over the sequenced molecule. When homoplasies are created by such tree-building programs as an artifact of recombination or gene conversion, physically close homoplasies are placed on the same branch of the tree. A runs test is then performed to see if the homoplasies are significantly clustered in the tree as an indicator of recombination and/or gene conversion, as detailed in TEMPLETON *et al.* (2000). Whenever statistically significant recombinants were identified, they and all their descendants were removed from the initial SP tree. Because the original recombinant haplotype can continue to evolve by mutational accumulation and diversification, branches removed from the SP tree but generated after recombination are classified in the standard manner, with the original recombinant being regarded as the root of the postrecombination tree. For simplicity, we still refer to a "tree" being estimated for each subject, but some of these "trees" represent a set consisting of the SP tree with all detected recombination events removed plus any postrecombination evolutionary tree structure that accumulated in the descendants of the original recombinant. The numbers of nonsynonymous and synonymous mutations on each branch were recorded for each subject. In subjects for whom there was no unique SP solution, we used the topology that gave the most conservative (least significant) results in the logit and contingency analyses described below.

Statistical analysis of selection and demography: The tools used here for detecting selection and testing the hypothesis of neutrality and demographic stability are based on an extension of the contingency table approach developed by TEMPLETON (1996), a logit analysis (FIENBERG 1977), and homogeneity tests of haplotype tree topology.

The neutral theory assumes that mutations fall into two categories with respect to natural selection: deleterious mutations and neutral mutations (KIMURA and OHTA 1971). Under this theory, deleterious mutations are rapidly eliminated by natural selection and therefore

do not contribute to DNA evolution. Neutral mutations are not subject to natural selection, but can increase in frequency and go to fixation due to the random evolutionary force of genetic drift. Hence, the neutral theory predicts that only neutral mutations contribute to DNA evolution. Under neutrality, the rate of evolution equals the neutral mutation rate (KIMURA and OHTA 1971). The rate of neutral mutation is not necessarily constant over all genes or over all types of mutational change within a gene because the probability of a mutation being deleterious *vs.* neutral is hypothesized to vary across genes and across nucleotide positions within a gene. For example, it is widely accepted that synonymous changes, because they do not affect the amino acid composition of the coded protein, are less likely to be deleterious and more likely to be neutral than are nonsynonymous changes (MCDONALD and KREITMAN 1991; NIELSEN 2001). Although the neutral mutation rate may vary between synonymous and nonsynonymous positions, the ratio in evolutionary substitution rates between these mutational classes should be the constant ratio of their underlying mutation rates under the null hypothesis of selective neutrality.

The mutations that accumulate in a DNA region without recombination also define a haplotype tree that reflects the evolutionary origins of each haplotype. If all accumulated mutations are neutral, all topological positions in the evolutionary tree should reflect the same underlying constant neutral rates of mutational accumulation. Putting these two predictions together leads to a simple test for neutrality: a contingency test of homogeneity in which one dimension consists of mutational categories and the other dimension consists of distinct topological positions in the evolutionary tree of mutational variation (TEMPLETON 1987; MCDONALD and KREITMAN 1991). Under neutrality, homogeneity is expected in such a contingency table even though the distinct mutational categories may have different underlying neutral mutation rates.

The original topological positions were "fixed" (a branch in the intra-interspecific tree that was a connection between two species) and "polymorphic" (branches of the tree found within a single species; TEMPLETON 1987; MCDONALD and KREITMAN 1991), but these categories are not applicable to the HIV-1 sequence data set. Another meaningful categorization of evolutionary position is those mutations falling on "tip" branches *vs.* "interior" branches (FU and LI 1993; TEMPLETON 1996). A tip haplotype is connected to only one other haplotype or node in the tree. An interior haplotype is connected to two or more other haplotypes or nodes in the tree and hence represents an interior node in a topological sense. Interior haplotypes tend to be older than tip haplotypes, to be more frequent in the gene pool, and to have left descendant haplotypes (CASTELLOE and TEMPLETON 1994). Hence, interior haplotypes have demonstrated a degree of evolutionary success that has not yet been demonstrated by tip haplotypes, many

of which may turn out to be evolutionary dead ends. The topological classes of tip *vs.* interior exist in both intraspecific and combined intra-interspecific trees and therefore can be applied to studies on HIV-1 evolution within subjects.

Another criterion of evolutionary success is temporal persistence. Our data consist of viral samples taken at multiple time points (visits) for each subject. Hence, the branches of the intraspecific tree of gp120 genes can also be subdivided into two other topological classes: intravisit *vs.* intervisit. Intravisit branches are those that interconnect haplotypes sampled at the same visit without connecting through a previously sampled haplotype or node in a previous intravisit tree. Intervisit branches are those that connect a haplotype or monophyletic clade of haplotypes observed in the sample at one visit to haplotypes or nodes in the haplotype tree defined by haplotypes sampled in earlier visits. Note that an intervisit branch will sometimes contain mutations that occurred during the same time period as some intravisit mutations, but the important feature is not the time *per se* during which a particular mutation occurred, but rather the qualitative feature that this branch marks a lineage that was able to persist over a longer period of time than an intravisit branch. Hence, mutations on the intervisit branches, regardless of the specific time period in which they arose, have demonstrated their evolutionary success in persisting through time from at least one visit to the next, whereas mutations on intravisit branches have not demonstrated such temporal persistence. Intervisit and intravisit branches can also be subdivided into tips and interiors, but sample size considerations led us to pool all intervisit branches together. We therefore examine how synonymous *vs.* nonsynonymous mutations are distributed over three topological categories of branches in the gp120 gene trees: intravisit tip, intravisit interior, and intervisit.

Under neutrality, all contingency homogeneity tables should be homogeneous across the mutational/phylogenetic classes. The null hypothesis of homogeneity could be tested with a standard contingency chi-square test for homogeneity (TEMPLETON 1996), but with three or more ordered tree topological categories, the Jonckheere-Terpstra (JT) test can be a more powerful alternative (HETTMANSPERGER 1984). The impact of selection should be strongest on the intervisit branches and weakest on the intravisit tip branches. The Jonckheere-Terpstra test takes into account this *a priori* ordination. However, we have no *a priori* expectation as to the relative numbers of nonsynonymous *vs.* synonymous mutations in each of these tree topology categories when selection is present. If selection were positive, we would expect increasing deviations in favor of nonsynonymous mutations in going from tip to intervisit branches, but if selection were purifying, we would expect the opposite pattern. Therefore, a two-tailed Jonckheere-Terpstra test is used. Because we frequently had some cells in the contingency tables with fewer than five observations,

all tests were implemented as exact tests using the program StatXact 5 (Cytel Software).

To increase power with sparse data sets, we sometimes pooled the intravisit interior and intervisit classes together, as both represent branches with proven evolutionary success. As there are only two topological categories in this case, the Jonckheere-Terpstra test is not needed, so only a standard contingency test is used. Because some cells had fewer than five observations, the significance levels of these two-by-two tables were evaluated with Fisher's two-tailed exact test.

A cross-classified table can be extended to more than two dimensions, and a useful method of analysis of higher-dimension contingency tables is a logit analysis. A logit analysis requires a dichotomous response variable and categorical explanatory variables. The response variable used to detect selection was the number of nonsynonymous *vs.* synonymous mutations. These mutational counts were cross-classified with respect to five explanatory variables: (1) topological position in the haplotype tree (tip, interior, intervisit); (2) disease progression category (rapid, moderate, and nonprogressors); (3) CD4 T-cell brackets based on the CD4 T-cell counts associated with the particular subject visit (or the root visit for intervisit branches); (4) CD8 T-cell brackets; and (5) branches in the haplotype trees evolving within NSI *vs.* SI forms. Those branches transitional between NSI and SI forms were used in a separate analysis and are not included in either the contingency or the logit analyses. The amount of data in any single tree from one subject was insufficient for meaningful statistical inference, so we pooled the cross-classified data over all subjects.

A preliminary inspection of the data revealed that the NSI and SI portions of the trees evolved extremely differently. Therefore, we performed separate logit analyses on the NSI and SI portions of the data, with each analysis resulting in a five-dimensional table (the response variable plus the remaining four explanatory variables). The resulting five-dimensional tables had many zero entries. Most of these were structural zeros (FIENBERG 1977); that is, these cells were empty by definition. For example, by definition a nonprogressor cannot have CD4 counts in the lower two categories. Many other cells had sample zeros; that is, these cells were empty in the sample but in theory observations could have occurred within them. These were associated primarily with biologically unlikely but not impossible combinations, such as simultaneously having the lowest CD4 category and the highest CD8 category in a rapid progressor. Many nonzero cells had few observations for similar reasons. Accordingly, an exact logistic regression was implemented using the program LogXact 5 (Cytel Software). As a result of structural and sample zeros, a logit model that included all explanatory variables and all possible interactions among them was greatly overdetermined. We therefore tried many models to find the one with the least number of parameters that still fit

the data well as judged by a nonsignificant (5% level) deviance statistic (the log-likelihood ratio test of the specified model *vs.* a saturated model that fits the data perfectly).

We used a chi-square test of homogeneity to investigate the relative fitnesses of NSI and SI strains and to explore potential demographic regimes under which viral subpopulations evolve. In the contingency and Jonckheere-Terpstra tests, mutations were classified into six mutational/tree classes (intravisit tip, intravisit interior, intervisit), which were then split into nonsynonymous and synonymous. Under neutrality and demographic equilibrium, the proportion of mutations that fall into these six classes should be the same regardless of disease progression category, CD4 bracket, CD8 bracket, or NSI *vs.* SI forms. However, the topology and relative lengths of tip *vs.* interior branches can be altered by different selective and/or demographic regimes (Figure 4, redrawn from PAGE and HOLMES 1998). For example, positive diversifying selection or demographic subdivision will cause the haplotype tree to have longer internal branches and more intervisit branches relative to tip branches when compared to neutral expectations with stable population sizes, whereas population growth or recent selective sweeps will cause a shift toward longer tip branches relative to internal branches (PAGE and HOLMES 1998). Hence, rejecting the null hypothesis that homogeneity for mutational/tree classes across various categories is sensitive to heterogeneity in both selective forces and demographic attributes. We focused on demography by considering only synonymous mutations that are *a priori* more likely to be selectively neutral.

We used the synonymous branch-length data in two distinct, statistically independent fashions. First, we examined the intravisit tip *vs.* intravisit interior synonymous branch lengths across different categories of HIV types or clinical conditions. This contrast tests the null hypothesis of homogeneity in relative growth rates (fitness). The second contrast pools all the intravisit synonymous branches together and contrasts them to the intervisit synonymous branch lengths. This test is statistically independent of the first because the categories tested for heterogeneity in the first test are pooled together in the second test, which is therefore mathematically invariant to the results of the first test. The relative mutational count on the intervisit to intravisit branches is an indicator of demographic subdivision because with increasing subdivision there should be more intervisit branches.

RESULTS

SP haplotype trees or sets of SP trees were estimated for all 15 subjects under the assumption of no recombination or gene conversion. Figure 1 shows the SP haplotype tree estimated for subject 7, which was chosen because this subject is intermediate in our set of 15 subjects
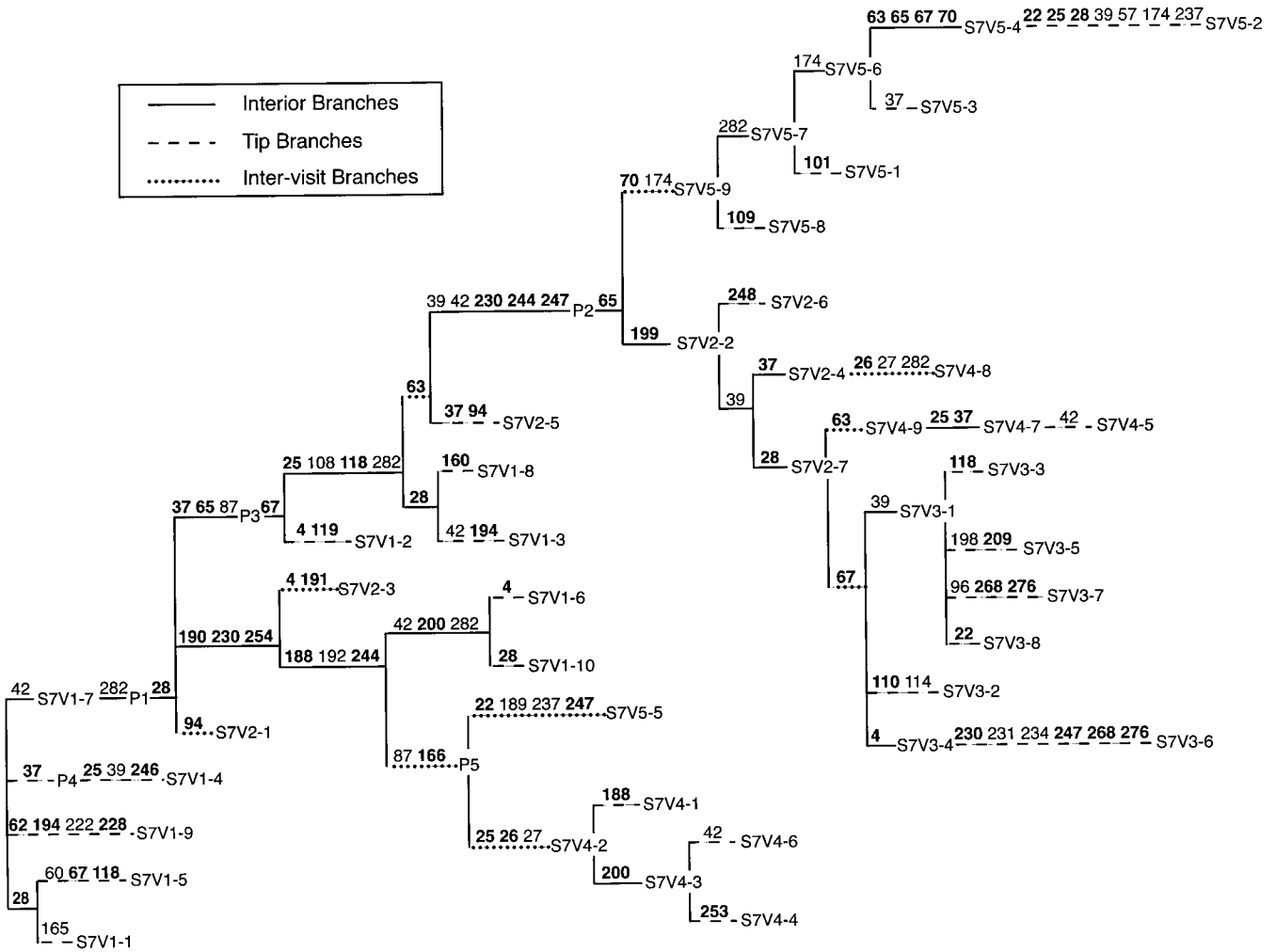
FIGURE 1.—The SP haplotype tree for subject 7 under the assumption of no recombination or gene conversion. Branches showing evolutionary change are given by horizontal lines. These lines are solid if the branch is an intravisit interior, dashed if an intravisit tip, and dotted if an intervisit branch. Vertical lines do not indicate any evolutionary change, but rather are used to show when multiple lineages diverge from a single ancestral haplotype or node in the tree. The numbers above the horizontal lines indicate the nucleotide positions that mutated on that branch. If the number is in boldface type, that mutation was nonsynonymous; otherwise, it was a synonymous mutation. Haplotypes are designated by the general format "S7Vx–y," where S7 indicates that the haplotype is from subject 7, Vx indicates that the haplotype was observed at visit number x of subject 7 (there were a total of five visits for this subject), and y is a number assigned to identify the distinct haplotypes observed at visit x. Nodes labeled Pz, where z is an integer from 1 to 5, refer to putative parental types involved in recombination and/or gene conversion events (see Figure 2).

with respect to disease progression, mutational diversity, and amount of recombination. The estimated SP trees or tree sets for the other 14 subjects are available on request to A. R. Templeton.

Following estimation of the SP trees under the assumption of no recombination or gene conversion, we tested for statistically significant recombination and/ or gene conversion events. A total of 33 events were detected, ranging from 0 to 10 per subject. At least 1 recombination/gene conversion event was detected in 11 of the 15 subjects. Figure 2 presents the 5 statistically significant recombination and gene conversion events detected in subject 7. As Figure 2 illustrates, additional evolutionary changes usually occurred after a recombi-

nant haplotype was created, including both the accumulation of additional mutational change and the production of new clades of haplotypes. The details of the other 28 recombination/gene conversion events are available on request to A. R. Templeton. After inferring the recombination and gene conversion events, we eliminated the "branches" in the original SP haplotype tree that contained the homoplasies inferred to be due to recombination, resulting in a tree that no longer displays connections to the inferred recombinants and their descendants. Figure 3 shows the SP tree for subject 7 after the five recombination/gene conversion events in Figure 2 have been removed from the original SP tree in Figure 1.
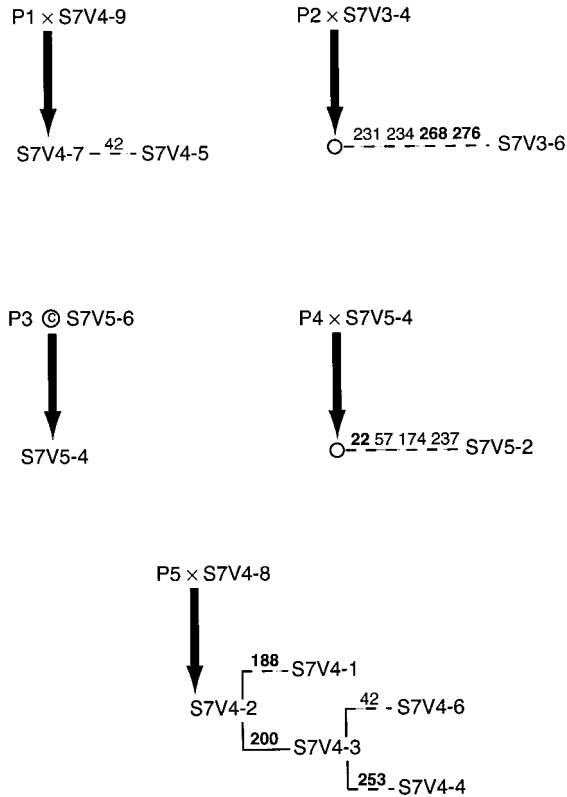
FIGURE 2.—The statistically significant ($P \leq 0.05$) recombination and/or gene conversion events inferred for subject 7. A total of five events were detected: four recombination events ($\times$ between the two putative parental types) and one gene conversion event (© between the two putative parental types). A thick, solid arrow points to the product of recombination and/or gene conversion, which can be either an observed haplotype (designated as in Figure 1) or a haplotype not directly observed in any sample but that left descendants affected by postrecombinational mutational change. Any evolutionary change that occurred in the descendants of the original recombinant is indicated as in Figure 1.

After removing the homoplasies due to recombination, mutational counts were made on the remaining cladistic structure, with each individual mutation characterized as synonymous or nonsynonymous and by branch location (intravisit tip, intravisit interior, and intervisit). For example, in subject 7 only those mutations shown in Figures 2 and 3 are counted because the additional "mutational" changes in Figure 1 are due to statistically significant recombination events and not to new mutational events. Similar counts were performed for the other 14 subjects. Also, the CD4 and CD8 counts for each visit were noted for the portion of the tree associated with that visit and for branches extending from that visit to future visits. Table 1 shows the Jonckheere-Terpstra test of the pooled data for all subjects of the mutational categories of nonsynonymous and synonymous *vs.* the tree topological positions of intravisit tip, intravisit interior, and intervisit branches.

Table 2 shows the results for the separate logit analyses of the NSI and SI portions of the haplotype trees. Only the logit model with the fewest parameters that fits the data is given in Table 2. In both cases, eliminating any of the regression parameters shown in Table 2 results in models that are rejected at least at the 5% level.

The simplest logit model that fits the NSI data includes only the topological categories of the branches. The simplest logit model that fits the SI data includes topological position in the tree, the interaction of topological position with CD4 category, and the three-way interaction of topological position, CD4, and CD8 (Table 2). Because this model indicates that CD4 and CD8 levels interact in inducing heterogeneity across topological positions, the SI data are partitioned into all eight, nonzero combinations of CD4 and CD8 levels (the combination of CD4 < 200 and CD8 > 1050 had no observations). Each combination is then analyzed with the Jonckheere-Terpstra test of nonsynonymous/synonymous mutations *vs.* tree topological categories and with the Fisher's exact test by pooling the intravisit interior and intervisit branches to yield a two-by-two table. The only significant test was for the combination CD4 > 650 and CD8 > 1050, which yielded a Jonckheere-Terpstra test statistic of -2.037 with a *P*-value of 0.0471 and Fisher's exact *P*-value of 0.0010.

As discussed below, the above results indicated that NSI forms are subject to positive selection in general, whereas the SI forms are subject to strong positive selection when the immune system is healthy as indicated by CD4 counts >650 and CD8 counts >1050. SI forms are also subject to positive, albeit weaker, selection when the immune system is severely compromised, as indicated by CD4 counts <200. We therefore performed standard contingency chi-square tests on SI *vs.* NSI synonymous branch-length counts under these two sets of clinical conditions, with respect to both intravisit tips *vs.* interiors and intravisit *vs.* intervisit branches (Table 3). A subpopulation growing more rapidly than a second subpopulation should have proportionally longer tip branches to interiors (Figure 4), so we measure the relative growth rates (fitnesses) of SI to NSI forms as (tip synonymous length/interior synonymous length)$_{SI}$/ (tip synonymous length/interior synonymous length)$_{NSI}$. These relative fitness estimates are also given in Table 3. Finally, we expect the ratio of intervisit synonymous branch length to intravisit synonymous branch length to increase with increasing intensity of diversifying selection and/or demographic subdivision. Therefore, we measure the relative amount of subdivision/diversification of SI to NSI forms as (intervisit synonymous length/ intravisit synonymous length)$_{SI}$/(intervisit synonymous length/intravisit synonymous length)$_{NSI}$. These estimators also appear in Table 3.

There are two potential complications in making demographic inferences from Table 3. First, although Table 3 presents a demographic contrast of SI *vs.* NSI forms, the SI form does not evolve in all subjects, and
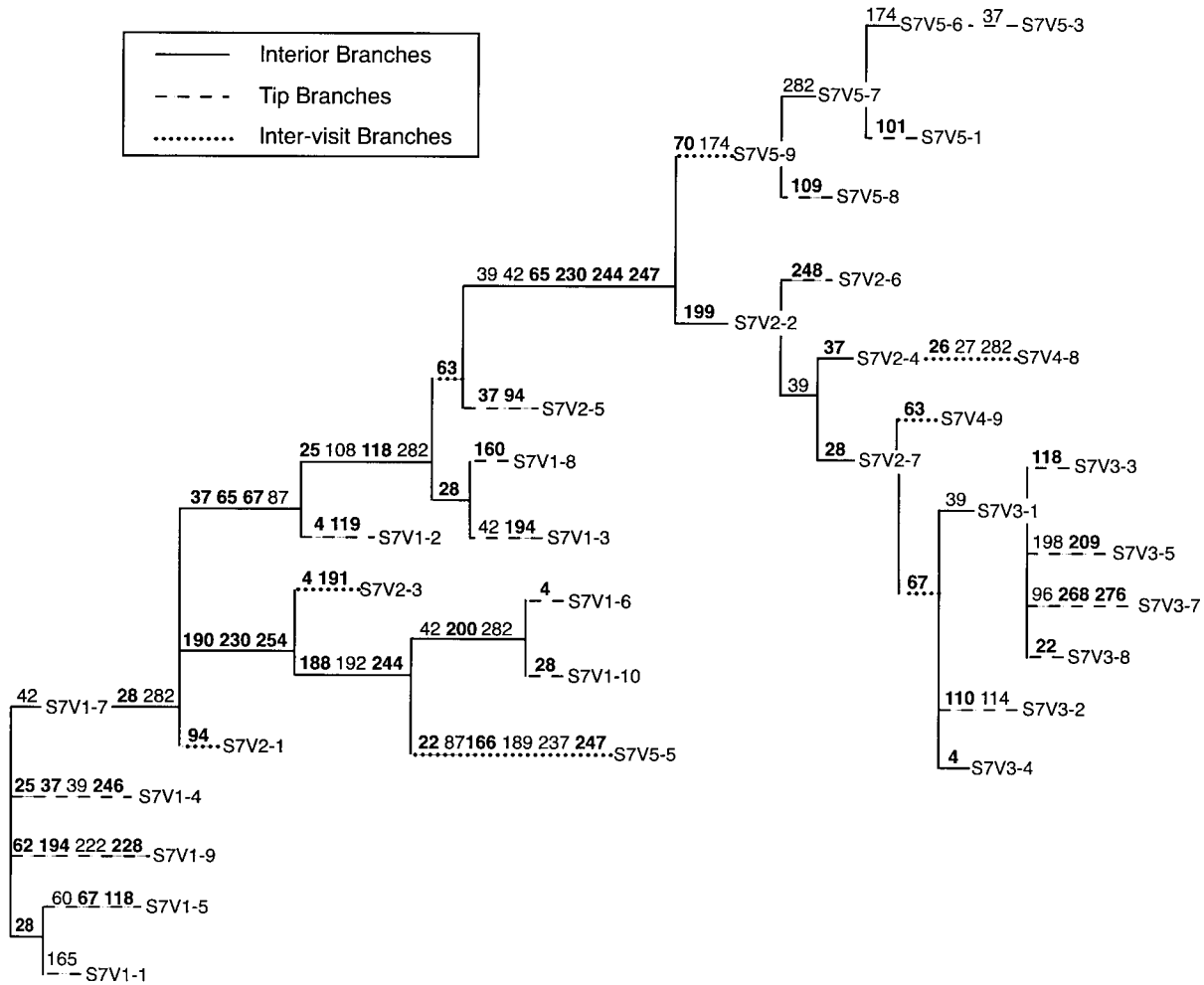
FIGURE 3.—The SP haplotype tree for subject 7 after the recombination and/or gene conversion events shown in Figure 2 have been peeled off the original SP tree (Figure 1).

it is not present during all visits (particularly early visits) in some of the subjects in which it evolves. Hence, many of the entries in Table 3 for NSI come from subjects and visits in which no SI were present. We therefore did a second analysis by limiting the NSI data to observations made in the presence of SI to examine the comparative demographics of NSI to SI when both coexist. All NSI

observations in subjects with a compromised immune system (CD4 < 200) were in the presence of SI, so the results given in Table 3 for a compromised immune system are unaltered. Neither of the tests in Table 3 for the healthy immune system was significant, and when inference was restricted to NSI forms in the presence of SI, both of the tests remained nonsignificant (P =

TABLE 1

The Jonckheere-Terpstra test of the pooled data for all subjects of the mutational categories of nonsynonymous and synonymous *vs.* the tree topological positions of intravisit tip, intravisit interior, and intervisit branches

	Phylogenetic position		
Mutational type	Intravisit tip	Intravisit interior	Intervisit
Nonsynonymous	552	290	219
Synonymous	247	75	61
Ratio nonsynonymous/synonymous	2.24	3.87	3.59

Standarized JT statistic, -3.929; two-tailed exact, $P = 0.0001$. P, the exact probability of the null hypothesis being true given the data. Negative JT statistics indicate positive selection, with increasing magnitudes and decreasing P-levels indicating increasing selection.

TABLE 2

The results of logit regression for NSI and SI forms of HIV

HIV type	Minimum parameter model that fits data with $P > 0.05$	Regression coefficients	Deviance	d.f.	P
NSI	Constant	-0.8614^a	0.2146	1	0.6432
	Tree topology	-0.1895^a			
SI	Constant	-0.7756^a	16.8181	11	0.1134
	Tree topology	-0.3882			
	Tree topology $\times$ CD4	-0.0264			
	Tree topology $\times$ CD4 $\times$ CD8	0.0172			

d.f., degress of freedom; P, the probability under the hypothesis that the data are drawn from the estimated model.

[a] Significantly different from zero at the 5% level.

0.2778 for the test of intravisit tips *vs.* intravisit interiors, and $P = 0.5453$ for the test of intravisit *vs.* intervisit branches). Hence, restricting the analysis to cases in which NSI coexists with SI alters none of the inferences in Table 3.

The other potential complication in making demographic inferences from Table 3 is the possibility of demographic heterogeneity induced by the use of antiviral drugs. Although *env* is not the target of the antiviral drugs used, it is possible that such drugs could affect the overall demography of SI and NSI forms and thereby influence the tests results in Table 3 that are designed to be sensitive to demographic conditions. The tests were therefore repeated by splitting each contingency contrast in Table 3 into two separate contrasts: one for observations made under antiviral drug therapy, and the other for observations made with no antiviral drug therapy (the numbers given in parentheses in Table 3). All SI observations made in individuals with healthy immune systems occurred in subjects who did not re-

TABLE 3

Tests of homogeneity in demography in SI *vs.* NSI forms when the immune system is healthy (CD4 > 650, CD8 >1050) and when the immune system is severely compromised (CD4 < 200)

Immune system state	HIV type	Total synonymous branch lengths		Two-tailed Fisher exact test results	Relative fitness of SI/NSI
		Intravisit tip	Intravisit interior		
Healthy	SI	9 (9)	4 (4)	$P = 1$	0.96
	NSI	28 (19)	12 (9)	(No AVD: $P = 1$)	(No AVD: 1.07)
				(AVD: not defined)	(AVD: not defined)
Compromised	SI	39 (23)	5 (1)	$P = 0.0195$	6.5
	NSI	6 (4)	5 (3)	(No AVD: $P = 0.0278$)	(No AVD: 17.25)
				(AVD: $P = 0.2513$)	(AVD: 4.0)

Immune system state	HIV type	Total synonymous branch lengths		Two-tailed Fisher exact test results	Relative subdivision of SI/NSI
		Intravisit	Intervisit		
Healthy	SI	13 (13)	5 (5)	$P = 0.2667$	2.56
	NSI	40 (28)	6 (1)	(No AVD: $P = 0.0249$)	(No AVD: 10.77)
				(AVD: not defined)	(AVD: not defined)
Compromised	SI	44 (24)	2 (1)	$P = 0.5080$	0.50
				(No AVD: $P = 0.4318$)	(No AVD: 0.29)
	NSI	11 (7)	1 (1)	(AVD: $P = 1$)	(AVD: not defined)

Numbers in parentheses are the mutational counts when inference is restricted to subjects and visits during which no antiviral drug (AVD) therapy was prescribed. Subtracting the number in parentheses from the corresponding number with no parentheses yields the mutational count for the cases in which AVD was prescribed. Because of small cell counts, all contingency tables are tested with the Fisher exact test. The test results and statistics given in parentheses are the values obtained when the analysis is split into those receiving AVD and those not.

Neutral Genetic Drift,	Neutral Genetic Drift,	Negative Selection	Positive (Directional)	Positive (Diversifying)
Stable Population Size	Expanding Population Size		Selection or	Selection or
				Bottleneck	Demographic Subdivision

Topological Changes Relative	More Mutations on Tip	More Mutations on	More Mutations on Tip	More Mutations
to Selective Neutrality and	Branches Relative to	Interior Branches	Branches Relative to	on Intervisit Branches
Demographic Stability:	Interior Branches	Relative to	Interior Branches
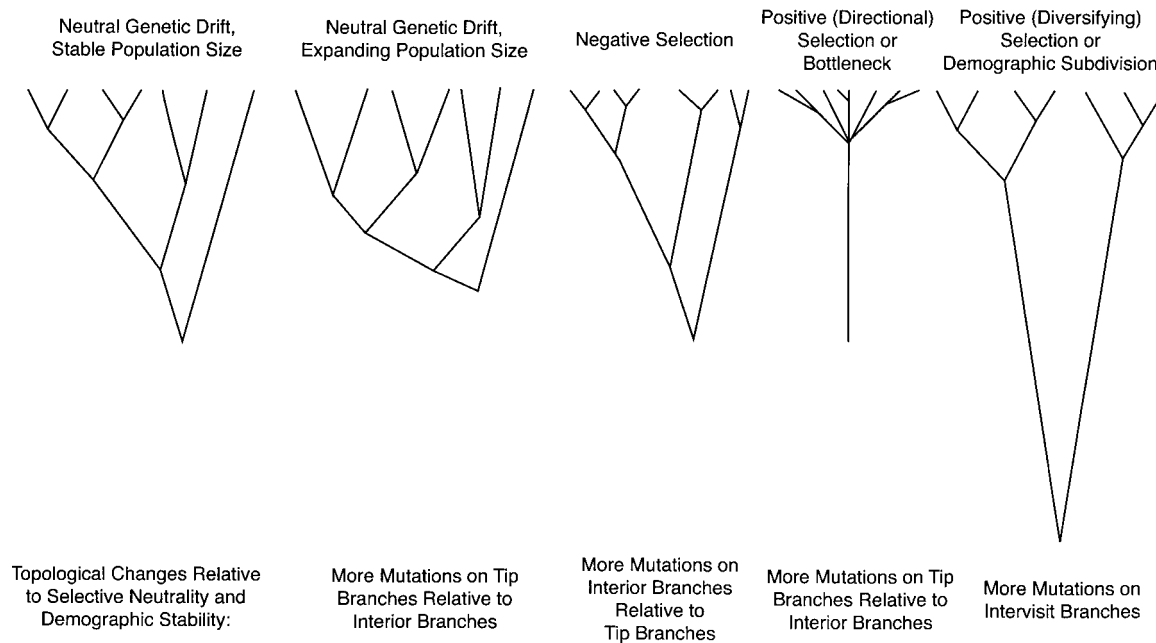			Tip Branches

FIGURE 4.—Expected haplotype tree topologies under various selective and demographic models (redrawn from PAGE and HOLMES 1998).

ceive antiviral drug therapy, so no tests of SI *vs.* NSI in individuals with healthy immune systems were possible. All test results and fitness and subdivision estimators for the analyses stratified by whether or not antiviral drug therapy was prescribed are given in parentheses in Table 3.

The data in Table 3 can be reorganized to test potentially different demographic regimes within SI and NSI forms when the human immune system is healthy *vs.* compromised. These test results are given in Table 4, as well as the estimators of relative fitness and subdivision within SI and NSI forms under healthy *vs.* compromised immune system conditions. All the tests obtained by confining the NSI inferences to those cases in which SI was present were not significant, so no alteration in the inferences shown in Table 4 occurred. Table 4 also shows the results of the analyses when stratified by whether or not antiviral drug therapy was prescribed.

All of the above analyses excluded those mutations found on the transitional branches between NSI and SI forms. Table 5 presents the contingency analysis of the branches containing amino acid replacement mutations that alter the NSI/SI phenotype status *vs.* all other branches. Because the branches altering NSI/SI phenotypes must have one nonsynonymous mutation, the mutation altering the phenotype is removed from the branches. Hence, the contingency analysis deals with only replacement mutations other than the defining ones at positions 306 or 320.

DISCUSSION

Table 1 shows strong evidence for selection when the data are pooled over all subjects and visits. The rejection

of the null hypothesis of neutrality is due to excesses of nonsynonymous changes for intravisit interior and intervisit branches, the branches of proven evolutionary success. Under the assumption that synonymous mutations are more likely than nonsynonymous mutations to be neutral, this analysis indicates the presence of overall positive selection that favors changing gp120 amino acid compositions in a diversifying and/or a directional manner.

The logit analyses in Table 2 reveal that this overall pattern of positive selection contains underlying heterogeneity. One source of heterogeneity is the distinction between NSI and SI forms. Both logit analyses reveal a significant effect of tree topological position on the ratio of nonsynonymous to synonymous substitutions, with the marginal logit-regression coefficient on tree topology being negative. The model was defined in such a way that a negative tree topology coefficient indicates increasing proportions of nonsynonymous to synonymous mutations from intravisit tips to intervisit branches. Hence, the logit analyses indicate that positive selection occurs within both viral forms. However, the logit analyses also indicate that the context in which this selection occurs is different in NSI and SI forms. For NSI forms, only the tree topology regression coefficient is needed to fit the data well, and the logit analysis reveals no significant effect on this positive selection by disease progression category, CD4 category, CD8 category, or their interactions. Hence, the NSI form is under positive selection over a broad range of conditions. In contrast, logit analysis (Table 2) indicates that selection of SI forms is context dependent as a function of both CD4 and CD8 levels.

No single regression coefficient is significant in the

TABLE 4

Tests of homogeneity in demography within SI and NSI forms when the immune system is healthy (CD4 > 650, CD8 >1050) *vs.* when the immune system is severely compromised (CD4 < 200)

| HIV type | Immune system state | Total synonymous branch lengths | | Two-tailed Fisher exact test results | Relative fitness: healthy/compromised |
		Intravisit tip	Intravisit interior		
SI	Healthy	9 (9)	4 (4)	$P = 0.1872$	0.29
	Compromised	39 (23)	5 (1)	(No AVD: $P = 0.0262$)	(No AVD: 0.10)
				(AVD: not defined)	(AVD: not defined)
NSI	Healthy	28 (19)	12 (9)	$P = 0.4719$	1.94
	Compromised	6 (4)	5 (3)	(No AVD: $P = 0.6697$)	(No AVD: 1.58)
				(AVD: $P = 0.5467$)	(AVD: 3.00)

| HIV type | Immune system state | Total synonymous branch lengths | | Two-tailed Fisher exact test results | Relative subdivision: healthy/compromised |
		Intravisit	Intervisit		
SI	Healthy	13 (13)	5 (5)	$P = 0.0157$	8.46
	Compromised	44 (24)	2 (1)	(No AVD: $P = 0.0672$)	(No AVD: 9.23)
				(AVD: not defined)	(AVD: not defined)
NSI	Healthy	40 (28)	6 (1)	$P = 1$	1.65
	Compromised	11 (7)	1 (1)	(No AVD: $P = 0.3904$)	(No AVD: 0.25)
				(AVD: $P = 0.5322$)	(AVD: undefined)

Because of small cell counts, all contingency tables are tested with the Fisher exact test. The counts for those subjects and visits receiving no prescription of antiviral drugs are shown in parentheses.

logit analysis of SI in Table 2. However, all three model parameters are needed together to fit the data; dropping any one results in a significant deviance. The slight negative regression coefficient associated with the tree topology × CD4 parameter suggests that there is a tendency for positive selection to become stronger with increasing levels of CD4. This result can be seen in the fact that all but one of the JT statistics were negative (data not shown), indicating positive selection. The strongest selection detected by the significant JT statistic is associated with the highest CD4 levels, where the fewest SI strains and smallest mutational sample sizes were observed. The lack of marginal significance of the tree topology × CD4 parameter is not surprising in light of the fact that intermediate levels of CD4 are associated with levels of selection (an average JT of -0.2997) weaker than those at either low CD4 levels <200 (an average JT of -1.2311) or high levels >650 (an average JT of -1.4878). The positive regression coefficient associated with the tree topology × CD4 × CD8 parameter suggests a synergistic effect between CD4 and CD8 levels in inducing positive selection. The strength of selection is greatly augmented by simultaneously having high levels of both CD4 and CD8. Overall, Tables 2 and the JT statistics indicate that SI forms are subjected to their strongest positive selection when the immune system is healthy as indicated by high counts of both CD4 and CD8. The SI forms are subject to little or weak selection in their V3 region when the immune system is at intermediate levels of functioning, but positive selection re-

turns when the immune system is severely compromised as indicated by CD4 counts <200.

The results in Tables 3 and 4 give additional insight into the differences between NSI and SI forms in these subjects. Table 3 shows that when the immune system is healthy, the fitnesses (relative growth rates) of the SI and NSI forms are not significantly different either for all subjects or for only those subjects not receiving antiviral drug therapy (no contrast between SI and NSI is possible in those subjects receiving such therapy as none of them had any SI forms). However, when the immune

TABLE 5

Contingency analyses of the nonsynonymous *vs.* synonymous mutations in NSI/SI transitional branches *vs.* all other branches for all subjects

| Mutational type | Phylogenetic position | |
	NSI/SI transitional branch	All other branches
Nonsynonymous	38	1004
Synonymous	6	377

Contingency chi-square = 4.05, d.f. = 1, and $P = 0.0442$. d.f., degrees of freedom; P, the probability of the null hypothesis being true given the data. Because all transitional branches by definition have a nonsynonmous mutation at either positions 306 or 320, that mutation was removed from all counts of nonsynonymous mutations to eliminate the ascertainment bias.

system is severely compromised, the SI form has a much higher and significantly different fitness than the NSI form (Table 3), and the effect is particularly strong in subjects not receiving antiviral drug therapy (over a 17-fold difference). The relative fitness of SI to NSI is 4.0 in those subjects receiving antiviral drug therapy, but in this case the value is not significantly different from 1 as shown by Fisher's exact test (Table 3). However, there were very few observations for this contrast, so the lack of significance may be due to small sample size. Indeed, the relative fitnesses in recipients and nonrecipients of antiviral drug therapy are not significantly different using an exact test, indicating that the SI form has in general a growth advantage over the NSI form when the immune system is severely compromised but not when the immune system is healthy. Table 4 indicates that this effect is due to a large 10-fold increase in the growth rate of the SI form as the immune system goes from a healthy to a compromised state in subjects not receiving antiviral drug therapy (once again, the absence of SI in healthy individuals receiving antiviral drug therapy precludes testing the change in relative SI growth rates in those subjects). In contrast, there is no significant change in the growth rates of the NSI form, either overall or as a function of antiviral drug therapy status, and the fitness values suggest that, if anything, NSI does better when the immune system is healthy than when it is compromised—exactly the opposite of the SI pattern.

Table 3 also indicates that there is no significant difference in the amount of subdivision between SI and NSI in a compromised immune system state, but the SI form shows significantly more demographic subdivision than the NSI form under a healthy immune system when no antiviral drug therapy is being given. Unfortunately, small sizes for SI precluded testing the relative degree of subdivision in recipients of antiviral drug therapy with healthy immune systems. Table 4 indicates that the changes in the relative degree of subdivision of SI to NSI forms as the immune system changes from a healthy to a compromised state are due to changes in the amount of subdivision in the SI form. No significant changes in subdivision are found for NSI, but the SI form shows more subdivision when the immune system is healthy than when it is compromised, with the change being significant in the overall data and nearly significant (0.0672) when restricted to subjects not receiving antiviral drug therapy (once again, we cannot test this situation in the recipients of antiviral drug therapy because of missing classes). Thus, just as with the logit model on selection, Tables 3 and 4 reveal little demographic context dependency for NSI but significant context dependency for SI.

The patterns discussed above indicate that both NSI and SI forms are subject to positive selection in the V3 region, but in different contexts. Efforts to understand the biological basis for differences in evolutionary patterns of SI and NSI variants should consider the host immune system, the coreceptor environment at different times within the host, and the impact of the host environment on other factors in the virus life cycle. It is important to keep in mind that these explanations are not mutually exclusive and that components of all could be shaping the evolution of the V3 region.

We first consider why SI viruses are under strong positive selection when the immune system is intact. Since gp120, particularly the V3 loop, is an immune system recognition site of HIV-1, selective pressure for "escape mutants" should affect this region (LEIGH BROWN and HOLMES 1994; MAY 1995; NOWAK *et al.* 1995; ROWLAND-JONES *et al.* 2001). In this regard, recent studies (CALLAWAY *et al.* 1999; SHIINO *et al.* 2000) provide evidence for the hypothesis that the V3 loop of the NSI form is more hidden from neutralizing antibody than is the SI V3 loop; that is, the amino acid changes necessary for the SI form simultaneously make SI more apparent to the immune system than NSI. Greater immune apparency is consistent with our observations that the SI forms experience their strongest positive selection when the immune system is healthy. As the immune system becomes compromised and then collapses, selection from this source should diminish and then disappear, and there should also be less of a growth inhibition on SI forms relative to NSI (CALLAWAY *et al.* 1999; SHIINO *et al.* 2000). The hypothesis of greater immune apparency of SI therefore could contribute to the large increase in growth rates of SI as the immune system collapses and to the achievement of greater fitness of SI relative to NSI after the immune system is severely compromised (Tables 3 and 4). This hypothesis could also explain the patterns of increased demographic subdivision in SI in individuals with healthy immune systems as the SI forms are driven to exist in those tissues and compartments where there is little antibody penetration. When the immune system collapses, the SI form could emerge from these tissues, resulting in a decrease in the amount of subdivision, as was observed (Table 4). This change in demographic and selective regimes with a collapsing immune system could contribute to the switch in phenotypic dominance from NSI to SI over the course of the infection in rapid progressors (CALLAWAY *et al.* 1999; SHIINO *et al.* 2000). However, immune apparency alone would not fully explain why SI viruses become dominant when the immune system collapses, since both viral phenotypes would be under less pressure from the immune system.

Another possible reason for the relative fitness of SI forms to NSI forms increasing when the immune system collapses is suggested by the work of XIAO *et al.* (2000), which indicates that NSI forms can suppress replication of SI forms when NSI are abundant. The transition from NSI to SI is marked by the evolution of coreceptor use, with the NSI form primarily using the human CCR5 coreceptor to gain entry into host cells but with the SI

form using both CCR5 and CXCR4 coreceptors (VAN RIJ *et al.* 2000). Indeed, the amino acid motif in the V3 region that we used to define SI is strongly associated with CXCR4 usage (SHANKARAPPA *et al.* 1999; BRIGGS *et al.* 2000). XIAO *et al.* (2000) report that the HIV-1 Tat protein, which is secreted from virus-infected cells, is a CXCR4-specific antagonist that can selectively inhibit the entry and replication of SI, but not NSI, forms in peripheral blood mononuclear cells, thereby restricting the target cell substrate preferentially for the SI form. Because SI forms also secrete Tat, there may be some degree of self-inhibition, but under this model the inhibition of SI is more severe when the NSI forms dominate in abundance over SI and create a level of inhibition far greater than SI self-inhibition. NSI forms predominate when the immune system is healthy. As the immune system declines, NSI is no longer as abundant, so the replication suppression on SI should decrease to the SI self-inhibition levels, leading to increased relative fitness of SI to NSI when the immune system is severely compromised. However, a recent analysis of the relationship between CCR5-expressing and CXCR4-expressing cells and viral phenotype has suggested that a restricted target cell substrate for viral replication is not a limiting factor in virus selection and therefore may not be a driving force in the shift from NSI to SI dominance (VAN RIJ *et al.* 2003). However, even if this inhibitory effect does not dominate the demographic shifts in NSI and SI, it could still induce selection on the V3 region and thereby contribute to the selective patterns that we observed.

The logit analysis (Table 2) indicates that CD8 plays an important role in defining the selective and demographic context for SI. This pattern may be a consequence of the evolution of coreceptor usage mentioned above. The CD4+ T cells are attacked by the NSI form through subversion of the CCR5 chemokine signaling pathway (ROWLAND-JONES *et al.* 2001). However, CD8+ T-cell survival is not significantly reduced by the initial HIV-1 infection, and CD8+ T-cell production is initially modestly increased (ROWLAND-JONES *et al.* 2001). However, CD8+ T-cell viability is affected by CXCR4-using viruses (BLANCO *et al.* 2001), and as the disease progresses the absolute number of CD8+ T lymphocytes is decreased in peripheral blood and their turnover rate is increased (HERBEIN *et al.* 1998). HERBEIN *et al.* (1998) report results that indicate that the increased cell death of CD8+ T cells in HIV-infected subjects is mediated by the HIV envelope protein through the CXCR4 chemokine receptor. All these results suggest that SI strains interact more with CD8+ T cells than do NSI strains, which is compatible with our observation in the logit analysis that the CD8+ T-cell counts are predictive of the selective and demographic attributes of SI but not of NSI.

Expression of the activation marker CD38 has also been shown to favor infection by CXCR4-using viral forms (HORIKOSHI *et al.* 2001). The fact that expression of this activation marker increases as disease progresses (KESTENS *et al.* 1994) indicates an increased substrate for replication of CXCR4 virus replication, perhaps reflecting the increasing relative abundance of CD8+ to CD4+ cells as the disease progresses. This in turn would facilitate the increased growth of SI relative to NSI forms as the disease progresses. As this favored target population of cells increases in abundance with disease progression, selective forces could favor increased specialization in SI viruses to use CXCR4. Indeed, the ability of SI forms to use CCR5 declines and is lost over time (VAN RIJ *et al.* 2000). These observations are consistent with the hypothesis that SI is selected to become more and more of a CXCR4-using specialist as the disease progresses. This in turn could also lead to less and less demographic subdivision as the SI forms become more homogeneous in the types of cells they target.

These studies collectively identify multiple sources of potential selection within and between the NSI and SI forms. None of the hypotheses outlined above can by themselves explain all our observations. However, these hypotheses are not mutually exclusive, and therefore multiple selective forces could be operating either together or in a shifting temporal dynamic. Indeed, shifts in temporal dynamics are inevitable for these selective components, given the collapse of the immune system as the disease progresses, the shift in relative abundance of NSI to SI forms, the changes in abundance of CD8+ T cells, and the increasing abundance of CD38-activated target cells. Moreover, other selective forces that we have failed to identify may be operating. Regardless, the results of this article show that the positive selection observed in the V3 region of HIV-1 is not one-dimensional or due to some uniform selective pressure. Rather, the patterns of positive selection are complex and temporally dynamic.

In addition to the varying selective and demographic patterns observed within the SI and NSI portions of the haplotype trees, Table 5 indicates that positive selection occurs on the branches that are transitional between the NSI and SI forms. The NSI/SI transitional branches have more than twice the rate of nonsynonymous substitutions as all other branches, even when the replacement mutations responsible for the NSI/SI transition are excluded. This result is even more biologically significant when the overall positive selection (Table 1) is taken into account; that is, the intensity of positive selection in the NSI to SI transition significantly exceeds that of the background positive selection and not just neutrality. Hence, the transition from the NSI to the SI phenotype (or its reverse in a few instances) is associated with strong positive selection on other amino acid sites, a result consistent with previous studies that indicate that although the switch from NSI to SI can be accomplished by a single amino acid substitution (DITTMAR *et al.* 1997), other amino acid sites in the V3 region and

elsewhere also influence the SI phenotype (Seillier-Moiseiwitsch *et al.* 1994). Our results indicate that once the SI-defining mutations occur, intense positive selection is induced on other amino acid sites, resulting in a burst of evolutionary change.

Overall, our results show that the V3 region is indeed subject to selection, but this selection cannot be meaningfully summarized in a marginal fashion or with respect to a single variable, such as disease progression or CD4 counts. Given our results, studies could observe very diverse patterns of V3 region evolution depending on whether they looked only at rapid or nonprogressors, whether they examined only subjects with NSI forms or subjects with SI forms, and whether they examined the disease early after seroconversion or only much later. Selection on the V3 region is intense, but highly context dependent. An understanding of this context dependency can yield greater insight into the evolutionary biology of this region of the HIV-1 genome.

LITERATURE CITED

Blanco, J. B., J. Barretina, C. Cabrera, A. Gutierrez and B. Clotet, 2001 CD4+ and CD8+ T cell death during human immunodeficiency virus infection in vitro. Virology **285:** 356–365.

Bonhoeffer, S., and M. A. Nowak, 1994 Mutation and the evolution of virulence. Proc. R. Soc. Lond. Ser. B **258:** 133–140.

Bonhoeffer, S., E. C. Holmes and M. A. Nowak, 1995 Causes of HIV diversity. Nature **376:** 125.

Briggs, D. R., D. L. Tuffle, J. W. Sleasman and M. M. Goodenow, 2000 Envelope V3 amino acid sequence predicts HIV-1 phenotype (co-receptor usage and tropism for macrophages). AIDS **14:** 2937–2939.

Callaway, D. S., R. M. Ribeiro and M. A. Nowak, 1999 Virus phenotype switching and disease progression in HIV-1 infection. Proc. R. Soc. Lond. Ser. B **266:** 2523–2530.

Carrieri, M. P., M. A. Chesney, B. Spire, A. Loundou, A. Sobel *et al.*, 2003 Failure to maintain adherence to HAART in a cohort of French HIV-positive injecting drug users. Int. J. Behav. Med. **10:** 1–14.

Castelloe, J., and A. R. Templeton, 1994 Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol. Phylogenet. Evol. **3:** 102–113.

Coffin, J. M., 1995 HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science **267:** 483–489.

Coffin, J. M., 1999 Molecular biology of HIV, pp. 3–40 in *The Evolution of HIV*, edited by K. A. Crandall. The Johns Hopkins University Press, Baltimore.

Crandall, K. A., 1994 Intraspecific cladogram estimation: accuracy at higher levels of divergence. Syst. Biol. **43:** 222–235.

Crandall, K. A., and A. R. Templeton, 1996 Applications of intraspecific phylogenetics, pp. 81–99 in *New Uses for New Phylogenies*, edited by P. Harvey, A. J. L. Brown, J. M. Smith and S. Nee. Oxford University Press, Oxford.

Crandall, K. A., and A. R. Templeton, 1999 Statistical approaches to detecting recombination, pp. 153–176 in *The Evolution of HIV*, edited by K. A. Crandall. The Johns Hopkins University Press, Baltimore.

De Jong, J. J., A. Deronde, W. Keulen, M. Tersmette and J. Goudsmit, 1992 Minimal requirements for the human-immunodeficiency-virus type-1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino-acid substitution. J. Virol. **66:** 6777–6780.

Dittmar, M. T., A. McKnight, G. Simmons, P. R. Clapham, R. A. Weiss *et al.*, 1997 HIV-1 tropism and co-receptor use. Nature **385:** 495–496.

Endo, T., K. Ikeo and T. Gojobori, 1996 Large-scale search for genes on which positive selection may operate. Mol. Biol. Evol. **13:** 685–690.

Fienberg, S. E., 1977 *The Analysis of Cross-Classified Categorical Data.* The MIT Press, Cambridge, MA.

Fu, Y. X., and W. H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Ganeshan, S., R. E. Dickover, B. T. M. Korber, Y. J. Bryson and S. M. Wolinsky, 1997 Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease. J. Virol. **71:** 663–677.

Goudsmit, J., 1997 *Viral Sex: The Nature of AIDS.* Oxford University Press, New York.

Herbein, G., U. Mahlknecht, F. Batliwalla, P. Gregersen, T. Pappas *et al.*, 1998 Apoptosis of CD8+ T cells is mediated by macrophages through interaction of HIV gp120 with chemokine receptor CXCR4. Nature **395:** 189–194.

Hettmansperger, T. P., 1984 *Statistical Inference Based on Ranks.* John Wiley & Sons, New York.

Horikoshi, H., M. Kinomoto, F. Sasao, T. Mukai, R. B. Luftig *et al.*, 2001 Differential susceptibility of resting CD4(+) T lymphocytes to a T-tropic and a macrophage (M)-tropic human immunodeficiency virus type 1 is associated with their surface expression of CD38 molecules. Virus Res. **73:** 1–16.

Kelly, J. K., 1994 An application of population genetic theory to synonymous gene sequence evolution in the human immunodeficiency virus (HIV). Genet. Res. **64:** 1–9.

Kestens, L., G. Vanham, C. Vereecken, M. Vandenbruaene, G. Vercauteren *et al.*, 1994 Selective increase of activation antigens HLA-DR and CD38 on CD4+ CD45RO+ T lymphocytes during HIV-1 infection. Clin. Exp. Immunol. **95:** 436–441.

Kimura, M., and T. Ohta, 1971 Protein polymorphism as a phase of molecular evolution. Nature **229:** 467–469.

Leigh Brown, A. J., and E. C. Holmes, 1994 Evolutionary biology of human immunodeficiency virus. Annu. Rev. Ecol. Syst. **25:** 127–165.

Liu, S. L., T. Schacker, L. Musey, D. Shriner, M. J. McElrath *et al.*, 1997 Divergent patterns of progression to AIDS after infection from the same source: human immunodeficiency virus type 1 evolution and antiviral responses. J. Virol. **71:** 4284–4295.

Markham, R. B., W. Wang, A. E. Weisstein, Z. Wang, A. Munoz *et al.*, 1998 Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline. Proc. Natl. Acad. Sci. USA **95:** 12568–12573.

May, R. M., 1995 The co-evolutionary dynamics of viruses and their hosts, pp. 553–585 in *Molecular Basis of Virus Evolution*, edited by F. García-Arenal. Cambridge University Press, Cambridge, UK.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

McDonald, R. A., D. L. Mayers, R. C.-Y. Chung, K. F. Wagner, S. Ratto-Kim *et al.*, 1997 Evolution of human immunodeficiency virus type 1 *env* sequence variation in patients with diverse rates of disease progression and T-cell function. J. Virol. **71:** 1871–1879.

Nielsen, R., 2001 Statistical tests of selective neutrality in the age of genomics. Heredity **86:** 641–647.

Nowak, M. A., R. M. May, R. E. Phillips, S. Rowland-Jones, D. G. Lalloo *et al.*, 1995 Antigenic oscillations and shifting immunodominance in HIV-1 infections. Nature **375:** 606–611.

Page, R. D. M., and E. C. Holmes, 1998 *Molecular Evolution: A Phylogenetic Approach.* Blackwell Science, Oxford.

Palepu, A., M. Tyndall, B. Yip, M. V. O'Shaughnessy, R. S. Hogg *et al.*, 2003 Impaired virologic response to highly active antiretroviral therapy associated with ongoing injection drug use. J. Acquir. Immune Defic. Syndr. **32:** 522–526.

Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard and D. D. Ho, 1996 HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science **271:** 1582–1586.

Rowland-Jones, S., S. Pinheiro and R. Kaul, 2001 New insights into host factors in HIV-1 pathogenesis. Cell **104:** 473–476.

Seibert, S. A., C. Y. Howell, M. K. Hughes and A. L. Hughes,

1995 Natural selection on the *gag, pol*, and *env* genes of human immunodeficiency virus 1 (HIV-1). Mol. Biol. Evol. **12:** 803–813.

Seillier-Moiseiwitsch, F., B. H. Margolin and R. Swanstrom, 1994 Genetic variability of the human immunodeficiency virus: statistical and biological issues. Annu. Rev. Genet. **28:** 559–596.

Shankarappa, R., P. Gupta, G. H. Learn, A. G. Rodrigo, J. C. R. Rinaldo *et al.*, 1998 Evolution of human immunodeficiency virus type 1 envelope sequences in infected individuals with differing disease progression profiles. Virology **241:** 251–259.

Shankarappa, R., J. B. Margolick, S. J. Gange, A. G. Rodrigo, D. Upchurch *et al.*, 1999 Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J. Virol. **73:** 10489–10502.

Shiino, T., K. Kato, N. Kodaka, T. Miyakuni, Y. Takebe *et al.*, 2000 A group of V3 sequences from human immunodeficiency virus type 1 subtype E non-syncytium-inducing, CCR5-using variants are resistant to positive selection pressure. J. Virol. **74:** 1069–1078.

Templeton, A. R., 1987 Genetic systems and evolutionary rates, pp. 218–234 in *Rates of Evolution*, edited by K. S. W. Campbell and M. F. Day. Allen & Unwin, London.

Templeton, A. R., 1996 Contingency tests of neutrality using intra-interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. Genetics **144:** 1263–1270.

Templeton, A. R., K. A. Crandall and C. F. Sing, 1992 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132:** 619–633.

Templeton, A. R., A. G. Clark, K. M. Weiss, D. A. Nickerson, J. Stengård *et al.*, 2000 Recombinational and mutational hot-spots within the human *lipoprotein lipase* gene. Am. J. Hum. Genet. **66:** 69–83.

van Rij, R. P., H. Blaak, J. A. Visser, M. Brouwer, R. Rientsma *et al.*, 2000 Differential coreceptor expression allows for independent evolution of non-syncytium-inducing and syncytium-inducing HIV-1. J. Clin. Invest. **106:** 1039–1052.

van Rij, R. P., M. Hazenberg, B. van Benthem, S. Otto, M. Prins *et al.*, 2003 Early viral load and CD4+ T cell count, but not percentage of CCR5+ or CXCR4+ CD4+ T cells, are associated with R5-to-X4 HIV type 1 virus evolution. AIDS Res. Hum. Retroviruses **19:** 389–398.

Wei, X. P., S. K. Ghosh, M. E. Taylor, V. A. Johnson, E. A. Emini *et al.*, 1995 Viral dynamics in human-immunodeficiency-virus type-1 infection. Nature **373:** 117–122.

Wolinsky, S. M., B. T. M. Korber, A. U. Neumann, M. Daniels, K. J. Kunstman *et al.*, 1996 Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. Science **272:** 537–542.

Wood, E., J. S. G. Montaner, B. Yip, M. W. Tyndall, M. T. Schechter *et al.*, 2003 Adherence and plasma HIV RNA responses to highly active antiretroviral therapy among HIV-1 infected injection drug users. Can. Med. Assoc. J. **169:** 656–661.

Xiao, H., C. Neuveut, H. L. Tiffany, M. Benkirane, E. A. Rich *et al.*, 2000 Selective CXCR4 antagonism by Tat: implications for in vivo expansion of coreceptor use by HIV-1. Proc. Natl. Acad. Sci. USA **97:** 11466–11471.

Yamaguchi, Y., and T. Gojobori, 1997 Evolutionary mechanisms and population dynamics of the third variable envelope region of HIV within single hosts. Proc. Natl. Acad. Sci. USA **94:** 1264–1269.