Evolution of the human immunodeficiency virus type 1 subtype-specific V3 domain is confined to a sequence space with a fixed distance to the subtype consensus

Lukashov, V.V.; Goudsmit, J.

Publication date
1997

Published in
Journal of Virology

JOURNAL OF VIROLOGY, Sept. 1997, p. 6332–6338 Vol. 71, No. 9
0022-538X/97/$04.00+0

Evolution of the Human Immunodeficiency Virus Type 1 Subtype-Specific V3 Domain Is Confined to a Sequence Space with a Fixed Distance to the Subtype Consensus

VLADIMIR V. LUKASHOV AND JAAP GOUDSMIT*

Department of Human Retrovirology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

Received 4 November 1996/Accepted 20 May 1997

Human immunodeficiency virus type 1 (HIV-1) strains can be separated into genetic subtypes based on phylogenetic analysis of the envelope gene. Once it had been shown that population-wide intrasubtype genetic variation of HIV-1 strains increases in the course of the AIDS epidemic, it remained uncertain whether HIV-1 subtypes are phenotypic entities spreading as distinct virus populations. To examine this, we applied Eigen's concepts of sequence geometry and fitness topography to the analysis of intrasubtype evolution of the gp120 V3 domain of HIV-1 subtypes A, B, C, and D in the course of the global AIDS epidemic. We observed that despite the high evolution rate of HIV-1, the nonsynonymous distances to the subtype consensus of sequences obtained early in the epidemic are similar to those obtained more than 10 years later, in contrast to the synonymous distances, which increased steadily over time. For HIV-1 subtype B, we observed that the evolution rate of the individual sequences is independent of their distance from the subtype B consensus, but for the individual sequences most distant from the consensus evolution away from the consensus is constrained. As a result, individual HIV-1 genomes fluctuate within a sequence space with fixed distance to the subtype consensus. Our findings suggest that the evolution of the V3 domain of HIV-1 subtypes A, B, C, and D is confined to an area in sequence space within a fixed distance to the consensus of a respective subtype. This in turn indicates that each HIV-1 subtype is a distinct viral quasispecies that is well adapted to the present environment, able to maintain its identity in the V3 region over time, and unlikely to merge during progression of the AIDS epidemic.

The human immunodeficiency virus type 1 (HIV-1) genome is rapidly evolving. Genetic variation has been observed among HIV-1 sequences obtained within a single infected individual at different stages of HIV-1 infection (10, 11, 15, 16, 24, 26, 27, 37). This intrahost evolution, together with transmission bottlenecks, results in genetic variation of HIV-1 sequences among infected individuals within a population (interhost or population-wide evolution) (15, 16).

Based on phylogenetic analysis of the envelope gene, HIV-1 strains can be separated into genetic subtypes A through J (31). HIV-1 subtype B is geographically the most widely distributed and is associated with AIDS epidemics in the Americas, Europe, and Asia. HIV-1 subtypes A, C, and D are found mainly in Africa. Approximately 90% of the HIV-1 sequences obtained so far belong to these four HIV-1 subtypes, which are considered dominant HIV-1 subtypes, while for the other six HIV-1 subtypes only few sequences, typically from epidemiologically closely related individuals, are known.

For HIV-1 subtype B, which has been the most widely studied, the evolution rate of the envelope gene has been shown to be about 1% per year (24, 37). Genetic analysis of this gene, the third variable domain (V3) in particular, showed that the interhost (intrasubtype) variation increased over the course of the AIDS epidemic (15, 16). The question remained whether HIV-1 subtypes are phenotypic entities able to maintain their genetic integrity and spreading as distinct virus populations. Our study specifically addressed this issue by analyzing the evolution of the V3 domain of the envelope glycoprotein gp120 of HIV-1 subtypes A, B, C, and D.

The term "quasispecies" is often used to encompass HIV-1 sequence heterogeneity within a single infected individual (27). According to Eigen et al., a viral quasispecies or self-sustained virus population that is able to maintain its collective identity over time could be described as a "swarm" in multidimensional sequence space (4–7). Each individual sequence occupies an unique position within this swarm. The geometric center of a viral swarm or quasispecies is the consensus sequence.

To anticipate future evolution of HIV-1 subtypes, two scenarios can be envisioned. The first assumes that intrasubtype HIV-1 variation has not yet reached a limit and will expand over time until the HIV-1 subtypes lose their collective identity and merge with each other. At the population level, intrasubtype sequence variation will continually grow, as will the size (radius) of the sequence swarms of HIV-1 subtypes. At the individual level (intrahost evolution), this scenario will result in continual sequence evolution away from the subtype consensus, irrespective of the initial distance from the subtype consensus. Although the nucleotide distance between an individual sequence and the subtype consensus can reach 15%, 85% of nucleotide positions within those sequences are still identical. The vast majority of these positions, which are identical in the tested and the subtype consensus sequences, nevertheless vary in other sequences belonging to the same subtype. Thus, in the case of absent evolutionary constraints, the chance occurrence of new substitutions is more likely at those identical positions, increasing the distance between a tested sequence and the subtype consensus.

The second scenario assumes that the level of HIV-1 intrasubtype variation we now observe is nearing the limit of sequence distance to the consensus that is compatible with HIV-1 fitness or survival. At the population level, this scenario assumes

* Corresponding author. Mailing address: Department of Human Retrovirology, Academic Medical Center, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands. Phone: (31-20) 566-4522. Fax: (31-20) 691-6531.

a set maximum distance from the subtype consensus (the size of the subtype sequence swarm). As in the first scenario, the evolution of the individual HIV-1 sequences is continual. However, in the second scenario, sequences farthest from the consensus at some point begin to evolve toward the consensus. This constraint is predicted to be more pronounced for nonsynonymous substitutions (which result in a change of an amino acid) than for synonymous ones (which do not change the amino acid). This view is based on the premise that HIV-1 subtypes represent self-sustained populations that can maintain their integrity over time.

To examine these scenarios, we analyzed the population-wide evolution of the V3 region of the HIV-1 subtypes A, B, C, and D in the course of the global HIV-1 epidemic, as well as the intrahost HIV-1 subtype B evolution in 55 individuals over 5 years.

MATERIALS AND METHODS

The HIV-1 sequences used in this study were approximately 270 nucleotides long and contained the genomic region of *env* gp120 coding for the V3 domain (2). The majority of sequences were generated earlier by us, including those from the Amsterdam cohorts of homosexual men and intravenous drug users (22, 24), the Baltimore cohort of intravenous drug users (19, 25), individuals LAI and MN (NM) (21), the World Health Organization UNAIDS network study (2), individuals in Russia (20), and HIV-1-infected women from The Netherlands (23). The sequences are available through GenBank (for the accession numbers, see the original publications). In addition, we used the sequences available from the Los-Alamos HIV-1 database (31) (subtype C sequences from Ethiopia were kindly provided by Almaz Abebe prior to publication). The description of the sequence sets, their origin, and references are shown in Results. To analyze intrahost evolution, seroconversion sequences from 55 individuals of the Amsterdam cohorts were compared with sequences derived from the same individuals after 5 years of follow-up (22, 24).

The nucleotide sequences representing HIV-1 subtypes A, B, C, and D were aligned manually together with the consensus sequence of a respective HIV-1 subtype obtained from the Los Alamos HIV-1 database (31). All positions with an alignment gap in at least one sequence were excluded from pairwise sequence cparisons. Nucleotide, synonymous, and nonsynonymous p-distances (the proportion of differences, Hamming distances) between the individual sequences belonging to a certain HIV-1 subtype and the consensus sequence of the same subtype were estimated by using the MEGA program (17). The synonymous and nonsynonymous p-distances were calculated by using the Nei-Gojobori method (32) from the formulas $D_s = S_d/S$ and $D_a = N_d/N$, respectively, where D_s and D_a are the synonymous and nonsynonymous p-distances between two sequences, S_d and N_d are the numbers of synonymous and nonsynonymous differences between two sequences, and S and N are the numbers of synonymous and nonsynonymous sites, respectively.

A relationship (correlation) between sequence distances to the consensus and their sampling years was examined by using the linear regression analysis. The *t* test was used to compare groups. All statistical calculations were done by using SPSS/PC + software version 5.0 (SPSS Inc., Chicago, Ill.).

Model of HIV-1 evolution in sequence space. Eigen describes a viral quasispecies a swarm in multidimensional sequence space, whose center is the consensus sequence of the quasispecies (4–7). Each individual sequence occupies a unique position within the swarm that can be characterized by a nucleotide, synonymous, and nonsynonymous distance from the consensus (D, D_s, D_a, respectively [Fig. 1]). The mean radius of a quasispecies at a certain time point can be calculated as the mean distance between individual sequences obtained at this time point to the consensus.

An individual host is considered to be infected by an HIV-1 subtype B strain that has a nucleotide distance of D_0 from the subtype B consensus (Fig. 1). This initial HIV-1 sequence will change in time and after 5 years of infection will have a distance of D_5 from the consensus sequence and a distance of D from the seroconversion sequence. Thus, intrahost evolution over 5 years can be represented as a vector between two points: the sequence positions at seroconversion and after 5 years. The length of this vector, which represents sequence evolution, can be calculated as the genetic distance between the sequences obtained at seroconversion and 5 years later. To estimate the direction of this evolutionary vector, the sequence comparison can be expanded to include a third point, the subtype consensus sequence. Intrahost evolution of a sequence can thus be studied relative to the subtype consensus, and its evolutionary direction can be deduced as being toward or away from the consensus sequence. The difference between D_5 and D_0 is the sequence evolution relative to the subtype consensus sequence, $dD_{05} = D_5 - D_0$. Note, that dD_{05} is negative when a sequence evolves toward the consensus.

RESULTS

Interhost HIV-1 subtype B evolution is bounded by a set nonsynonymous distance from the subtype B consensus. Since most complete data sets are available for HIV-1 subtype B, we first analyzed the genetic distances of the V3 sequences to the subtype B consensus (sequence swarm radius) in an infected population, year by year. Our population comprised 124 individuals from the Amsterdam cohorts infected in the period between 1985 and 1992, from whom sequences were obtained at the time of seroconversion.

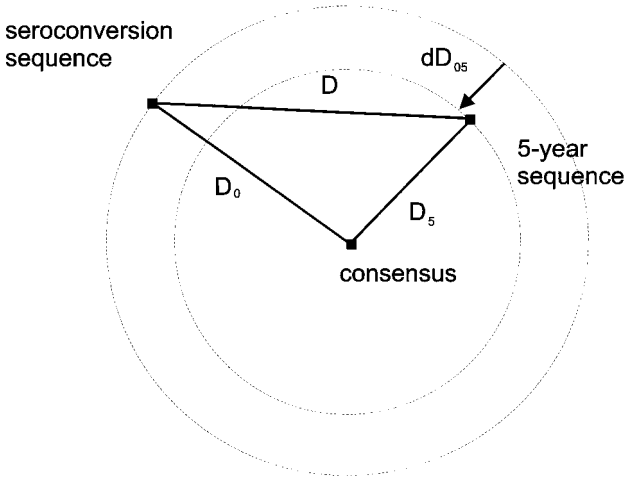

FIG. 1. HIV-1 evolution in sequence space. D_0, nucleotide distance of an individual HIV-1 sequence at seroconversion to the subtype consensus; D_5, nucleotide distance of the HIV-1 sequence obtained from the same individual after 5 years to the subtype consensus; D, nucleotide distance between HIV-1 sequences obtained at seroconversion and 5 years later; $dD_{05} = D_5 - D_0$, intrahost evolution over 5 years relative to the consensus.

Overall, the nucleotide distance between V3 sequences and the subtype B consensus increased significantly over time (Fig. 2). The mean nucleotide distance to the consensus was 0.052 ± 0.019 (p-distance ± standard deviation) in 1985 ($n = 27$) and 0.071 ± 0.022 in 1991 to 1992 ($n = 17$) ($P < 0.01$). Separate analysis of synonymous and nonsynonymous substitutions revealed that the increasing distance to the consensus was influenced far more by the increasing synonymous distance (1985; 0.040 ± 0.021; 1991 to 1992, 0.081 ± 0.036 [$P < 0.001$]) than by the nonsynonymous distance, which did not change significantly during the progression of the epidemic (1985, 0.055 ± 0.023; 1991 to 1992, 0.068 ± 0.026 [$P > 0.1$]). This finding points to selective pressures against nonsynonymous evolution away from the consensus, especially since most nucleotide substitutions in the V3 domain are nonsynonymous. Thus, the increasing nucleotide distance to the consensus over time has to be considered the result of accumulation of synonymous substitutions. Evolution of HIV-1 subtype B over 8 years did not show a significant increase of the nonsynonymous radius of the subtype B sequence swarm.

To exclude geographic bias in our results, we subsequently analyzed the evolution of the V3 domain in the US HIV-1 subtype B strains. For our analysis, we collected the HIV-1 sequences from the Los Alamos database (31) obtained in the early U.S. epidemic (1983 to 1986), i.e., ASP1, ALA1, CDC42, JRCSF, NY5CG, MN, RF, RJS, SBC, SC, SF33, for which sampling dates are available. We also used the sequences from the Florida dentist case (1990 to 1991) (33), a set of the early U.S. HIV-1 sequences (1983) (34a), and the sequences from the Baltimore intravenous drug user cohort (1989 to 1994) (19, 25). In total, 81 U.S. sequences obtained in a period between 1983 and 1994 were used.

The patterns of interhost evolution of HIV-1 in the United States corresponded to those observed for the Amsterdam cohorts (Fig. 3). The synonymous distance to the subtype B

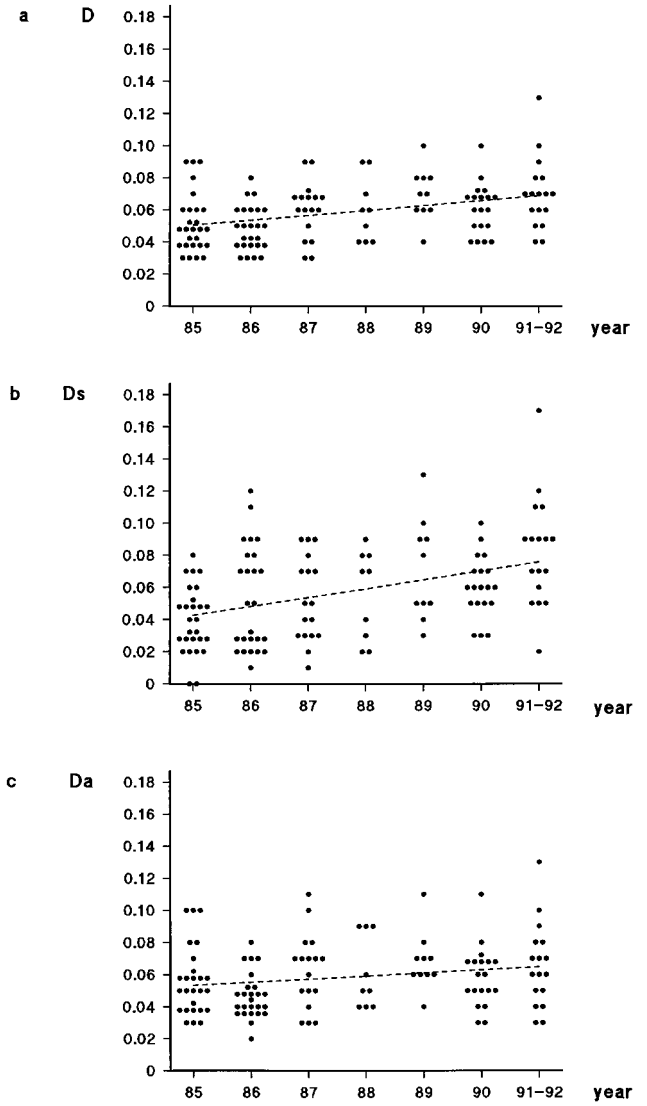

FIG. 2. Nucleotide (a), synonymous (b), and nonsynonymous (c) distances between the V3 sequences and the subtype B consensus in relation to seroconversion year in Amsterdam cohorts. Regression lines are shown. Statistics: (a) $r = 0.351$, $r^2 = 0.123$, $P < 0.001$; (b) $r = 0.402$, $r^2 = 0.162$, $P < 0.001$; (c) $r = 0.195$, $r^2 = 0.038$, $P = 0.03$.

consensus increased significantly (1983 [$n = 14$], 0.046 ± 0.024; 1991 to 1994 [$n = 14$], 0.071 ± 0.025 [$P = 0.01$]). The mean nonsynonymous distance to the subtype B consensus did not change significantly (1983, 0.076 ± 0.037; 1991 to 1994; 0.072 ± 0.021 [$P > 0.1$]).

It has previously been shown that the HIV-1 subtype B strains in Thailand differ from those in the United States and Europe, forming a separate phylogenetic cluster (12, 31, 34). To determine whether this genetic distinction is associated with the expansion of the subtype B sequence swarm, we analyzed the synonymous and nonsynonymous distance of the Thai sequences (12, 34) to the subtype B consensus. Despite their distinct genetic makeup, the U.S. and Thai HIV-1 subpopulations showed no difference in synonymous or nonsynonymous distances to the subtype B consensus over the same period. The mean nonsynonymous distance of the 29 Thai sequences (1991 to 1992) to the subtype B consensus was 0.087 ± 0.034 ($P > 0.1$ for the comparison with the 14 U.S. sequences obtained in 1991 to 1994). The U.S./European and Thai HIV-1 subtype B sequences thus appear to occupy separate locales in sequence space that are equidistant from the consensus. This was also the case when we analyzed the HIV-1 sequences obtained from Brazil (2, 28), Russia (20), Ivory Coast, and Gabon (31) (data not shown).

To locate the center of HIV-1 subtype B quasispecies in sequence space, we used the Los Alamos subtype B consensus, which could be biased by its overrepresentation of U.S. and European sequences. To examine whether this could influence our results, we also used an alternative subtype B consensus sequence, calculated as an arithmetic mean of a set of 100 subtype B sequences equally representing all major geographic areas of the HIV-1 subtype B epidemic (United States, Europe, Brazil, and Thailand). This alternative subtype B consensus was similar to the Los Alamos subtype B consensus (nucleotide p-distance, 0.022; synonymous p-distance, 0.000; nonsynonymous p-distance, 0.027). When we analyzed our sequence sets against the alternative subtype B consensus, similar results were observed (data not shown), indicating that both consensus sequences are close to the actual center of HIV-1 subtype B quasispecies in sequence space. Our data are in accord with the star-like topology of the phylogenetic tree of HIV-1 subtypes as well as with previous observations of the stability of the consensus sequence over time, which has been demonstrated by comparing the Dutch consensus sequences year by year over the period from 1982 to 1994 (15).

Sequence space of HIV-1 subtypes A, C, and D. To study the sizes of HIV subtype A, C, and D sequence swarms, we analyzed the genetic distances of 122 subtype A (1976 to 1995), 126 subtype

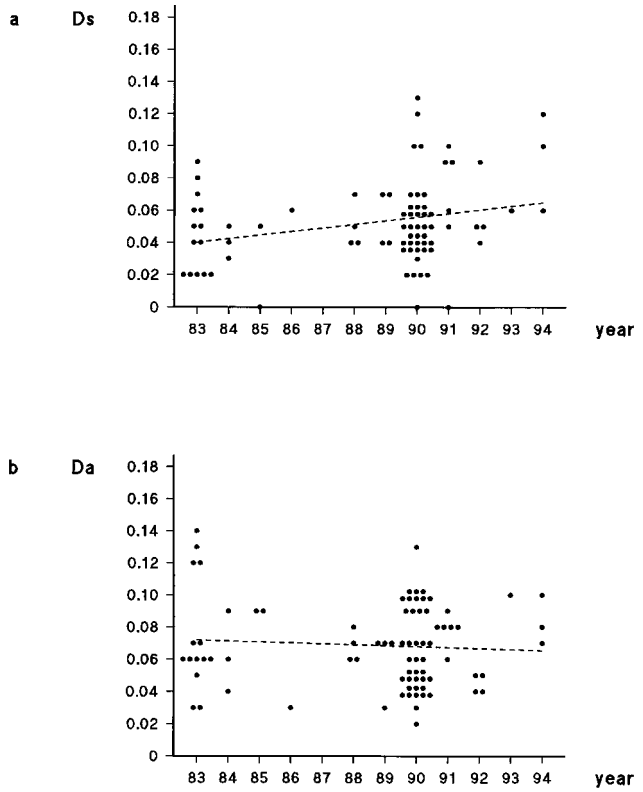

FIG. 3. Synonymous (a) and nonsynonymous (b) distances between the US V3 sequences and the subtype B consensus in relation to sampling year. Regression lines are shown. Statistics: (a) $r = 0.269$, $r^2 = 0.072$, $P = 0.02$; (b) $r = -0.071$, $r^2 = 0.005$, $P > 0.1$.

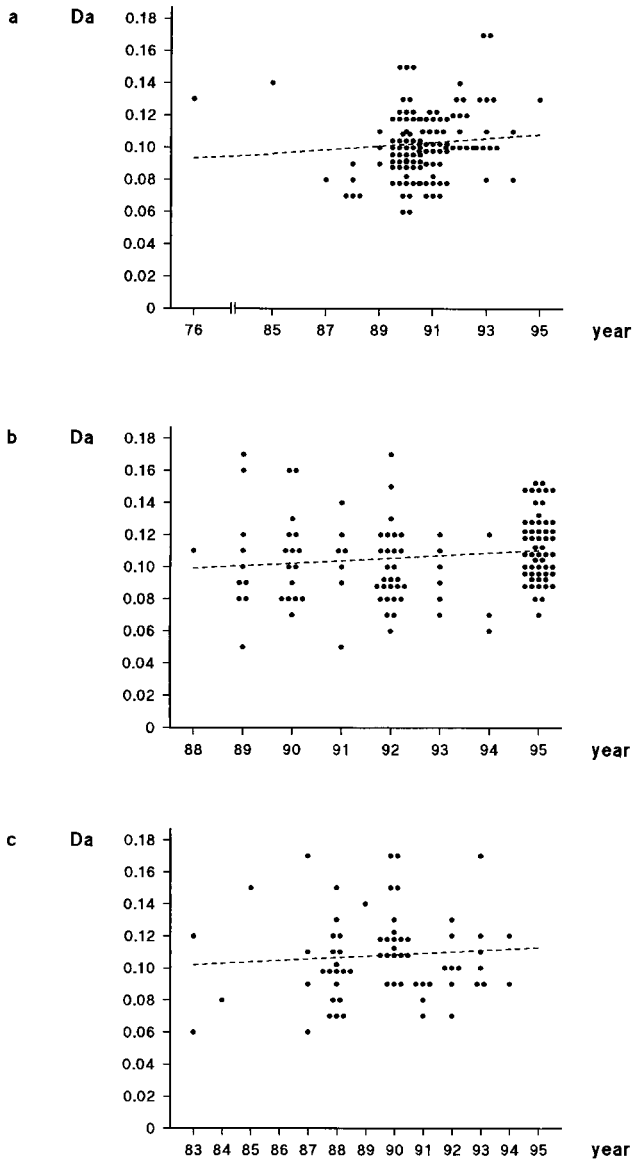

FIG. 4. Nonsynonymous distances of the HIV-1 subtype A (a), C (b), and D (c) sequences to the consensus sequence of the same HIV-1 subtype in relation to sampling year. Regression lines are shown. Statistics: (a) $r = 0.110$, $r^2 = 0.012$, $P > 0.1$; (b) $r = 0.140$, $r^2 = 0.019$, $P > 0.1$; (c) $r = 0.078$, $r^2 = 0.006$, $P > 0.1$.

C (1988 to 1995), and 67 subtype D (1983 to 1994) V3 sequences (references 2, 3, 14, 18, 20, 23, 29 and 31 and kindly provided by Almaz Abebe prior to publication) to the consensus of a respective subtype in relation to sequence sampling year.

Similar to our findings for HIV-1 subtype B, the nonsynonymous distances of the individual subtype A, C, and D sequences to the consensus of a respective subtype did not increase significantly over time (Fig. 4). The mean nonsynonymous distances to the consensus of earlier sequences were similar to those of the sequences obtained later: subtype A, 0.100 ± 0.021 (1976 to 1990) and 0.106 ± 0.021 (1991 to 1995); subtype C, 0.105 ± 0.035 (1988 to 1989) and 0.113 ± 0.021 (1995); subtype D, 0.105 ± 0.038 (1983 to 1987) and 0.111 ± 0.025 (1993 to 1994) ($P > 0.1$ for all comparisons).

The direction of the HIV-1 intrahost evolution, but not its extent, is associated with sequence distance to the subtype B consensus. If interhost V3 nonsynonymous variation is limited, the same must be true for the intrahost evolution. If transmission bottlenecks play only a random and neutral selective role, the constraints in nonsynonymous evolution at the population level can be explained by limitations in evolution within individuals during periods between transmissions. To test these presumptions, we studied the intrahost evolution of the V3 regions in 55 infected individuals over their first 5 years of infection, the period in which most infected individuals are healthy and can be assumed to cause most of the new infections (Fig. 5 and 6).

Our analyses revealed no significant association between the numbers of nucleotide substitutions per site accumulated by 55 HIV-1 V3 sequences over 5 years (D) and the nucleotide distances between those sequences and the subtype B consensus at seroconversion (D_0) (Fig. 5). Neither the numbers of synonymous substitutions per synonymous site (D_s) nor the numbers of nonsynonymous substitutions per nonsynonymous site (D_a) accumulated over 5 years were significantly associated with the synonymous and nonsynonymous distances to the subtype B consensus at seroconversion (D_{0s} and D_{0a}, respectively). Thus, the evolution rate of the V3 region appears to be independent of sequence distance from the consensus. Moreover, when we studied the ratio of synonymous to nonsynonymous substitutions accumulated over 5 years (D_s/D_a) in relation to the sequence distances to the consensus at seroconversion (D_0, D_{0s}, D_{0a}), we found no association (data not shown). This confirms our previous finding that host selection pressure results in continuous evolution, irrespective of the distance between the sequence and the subtype B consensus (24).

However, when we analyzed the direction of the evolution of the V3 region over 5 years relative to the consensus sequence, we found a significant negative correlation between the direction of the evolution of the V3 sequences and their distance from the subtype B consensus at seroconversion (Fig. 6). The largest nucleotide distances to the consensus at seroconversion were associated with restriction of further sequence evolution away from the consensus.

DISCUSSION

A high level of genetic variation is observed in HIV-1 genomes both at the population level and within individual hosts. The HIV-1 strains can be separated into subtypes (A to J) based on phylogenetic analysis of the envelope sequences irrespective of whether the complete gp120 or the V3 region is analyzed (31, 36). For the HIV-1 gp120 V3 region, the intrahost evolution rate has been estimated to be about 1% per year, with the nonsynonymous evolution rate being higher than the synonymous evolution rate (24, 37). Envelope sequence heterogeneity has been demonstrated to increase in an infected population over time (15, 16). Given no evolutionary constraints, this high evolution rate could be expected to lead to a continuous expansion of the subtype swarms in sequence space and, eventually, predicts the merger of HIV-1 subtypes.

To examine this scenario, we analyzed the population-wide evolution of the V3 region of the HIV-1 subtypes A, B, C, and D over the course of the global HIV-1 epidemic. We observed that the high evolution rate of HIV-1 did not result in a significant increase of the intrasubtype nonsynonymous variation (heterogeneity) of HIV-1 sequences, in contrast to the synonymous variation, which increased steadily over time. The nonsynonymous distances to the subtype consensus of the sequences obtained earlier in the course of the global AIDS epidemic are similar to those of the sequences obtained more than 10 years later. These observations indicate evolutionary constraints against nonsynonymous evolution away from the

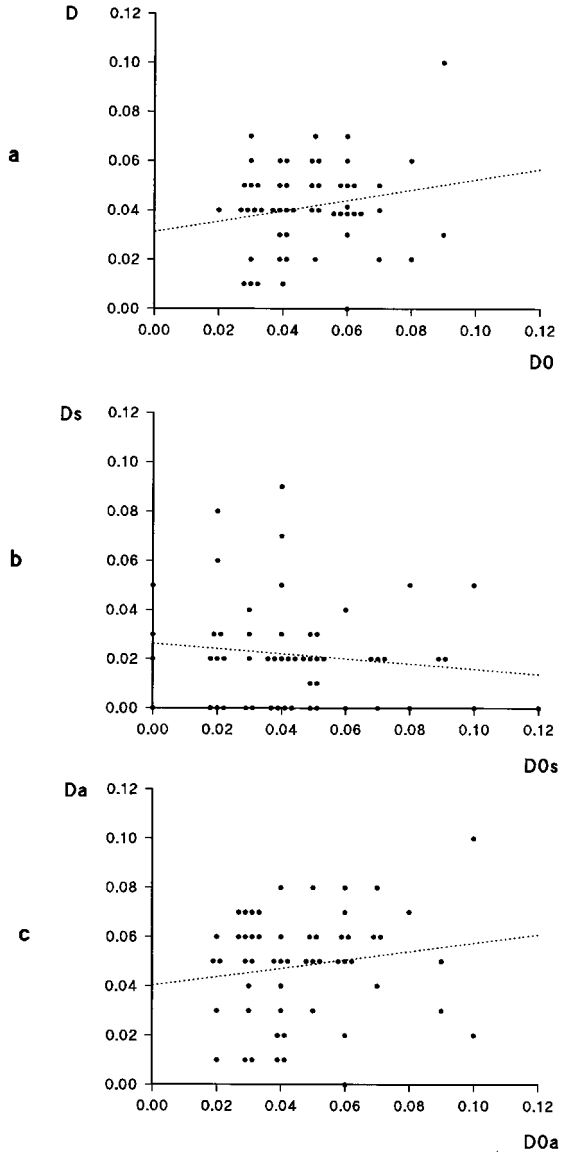

FIG. 5. The rate of the HIV-1 V3 nucleotide (a), synonymous (b), and nonsynonymous (c) intrahost evolution over 5 years is not associated with sequence distance to the subtype B consensus at seroconversion (nucleotide, synonymous, and nonsynonymous distances, respectively). Statistics: (a) $r = 0.192$, $r^2 = 0.037$, $P > 0.1$; (b) $r = -0.128$, $r^2 = 0.016$, $P > 0.1$; (c) $r = 0.160$, $r^2 = 0.025$, $P > 0.1$.

subtype consensus and suggest that a HIV-1 subtype swarm is a target for purifying selection.

The V3 domain is subjected to a strong immune pressure in the host, which drives the intrahost HIV-1 evolution (24). Most of the nucleotide substitutions in this region are nonsynonymous. However, the nonsynonymous distance between individual sequences and the subtype consensus is limited, an observation that is most probably related to the functional importance of this domain. A large nonsynonymous distance of an individual sequence to the subtype consensus could well be associated with a functional deterioration and decrease of viral fitness (adaptation). Continuous virus evolution irrespective of the sequence distance to the consensus (Fig. 5) combined with limited nonsynonymous distance to the subtype consensus (Fig. 6) can be explained as a result of multiple hits at one site, including back substitutions. Operating simultaneously, the evolutionary forces of immune pressure and fitness limitations cause fluctuations of individual HIV-1 V3 sequences within a limited nonsynonymous sequence space around the stationary subtype consensus sequence.

It is important to consider this circumstance in the analysis of the ratio of synonymous to nonsynonymous substitutions (D_s/D_a), which is often used to study the evolutionary trends of a certain gene (purifying versus positive selection). For the intrahost, relatively short-term evolution of the V3 region, the D_s/D_a ratio is about 0.4 (24), while for the intrasubtype, relatively long-term evolution, it is generally above 1 and is even higher for the comparisons of HIV-1 subtypes with HIV-1 group O sequences (13, 30). This phenomenon could be directly related to fluctuating evolution of the V3 sequences within a limited sequence space. It is known that the increasing number of nucleotide substitutions between two sequences

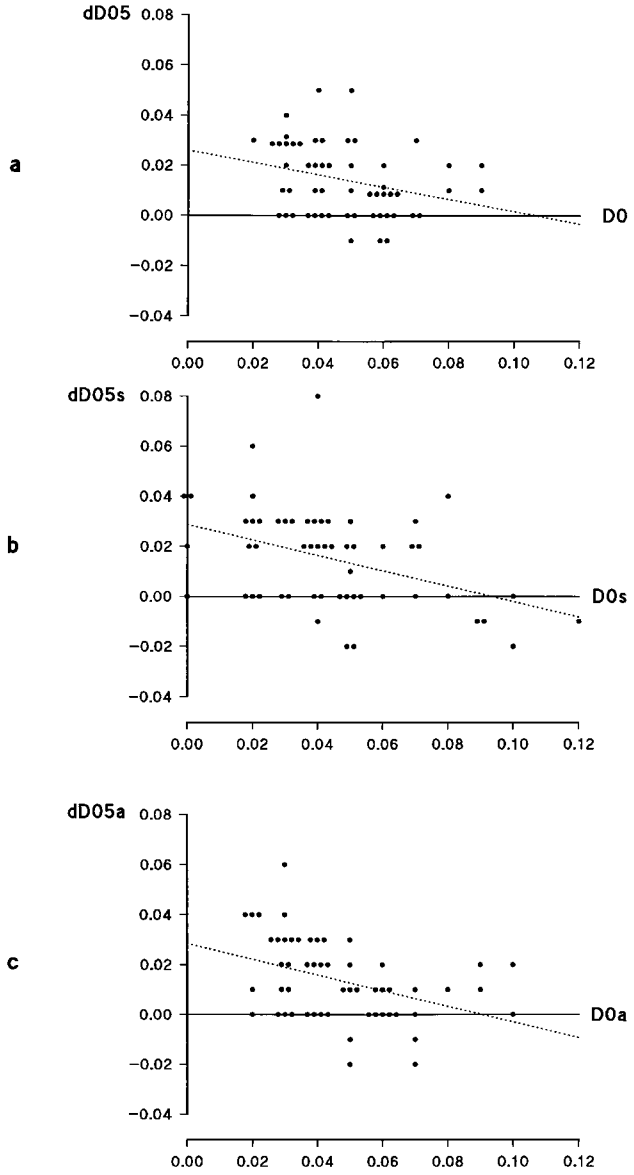

FIG. 6. The direction of the intrahost evolution of the V3 sequences relative to the subtype B consensus is associated with their nucleotide (a), synonymous (b), and nonsynonymous (c) distances to the subtype B consensus at seroconversion. Statistics: (a) $r = -0.278$, $r^2 = 0.077$, $P = 0.04$; (b) $r = -0.405$, $r^2 = 0.164$, $P < 0.01$; (c) $r = -0.400$, $r^2 = 0.157$, $P < 0.01$.

obstructs the precise measurement of the evolutionary distances between these sequences, making it impossible after a certain saturation threshold, when the nucleotide distance is, say, above 50% ("mechanical" saturation). Limited nonsynonymous sequence space creates circumstances whereby the impact of the saturation processes may be manifest at much lower nonsynonymous distances ("functional" saturation). If so, the changes in the D_s/D_a ratio related to the evolutionary scale provide additional support for the limited sequence space of HIV-1 subtypes. When we compared the HIV-1 subtype B V3 sequences obtained early versus late in the epidemic with the subtype B consensus, we observed an increase of the D_s/D_a ratios over time (for instance, in the Amsterdam epidemic, 1985, 0.73; 1991 to 1992, 1.19; in the U.S. epidemic, 1983, 0.61; 1991 to 1994, 0.99 [see Results]). In our opinion, this increase does not indicate declining selection pressure in the V3 region over the observation period but, rather, is an effect of the functional saturation of nonsynonymous substitutions.

Compared to the nonsynonymous substitutions, the synonymous ones are generally considered to be subjected to a lower selection pressure. We found that the mean synonymous distance of the individual HIV-1 sequences to the subtype consensus is increasing in the course of the AIDS epidemic (Fig. 2 and 3).

Our data show that the quasispecies concept can be applied to an HIV-1 genetic subtype. Unlike the rapidly changing HIV-1 quasispecies within an infected individual, the entire quasispecies of an HIV-1 subtype appeared to be stabilized and able to maintain its collective identity over time in the V3 region. Such stability suggests that HIV-1 subtypes are optimally adapted to the present environment. This is shown schematically in Fig. 7. We suggest that after the introduction(s) of pro-HIV-1 (simian immunodeficiency virus) into the human population, the adaptation of this virus to the new host went on in several evolutionary directions (irrespective of whether one or multiple cross-species transmissions took place). In this model, HIV-1 subtypes represent separate genetic niches in which virus fitness can be maintained most, although not necessarily equally, effectively. We propose that the distribution of individual sequences within one subtype is uneven, i.e., that the fitness landscape of a HIV-1 subtype is not smooth (as an example, HIV-1 B′ is shown in Fig. 7). Moreover, the observation that certain amino acid positions within the V3 region are less variable and even conserved, most probably because of their functional importance, indicates that certain minimal requirements for the maintenance of the tertiary structure of the V3 domain have to be met. Our model does not prohibit the evolution of an individual sequence far away from the subtype consensus, even from one subtype to another, but it predicts that such evolution could be associated with a decrease of virus fitness and makes it unlikely in the present environment. However, several factors could perturb the current balance in virus-host interactions. For instance, anti-V3 vaccines could potentially decrease the fitness of the viruses with the V3 regions that are homologous to the subtype consensus. Also, the changes in the genetic makeup of the other, more slowly evolving HIV-1 genomic regions could influence the current trends of the V3 evolution (8). In particular, antiviral therapy and HIV-1 drug-resistant mutants could contribute to this process.

Our observations indicate that the subtype consensus sequence could be considered to be a master sequence, i.e., the fittest sequence at the population level, although in particular HIV-1-infected individuals, the fittest sequence may differ from the consensus. When we analyzed the genetic distances between the subtype B consensus and the V3 sequences in participants of the Amsterdam cohorts who did or did not progress to AIDS within 5 years after seroconversion (24), we observed no significant difference between groups (data not shown), suggesting that HIV-1 sequence fluctuation within a limited sequence space around the consensus does not influence HIV-1 pathogenicity.

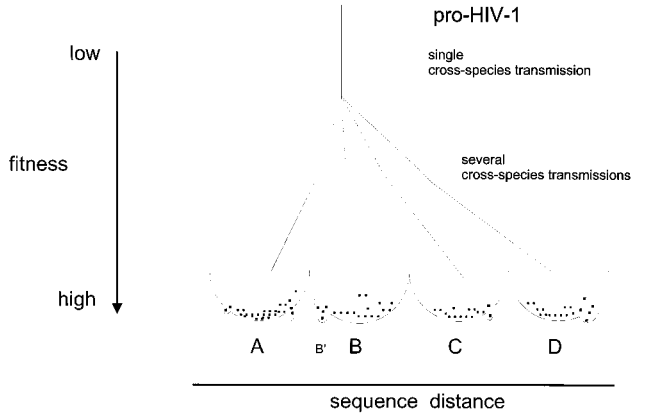

FIG. 7. Fitness landscape of HIV-1 subtypes. The model shows HIV-1 subtypes as separate genetic niches in which the virus phenotype can be realized most effectively (virus fitness is high). Although evolution of individual sequence (dots) far away from consensus, from one subtype to another, is not prohibited, it is predicted to be associated with a decrease of virus fitness. Note that in multidimensional sequence space, the genetic distance between any two HIV-1 subtypes is approximately equal. For more explanations, see Discussion.

Recombination between different HIV-1 subtypes is a significant source of HIV-1 variability (35). In our earlier studies, we provided the results of the phylogenetic analysis of the V3 sequences obtained from 55 individuals at seroconversion and 5 years later (15, 24). All late sequences (with single exceptions) clustered with the seroconversion sequences obtained from the same individual. These clusters were statistically supported by bootstrap analysis. Those individuals (homosexual men and intravenous drug users) have continued the AIDS risk-associated activity and may therefore be superinfected or coinfected by another HIV-1 subtype B strain. Intrasubtype recombination, if it occurs, would result in a loss of statistical support of the phylogenetic relations between early and late sequences from that particular individual. The strong statistical support we observed for the clusters of the V3 sequences from the same patient, together with previous observations of in vivo recombination between HIV-1 genes or the functional domains of the envelope gene rather than within them (1, 9), suggests that intrasubtype recombination within the V3 domain is not likely.

The extent of intrasubtype HIV-1 variability, when the nucleotide distances of certain sequences to the consensus can be as high as 15%, allows theoretically for subtype clouds to overlap (the nucleotide distance between the subtype consensus sequences is close to 25%). The fact that no V3 sequences equidistant to two HIV-1 subtypes were found may indicate low fitness or functional defects of HIV-1 variants in subtype-overlapping sequence space. This assumption is also supported by the observation that sequence convergence between HIV-1 subtypes is relatively rare within the V3 region (13, 30). Although intrasubtype recombination within the V3 domain seems to be not likely, one cannot exclude that certain intersubtype recombinations may result in increasing fitness of recombinant viruses. This intersubtype recombination within the V3 domain may lead to occupation of a new genetic niche and to emergence of a new genetic subtype.

In conclusion, our results suggest that the evolution of the V3 domain of HIV-1 subtypes A, B, C, and D is limited to an

area within a set nonsynonymous distance from the subtype consensus sequence. HIV-1 subtypes are thus phenotypic entities that are able to maintain their integrity and are unlikely to merge during progression of the AIDS epidemic.

ACKNOWLEDGMENTS

This study is funded in part by The Netherlands Council of Health Research (RGO; grant 94019) and The Netherlands Foundation for Preventive Medicine (grant 28-2371).

We thank Roel Coutinho, Anneke van der Hoek, David Vlahov, and Almaz Abebe for providing valuable data; Carla Kuiken for stimulating discussions; and Lucy Phillips for editing the manuscript.

REFERENCES

1. **Cornelissen, M., G. Kampinga, F. Zorgdrager, J. Goudsmit, and the UNAIDS Network for HIV Isolation and Characterization.** 1996. Human immunodeficiency virus type 1 subtypes defined by *env* show high frequency of recombinant *gag* genes. J. Virol. **70:**8209–8212.
2. **de Wolf, F., E. Hogervorst, J. Goudsmit, E.-M. Fenyö, H. Rübsamen-Waigmann, H. Holmes, B. Galvao-Castro, E. Karita, C. Wasi, S. D. K. Sempala, E. Baan, F. Zorgdrager, V. Lukashov, S. Osmanov, C. Kuiken, M. Cornelissen, and WHO Network on HIV Isolation and Characterization.** 1994. Syncytium-inducing and non-syncytium-inducing capacity of human immunodeficiency virus type 1 subtypes other than B: phenotypic and genotypic characteristics. AIDS Res. Hum. Retroviruses **10:**1387–1400.
3. **Dietrich, U., M. Grez, H. von Briesen, B. Panhans, M. Geisendorfer, H. Kuhnel, J. Maniar, G. Mahambre, W. B. Becker, M. L. B. Becker, and H. Rubsamen-Waigmann.** 1992. HIV-1 strains from India are highly divergent from prototypic African and US/European strains, but are linked to a South African isolate. AIDS **7:**23–27.
4. **Eigen, M.** 1971. Self-organization of matter and the evolution of biological macromolecules. Naturwissenschaften **58:**465–523.
5. **Eigen, M.** 1987. New concepts for dealing with the evolution of nucleic acids. Cold Spring Harbor Symp. Quant. Biol. **LII:**307–320.
6. **Eigen, M., J. McCaskill, and P. Schuster.** 1988. Molecular quasi-species. J. Phys. Chem. **92:**6881–6891.
7. **Eigen, M., R. Winkler-Oswatitsch, and A. Dress.** 1988. Statistical geometry in sequence space: a method of quantitative comparative sequence analysis. Proc. Natl. Acad. Sci. USA **85:**5913–5917.
8. **Fitch, W. M., and E. Markowitz.** 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochem. Genet. **4:**579–593.
9. **Groenink, M., A. C. Andeweg, R. A. M. Fouchier, S. Broersen, R. C. M. van der Jagt, H. Schuitemaker, R. E. Y. de Goede, M. L. Bosch, H. G. Huisman, and M. Tersmette.** 1992. Phenotype-associated *env* gene variation among eight related human immunodeficiency virus type 1 clones: evidence for in vivo recombination and determinants of cytotropism outside the V3 domain. J. Virol. **66:**6175–6180.
10. **Hahn, B. H., G. M. Shaw, M. E. Taylor, R. R. Redfield, P. D. Markham, S. Z. Salahuddin, F. Wong-Staal, R. C. Gallo, E. S. Parks, and W. P. Parks.** 1986. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. Science **232:**1548–1553.
11. **Holmes, E. C., L. Q. Zhang, P. Simmonds, C. A. Ludlam, and A. J. Leigh Brown.** 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. Proc. Natl. Acad. Sci. USA **89:**4835–4839.
12. **Kalish, M. L., C.-C. Luo, B. G. Weniger, K. Limpakarnjanarat, N. Young, C.-Y. Ou, and G. Schochetman.** 1994. Early HIV type 1 strains in Thailand were not responsible for the current epidemic. AIDS Res. Hum. Retroviruses **10:**1573–1575.
13. **Korber, B. T. M., K. MacInnes, R. F. Smith, and G. Myers.** 1994. Mutational trends in V3 loop protein sequences observed in different genetic lineages of human immunodeficiency virus type 1. J. Virol. **68:**6730–6744.
14. **Kostrikis, L. G., E. Bagdades, Y. Cao, L. Zhang, D. Dimitriou, and D. D. Ho.** 1995. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. J. Virol. **69:**6122–6130.
15. **Kuiken, C. L., V. V. Lukashov, E. Baan, J. Dekker, J. A. M. Leunissen, and J. Goudsmit.** 1996. Evidence for limited within-person evolution of the V3 domain of the HIV-1 envelope in the Amsterdam population. AIDS **10:**31–37.
16. **Kuiken, C. L., G. Zwart, E. Baan, R. A. Coutinho, J. A. R. van den Hoek, and J. Goudsmit.** 1993. Increasing antigenic and genetic diversity of the V3 variable domain of the human immunodeficiency virus envelope protein in the course of the AIDS epidemic. Proc. Natl. Acad. Sci. USA **90:**9061–9065.
17. **Kumar, S., K. Tamura, and M. Nei.** 1993. Molecular evolutionary genetics analysis (MEGA). Version 1.01. Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park.
18. **Louwagie, J., W. Janssens, J. Mascola, L. Heyndrickx, P. Hegerich, G. van der Groen, F. E. McCutchan, and D. S. Burke.** 1995. Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. J. Virol. **69:**263–271.
19. **Lukashov, V. V.** 1997. Unpublished data.
20. **Lukashov, V. V., M. T. E. Cornelissen, J. Goudsmit, M. N. Papuashvilli, P. G. Rytik, R. M. Khaitov, E. V. Karamov, and F. de Wolf.** 1995. Simultaneous introduction of distinct HIV-1 subtypes into different risk groups in Russia, Byelorussia and Lithuania. AIDS **9:**435–439.
21. **Lukashov, V. V., and J. Goudsmit.** 1995. Increasing genotypic and phenotypic selection from the original genomic RNA populations of HIV-1 strains LAI and MN (NM) by peripheral blood mononuclear cell culture, B-cell-line propagation and T-cell-line adaptation. AIDS **9:**1307–1311.
22. **Lukashov, V. V., and J. Goudsmit.** 1997. Founder virus population related to route of virus transmission: a determinant of intrahost human immunodeficiency virus type 1 evolution? J. Virol. **71:**2023–2030.
23. **Lukashov, V. V., C. L. Kuiken, K. Boer, and J. Goudsmit.** 1996. HIV type 1 subtypes in The Netherlands circulating among women originating from AIDS endemic regions. AIDS Res. Hum. Retroviruses **12:**951–953.
24. **Lukashov, V. V., C. L. Kuiken, and J. Goudsmit.** 1995. Intrahost human immunodeficiency virus type 1 evolution is related to length of the immunocompetent period. J. Virol. **69:**6911–6916.
25. **Lukashov, V. V., C. L. Kuiken, D. Vlahov, R. A. Coutinho, and J. Goudsmit.** 1996. Evidence for HIV type 1 strains of U.S. intravenous drug users as founders of AIDS epidemic among intravenous drug users in Northern Europe. AIDS Res. Hum. Retroviruses **12:**1179–1183.
26. **McNearney, T., Z. Hornickova, R. Markham, A. Birdwell, M. Arens, A. Saah, and L. Ratner.** 1992. Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. Proc. Natl. Acad. Sci. USA **89:**10247–10251.
27. **Meyerhans, A., R. Cheynier, J. Albert, M. Seth, S. Kwok, J. Sninsky, L. Morfeldt-Månson, B. Åsjö, and S. Wain-Hobson.** 1989. Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. Cell **58:**901–910.
28. **Morgado, M. G., E. C. Sabino, E. G. Shpaer, V. Bongertz, Brigido, M. D. C. Guimaraes, E. A. Castilho, B. Galvao-Castro, J. I. Mullins, R. M. Hendry, and A. Mayer.** 1994. V3 region polymorphism in HIV-1 from Brazil: prevalence of subtype B strains divergent from North American/European prototype and detection of subtype F. AIDS Res. Hum. Retroviruses **10:**569–576.
29. **Murphy, E., B. Korber, M. Georges-Courbot, B. You, A. Pinter, D. Cook, M. Kieny, A. Georges, C. Mathiot, F. Barre-Sinoussi, and M. Girard.** 1993. Diversity of V3 region sequences of human immunodeficiency viruses type 1 from the Central African Republic. AIDS Res. Hum. Retroviruses **9:**997–1006.
30. **Myers, G.** 1994. HIV: between past and future. AIDS Res. Hum. Retroviruses **10:**1317–1324.
31. **Myers, G., B. Korber, B. H. Hahn, K. Jeang, J. W. Mellors, F. E. McCutchan, L. E. Henderson, and G. N. Pavlakis.** 1995. Human retroviruses and AIDS: a compilation and analysis of nucleic acid and amino acid sequences. Theoretical Biology and Biophysics, Group T-10, Los Alamos National Laboratory, Los Alamos, N.M.
32. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
33. **Ou, C.-Y., C. A. Ciesielski, G. Myers, C. I. Bandea, C.-C. Luo, B. T. M. Korber, J. I. Mullins, G. Schochetman, R. L. Berkelman, A. N. Economou, J. J. Witte, L. J. Furman, G. A. Satten, K. A. MacInnes, J. W. Curran, H. W. Jaffe, Laboratory Investigation Group, and Epidemiologic Investigation Group.** 1992. Molecular epidemiology of HIV transmission in a dental practice. Science **256:**1165–1171.
34. **Ou, C.-Y., Y. Takebe, B. G. Weniger, C. C. Luo, M. L. Kalish, W. Auwanit, S. Yamazaki, H. D. Gayle, N. L. Young, and G. Schochetman.** 1993. Independent introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand. Lancet **341:**1171–1174.

34a. **Reitz, M. S., M. Popovic, W. C. Saxinger, L. Hall, E. Read-Connole, P. Markham, and R. C. Gallo.** 1994. Partial envelope sequences from some of the earliest isolates of HIV-1. AIDS Res. Hum. Retroviruses **10:**621–623.

35. **Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn.** 1995. Recombination in HIV-1. Nature **374:**124–126.
36. **WHO Network on HIV Isolation and Characterization.** 1994. HIV type 1 variation in World Health Organization-sponsored vaccine evaluation sites: genetic screening, sequence analysis, and preliminary biological characterization of selected viral strains. AIDS Res. Hum. Retroviruses **10:**1327–1344.
37. **Wolfs, T. F. W., J. J. de Jong, H. Van den Berg, J. M. G. H. Tijnagel, W. J. A. Krone, and J. Goudsmit.** 1990. Evolution of sequences encoding the principal neutralization epitope of HIV-1 is host-dependent, rapid and continuous. Proc. Natl. Acad. Sci. USA **87:**9938–9942.