Structured-tree topology and adaptive evolution of the simian immunodeficiency virus SIVsm envelope during serial passage in rhesus macaques according to likelihood mapping and quartet puzzling

Valli, P.J.S.; Goudsmit, J.

Published in:
Journal of Virology

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Zootonic transmission of some lentiviruses causes AIDS. The two principal etiologic agents of AIDS in humans are the human immunodeficiency viruses types 1 and 2 (HIV-1 and HIV-2), whose probable ancestors are found in two species of feral African primates (simian immunodeficiency virus SIVcpz and SIVsm) (17). HIV-2 appears to be descended from the SIVsm lineage (Fig. 1), as shown not only by sequence homology but also by the geographically localized infections (27) of humans and sooty mangabeys (Cercocebus atys) with this virus. However, the issue of HIV origins remains open to question (33a). The SIV strains do not induce disease in their natural hosts. The relation between pathogenesis and virus strain is not well understood but appears to be both host and pathogen related. When Asian macaques are experimentally infected with SIVsm, SIVagm, SIVmac, or a related molecular clone, the disease course mimics that seen in human AIDS (41) and thus provides a useful model for vaccine testing.

Although the length of the asymptomatic period of SIV infection varies greatly among individual Asian primates, the numerous SIVsm-derived viruses show a wide spectrum of pathogenicity or virulence. The most pathogenic SIVsm strain is the SIVsmPh14 (14, 20, 37), with an average time to death of 14 days. SIVsmH4 and SIVmac239 are both pathogenic during experimental infection but have a mean time to death of 1 year (19, 21–24). The related clones SIVsmm9 and SIV1A11 are relatively nonpathogenic, with times to death of several years. The length of the asymptomatic period during experimental infection with the many SIVsm-derived clones (11, 25, 28, 31, 37, 41, 42, 46) is presumably virus strain dependent, as shown by repeated infections with the same clones.

Viral variation is known to occur within a single cycle of intracellular replication (40), and the progeny virus from an infected cell may vary genetically from the parental virus. This genetic flexibility enables the rapid development of new mutants (3) by means of subtle changes in the env surface glycoprotein (3, 9, 12). The env genes of HIV-2 and SIVsm have the same functional layout, with five variable and constant domains and discrete (over 98% nonsynonymous [35, 38, 55]) nucleotide changes within the variable regions (V1, V2, V3, V4, and V5). The high nonsynonymous-substitution rate supports positive selection for nonfounder or adapted mutant virus by outgrowth of these new variants from the circulating virus population of founder or primary inoculum genotypes (14, 33, 47). Thought to be driven by the antigenic selection (14, 47) of new variants, constant evolution of new genotypes and the coexistence of distinct strains (54) during infection is believed to be sequence variation dependent, due to the number of different genomic variants that exist at any one time.

The goal of this study was to examine viral sequence variation during adaptation of the V1 to V4 regions of the env gene of SIVsm in the serum of immunodeficient rhesus macaques that were experimentally infected. The experiment was carried out by serial population passaging of a known SIVsm strain (SIVsmB670), accentuating the selection for adapted mutants in vivo (10), to elucidate the changes in the env gene that influence the marked increases in pathogenicity and virulence of SIVsm infection in the rhesus macaque model. Phylogenetic inference methods have been used to study the origins of HIV (33a, 34, 36) and the epidemiology of HIV infections (17) and to determine definitive transmission patterns (27). The use of phylogeny reconstruction methods with a known infection his-
tion of the phylogeny was tree-like. Construction of trees by quartet puzzling (QP) (48–50) produced a phylogeny that showed the adaptive process to be highly structured, since the divergence increases with passageing. The ordered divergence from the primary inoculum occurred in the same time, at the same rate, as the shortening of the asymptomatic period or the increase in pathogenicity of the passed SIVsm strain. The adaptation of SIVsm to Asian rhesus macaques is thus a structured, quickly occurring process in which the env gene evolves in a tree-like manner.

**MATERIALS AND METHODS**

**Virus.** Delta B670 (2, 5–7, 32, 35, 55) is an SIV strain that was originally discovered in a sooty mangabey presenting with a cutaneous lepromatous lesion. When tissue from this animal was used to inoculate rhesus macaques, it resulted in a syndrome that closely reproduced the effects of AIDS in humans (55). The virus isolate was proven to be type D negative and was extensively characterized (2, 5, 55).

**Passage.** The passaging was carried out with seven age-matched Asian rhesus macaques, all 2 years old. Their experimental infection with an SIVsm strain was performed in six consecutive steps. The first monkey was infected intravenously with $5 \times 10^3$ infectious doses of SIV Delta B670. The next five monkeys were intramuscularly inoculated in a serial fashion with $2 \times 10^3$ infectious doses of peripheral blood mononuclear cells (PBMC) taken at the asymptomatic stage of disease from the preceding rhesus macaque. For passage 5, PBMC of the passage 5 monkey (P5) were used to infect two monkeys (P6A and P6B) because little blood was available from monkey P5 due to its rapid progression to AIDS and sudden death. The time to death postinfection (tdpi) and moment of sampling for the reverse transcriptase PCR (RT-PCR) were as follows: P1, tdpi = 18 months, RT-PCR at 18 months; P2, tdpi = 12 months, RT-PCR at 3 and 7 months; P3, tdpi = 9 months, RT-PCR at 2 and 4.1 months; P4, tdpi = 4 months, RT-PCR at 1 and 1.8 months; P5, tdpi = 2 months, RT-PCR not done since no sample was available; P6A, tdpi = 2 months, RT-PCR at 2 months; P6B, tdpi = 2 months, RT-PCR at 2 months. The animals were euthanized upon evidence of undue discomfort. PBMC used for serial passage were not cultured or cocultivated with any other cells.

Viral RNA was harvested with silica in the presence of a chaotropic agent (8) from the serum of the infected macaques for use as a template in the RT-PCR. Viral RNA was isolated from 20 m of sera and resuspended in 20 m of RNasin in H_2O (1 U/mL) and used in a reverse transcription reaction (5 m of viral RNA, 250 n of each deoxynucleoside triphosphate, 2 ng of 3 RT-PCR primer [SIV4Not1: TTATATGCGGCCGCCTACTTTGTGCCACGTGTTG] per m, 2.5 mM MgCl_2, 1 U of RNasin [Promega] per m, 10 m of Super Script I [Gibco-BRL], and 1 X reaction buffer [45] in a 20-mL volume). The components were assembled at 37°C and incubated at that temperature for 90 min. The PCR mixture consisted of 250 m each deoxynucleoside triphosphate, 2 ng of 3 RT-PCR primer [SIV4Not1: TTATATGCGGCCGCCTACTTTGTGCCACGTGTTG] per m, 2.5 mM MgCl_2, 1 U of RNasin [Promega] per m, 10 m of Super Script I [Gibco-BRL], and 1 X reaction buffer [45] in a 20-mL volume). The components were assembled at 37°C and incubated at that temperature for 90 min. The PCR mixture was overlaid with paraffin, heated to 95°C for 5 min, and subjected to 35 cycles of 1 min at 95°C, 1 min at 55°C, and 1 min at 72°C, ending with 10 min at 72°C in a Perkin-Elmer Cetus DNA thermocycler. The reverse transcription reaction and PCR were done in duplicate for each sample to prevent mispriming and to ensure the fidelity of the PCR. The virus genotypes sampled. After PCR, the samples were combined and size selected on 0.8% agarose gels. The 1,151-bp band was excised and isolated from the gel slice, digested with NotI and HindIII, and again isolated from the agarose gel by silica particle isolation. Size-selected, digested, purified RT-PCR product was ligated overnight into plasmid pSP64 (Promega) containing a NotI site. Ligated products were electroporated into electrocompetent Escherichia coli C600, and double-stranded plasmid DNA for sequencing was isolated on Quagen columns. Five clones per sample time point were sequenced, except for the primary inoculum (P1), for which 10 clones were sequenced. The clones were named by the passage number (P), and by the infection point, i.e., seroconversion (S) and death (D).

**Sequencing and analysis.** The double-stranded plasmid DNA was sequenced by using custom-labelled dye primers (Applied Biosystems Inc.) with an automated sequencer (model 373A; Applied Biosystems) and version 1.2.0 software. Clones were assembled and aligned with the Sequence Navigator program (Applied Biosystems) or Clustal V (18) and further optimized manually. Phylogenetic analysis was conducted with MEGA (26), Clustal V (18), PAUP 3.1.1 (52), and PUZZLE 3.0 (49). Neighbor-joining, maximum-likelihood (ML), and puzzle-based trees were produced by using the Tamura-Nei distance estimation with pairwise comparison and then bootstrapped with 1,000 replications before tree construction with and without preselected outgroups (P1 sequences). The ML mapping and QP were carried out by using the Tamura-Nei distance estimation (49). Briefly, likelihood mapping is the construction of all quartets of sequences in the analysis have been compared) and calu-

FIG. 1. Phylogeny of the African SIV and HIV-2 strains commonly used in experimental infections. Known pathogenic and less pathogenic SIVs are seen to cluster together (e.g., SIVH4 and SIVSM9, and SIVMM239 and SIVMM1A11) according to virus strain and not by pathogenicity. The P1 and P6 sequences are more divergent than some of the other clades. It is also noteworthy that the pathogenicity of the P6A and P6B viruses is equal to that of PB14 but there is no tendency for these two SIVMM derived env sequences to cluster together. The same results were achieved with different P1 and P6A or P6B sequences as examples. This tree was constructed with PUZZLE 3.0 with the smallest information-containing subset of a tree, the quartet of sequences (16, 49) allows efficient visualization of the estimated phylogenetic content. Our application of this method to the SIVsm passage alignment showed it to be highly informative compared to HIV alignments, and the reconstruc-
lation of the relative weights of the probabilities for all three resolved tree structures (1: A is like B but not C or D; 2: A is like C but not like B or D; and 3: A is like D but not like B or C) (48–50) and four unresolved structures (partially resolved 1: A is like B or C but not D; partially resolved 2: A is like B or D but not C; partially resolved 3: A is like C or D but not B; and unresolved: A is not like B, C, or D). Thus, for the first four sequences analyzed by likelihood mapping, the seven possible locations of the probability vectors (49) include three with a completely resolved placement of the four taxa, three with one taxon split between attraction to two others and not like the fourth taxon (partially resolved), and one of star-like form or no favored relation between the four taxa (unresolved). These seven relations can be plotted in a triangle to show the phylogenetic signal (see Fig. 3, 5, and 7). The corners represent areas of completely resolved (tree-like) quartets (A is most like one of B, C or D, resolved 1, 2, and 3); the sides represent quartets with a split between equal attraction of one taxon for two other taxa (A is like B and C, B and D, or C and D, partially resolved); and the center of the triangle represents an area of equal likelihood or distance between the four taxa (A is not like B, C, or D, unresolved, giving a star-like formation). If each possible quartet vector (resolved, partially resolved, and unresolved quartet comparison) is plotted as a point in the triangle, it gives a rapid visualization of the phylogenetic content of the alignment data (robustness of the tree, or lack thereof). Tree construction was carried out with the ML-based PUZZLE 3.0 (48). The ML analysis is used for determination of a tree and its corresponding branch lengths that have the greatest likelihood of reconstructing the correct phylogeny. The numbers of possible topologies increases exponentially as the number of taxa increases, making heuristic searches very slow. Reconstructing a tree from all of the possible sets of four taxa, or quartets, allows the heuristic search to proceed efficiently in an ML procedure (50). QP reconstructs the ML tree for all possible quartets. The total set of quartet trees is subsequently combined to form a complete tree in the QP step. This procedure was carried out 1,000 times. The QP tree is a majority-rule consensus (30) of the set number of puzzling steps. There are three main steps to the QP method: ML determination of all possible quartet trees, combining all of the quartet trees 1,000 times to form a complete tree (puzzling step), and the final majority-rule consensus computation of all intermediate trees (49). The HIV sequences from 44 progressors and nonprogressors at seroconversion and after 5 years of infection (29) are available upon request.

RESULTS

Construction of molecular phylogenetic trees. Our study included seven Asian rhesus macaques whose serial infection involved a well-characterized viral isolate (SIVsm B670) (2, 5, 6, 35), and known times of infection, seroconversion, and direction of transmission. It thus relates a serial population passage and known phylogeny of an SIV. The use of phylogenetic analysis on the quasispecies sequences allows evaluation of the inference methods and insight into the pattern of viral evolution. The inference of the correct molecular phylogeny has
FIG. 4. HIV-1B phylogenetic inference for the 88 sequences from the 44 HIV-1-infected progressors and nonprogressors, 5 years apart, (29) by ML determination (PUZZLE 3.0) with the Tamura-Nei distance approximation and 1,000 tree-puzzling steps. The QP percentages of the important branches are in bold numbers. The scale bar is shown in the bottom left-hand corner. The topology of the HIV-1B tree is star-like as predicted by likelihood mapping and shows little relation between the sequences, except for the pairs of sequences from individual patients, which group together.
important implications for our understanding of transmission patterns and quasispecies evolution. We used four different tree-building methods: neighbor-joining, ML, maximum parsimony (MP), and QP. The results of the four methods were similar only if the number of taxa used in the analysis was reduced. The neighbor-joining, ML, and MP methods accommodated all 50 clones and maintained a recognizable portrayal of the known phylogeny (Fig. 2), albeit with discrepant branch placement. Inordinate amounts of time were needed with MP for branch swapping to prevent the irreversible misplacement of branches early on in the stepwise addition. The QP reconstruction produced the most true phylogeny, with knowledge of the relation of the viruses present at the time points sampled during the serial passage. The more closely related sequences (less heterogeneity between the sampled time points; e.g., the genetic distance between P1 to P2S is more than from P3D to P4S) after the third passage present a resolution problem for all of the known methods. The P3S, P3D, and P4D sequences clustered consistently. The discrepancy was the branching order of the P6A and P6B sequences, which were closer to the founder P1 and P2 sequences than to their precursor P3 and P4 sequences. This problem was not resolved by the use of other distance estimation methods or increased bootstrap resamplings.

The problem with tree construction with a large number of taxa is the inability to calculate all of the possible tree combinations. It is further complicated by the long-standing problems in phylogenetic inference. The evolutionary history of retroviruses is difficult to reconstruct because of recombination between similar but unlike RNA strands, selection of progeny during rapid turnover of viral genomes, differences in the rates of mutation among strains, unbalanced nucleotide frequencies, and differences in the individual substitution rates of the already skewed nucleotide concentrations. The use of quartet analysis, with its focus on the smallest informative subtree (27, 48–50), has been applied to an inference method that takes advantage of the determination of all possible quartets, or heuristic search, which is unavailable to the other methods, too time-consuming, or available only as a statistical test of best fit (27). PUZZLE 3.0 (48–50) incorporates a first-step plot of all the possible four taxon quartets as a means of weighting the best placement of them all within a phylogeny. With seven possible organizations for each quartet, calculation of all possibilities is feasible. The ML of the grouping of a quartet can also be plotted graphically to give an idea of the “tree likeness” of the data known as likelihood mapping (49) (see Fig. 3, 5, and 7). The four taxa are compared by grouping three and seven possible ML areas as a dot: with taxon D most like taxon A and C, or B and C (the three sides: unresolved between the two), and, lastly, unlike A, B, or C and placed in the middle as a neutral attractor (star-like formation). The centers of the ML triangles are inhabited by star-like phylogenies lacking tree-like content, i.e., having no attraction to cluster with the other taxa (Fig. 3 and 4). The four taxa are compared by grouping three and seven possible ML areas as a dot: with taxon D most like taxon A, B, or C (the three corners: tree-like), split between A and B, A and C, or B and C (the three sides: unresolved between the two), and, lastly, unlike A, B, or C and placed in the middle as a neutral attractor (star-like formation). The centers of the ML triangles are inhabited by star-like phylogenies lacking tree-like content, i.e., having no attraction to cluster with the other taxa (Fig. 3 and 4). The percentages of the contents of the complete set of quartets (≥300,000 for 50 taxa) are displayed in the numbered triangles. Each of the seven areas has a finite number of the ML dots, with the total of the seven being 100%.

The utility of likelihood mapping is shown by analysis of an extensive set of HIV-1B sequences (Fig. 3) (29) from 44 infected progressors and nonprogressors sampled at seroconversion and 5 years thereafter. The three corners contain 74% of the quartets, and fully 26% of the quartets are unresolved, with points in the center of the triangle (16.3%) being neutral attractors or star forms. The phylogeny of this alignment (Fig. 4) shows intrapatient sequence clustering and no definition of evolution of the HIV-1 B env gene over time. Evolution of HIV-1 during infection is not as structured as the cross-species passage because several, or many, infections have already taken place between the progenitor virus (presumably a relation of SIV from chimpanzees) and the modern HIV-1. The lack of sequences from the “original” founder of HIV-1 leaves a gap in the phylogeny. Without the founder and early infection sequences, the variation is seen to be nonstructured, since there is no directed evolution or divergence over time.

Figure 5 shows the tree likeness of all 50 taxa of the SIVsm passage. The sequence alignment contains a very high 96.5% of the phylogeny in the three corners, or “in” the tree. There are 241 variable and 135 parsimonious sites of 1,119 nucleotides in this alignment. The split-between-partners quartets or sides of the triangle include 3.1% that cluster between two other taxa and only 0.4% that are unresolved (“out” of the tree) in the “star” area where HIV-1 phylogenies exist (Fig. 4).

The phylogeny derived from generation of all possible quartets, QP (Fig. 6), tells the complete story of the serial passage in graphic form. As the infection history is known, the tree is rooted in the P1-1 clone as the founder sequence. The heterogeneity of the P1 sequences is evidenced by the long branches that contain only pairs of taxa. The initial adaptive effect is seen in the large differences in the lengths of the P1 branches, varying from P1-6 to P1-8 and P1-9 (which are the same distance from P1-1). P2S and P2D are not completely separate, with two clusters forming, but four of the five P2-D taxa are distinctly divergent from the P2-S cluster, and P2D-2 is a likely founder for the next passage, although P2S-4 is actually closer to the P3 to P6 sequences. The branch lengths and the clustering numbers place the P2 taxa between the P1 and the P3 to P6 taxa. This pattern follows the changes in pathogenicity during passage and holds for the P1 to P2 clustering as well. The P3 to P6 taxa are placed on the same branch, with a
clustering value of 64, since the P3 and P4 taxa do not carry much information that could cluster them separately as distinct clades. The differences in branch lengths are indistinct because of minimal sequence variation between the P3 and P6 clones. There is an almost quantum-like effect, with the branch lengths of the P3S-1 and seven other taxa all being equal. The P4D-4 and six other clones also have equal branch lengths, seemingly twice as long as those of the P3S-1 length. The tree continues to grow outward from the root, with the P6 sequences having longer branch length than those of the P3 and P4 sequences. The P6B sequences form a distinct branch, except for the P6B-4 clone, which has seven amino acid substitutions not found in the other P6B sequences.

ML mapping was used to display the amount of tree-like information held by a separation of the sequences (P1 and P2 as one cluster and the P3 to P6A and P6B sequences as another). The robustness of the branching separation of the two clusters (Fig. 7) is seen as the absence of any shared quartets which would appear in the center of the triangles. The mapping shows the correlation of the branching order and distinctly altered pathogenicities of these quasispecies. The taxa were set in two distinct clusters, with 97.6% of the phylogenetic information confirming the separation of these subsets of the passage clones as evolutionarily distinct.

We divided variable and constant regions into two separate phylogenies to assess the pattern of evolution of these two functional types of env domains during adaptation and increase in pathogenicity (Fig. 8 and 9). The variable-region phylogeny

FIG. 6. SIVsmB670 passage complete clone phylogenetic inference for the passage alignment. The tree was constructed by ML determination (PUZZLE 3.0) with the Tamura-Nei distance approximation and 1,000 tree-puzzling steps. The QP percentages of the important branches are shown by boldface numbers. The scale bar is in the lower left-hand corner. The tree shows the pattern of evolution during adaptation of SIVsm to a new species of nonhuman primate and the effect of the serial passages upon env during the increases of pathogenicity.
VOL. 72, 1998 SIV sm env EVOLUTION DURING PASSAGE IN RHESUS MACAQUES 3679

FIG. 7. SIVsmB670 passage complete clone two-cluster likelihood-mapping analysis. The sequences were split into two disjoint groups: a (P1 to P2) and b (P3 to P6). The corners of the triangles are labelled with the corresponding tree topologies. The top triangle shows the distribution pattern of the quartet analysis, and the bottom triangles show the distribution of the data in the seven basins of attraction or quartet possibilities. The likelihood-mapping analysis can also be applied to the testing of an internal edge of a tree. The clusters chosen separate the early, less pathogenic env clones of the virus from the later, more pathogenic env clones.

shows the large heterogeneity within the P1 sequences that was observed above. With 65% of the variable sites being found within the variable regions, the P1 sequences are split into three separate branches. Cluster definition decreases after seroconversion of passage 3 (P3S clones), except for the P6B sequences. The heterogeneity appears to be very large, but, as the distance bar shows, the distances are far greater than those encountered in the constant regions. Misclustering was found due to the enormous variation, and not the similarity, of the sequences. In the variable regions, the average diversity was limited after the P3D sequences.

The constant-region tree structure places the different passages in order (Fig. 9). The P1 heterogeneity is still large, but not large enough to reduce the cluster definition of the whole tree, as in the variable regions. The P3S sequences are the most clustered and mark the greatest increase in pathogenicity, with viral diversity increasing as the passage progresses toward the P6 sequences.

DISCUSSION

Serial passage of SIVsmB670 from sooty mangabeys in rhesus macaques has drastically shortened the time of disease progression from 1.5 years to 1 month during adaptation of the retrovirus to the Asian rhesus macaque hosts. The increase in pathogenicity has been correlated with higher antigen levels in plasma, persistent plasma antigenemia, and a more rapid loss of T-helper/memory cells (53a). The more rapid disease progression of the P6 inoculum produces clinical symptoms similar to those due to the slower progressing P1 inoculum. The SIVsm-related strain PBJ14 (Fig. 1) is pathogenic within 2 weeks in pig-tailed, cynomolgus, and some rhesus macaques (43) but does not produce the immunodeficiency seen in HIV-1 infections. It caused massive T-cell proliferation, increased cytokine release, and mucoid diarrhea, and its env glycoprotein has some superantigen-like properties (44). The P6 inoculum reproduced the relevant immunodeficiency symptoms to model AIDS in humans, except for lack of seroconversion.

The present study was used to analyze the genetic adaptation of the SIVsm env gene during serial population passage in rhesus macaques. The clones sequenced were all unique, with decreased heterogeneity or viral diversity in the P6 sequences compared to those of P1. The heterogeneity or divergence seen may not be representative of genotypes present in tissue, but, for the purpose of identifying the evolutionary path during the adaptation of the env gene between the P1 and P6 sequences, this single compartment sampling suffices. The major areas of variability were observed in the V1 and V4, as is the norm in SIV infections (10, 19, 39), but mutations were seen also within the V2 and V3 regions and the C3 region just C-terminal to the V3 loop.

Phylogenetic inference methods were used to analyze the evolution of the env gene of SIVsm during adaptation in rhesus macaques. The QP method was used to allow the rendering of phylogenetic content of the data set and for the two-cluster mapping of the variable versus the constant domains of the env. Four-taxon trees may, in some purposely selected cases, be difficult to resolve (18a), but a one-tree example is not a defining study (54). Sequence length is always a crucial factor in tree building, but the simultaneous analysis of many lineages tends to improve phylogenetic estimation considerably (18b). Quartets can be hard, but extra information helps. If all that is available consists of data on species A, B, C, and D, it might be relatively difficult to find the correct tree for them. However, if additional data are available (species E, F, G, ...), and an attempt is made to find a tree for all the species, that part of the tree relating A, B, C, and D will be expected to be more accurate than if just the data for A, B, C, and D were available. There are many examples of subsets of four species which in themselves might be hard to resolve correctly but which are correctly resolved thanks to additional data (e.g., inordinately long sequences or many lineages). PUZZLE 3.0 gains advantage from extra data in the same way. Its “understanding” or resolution of the quartet A, B, C, and D may be incorrect, but the information on the relationships of A, B, C, and D implicit in its treatment of A, B, C, and E, of A, B, D, and E, of A, C, D, and E, of B, C, D, and E, of A, B, C, and F, of A, B, D, and F, of A, C, D, and F, of B, C, D, and F, of A, B, C, and E, etc., should overcome this problem (17a). Using this quartet-based heuristic search algorithm (50) for the best tree fit and 50 sequences of 1,119 bp each, we found that the sampled sequences diverged from that of the primary inoculum in a time-dependent fashion, following the path of the serial passage. The topology of the tree placed the env sequences in clusters roughly according to the pathogenicity of the virus and the history of the passage infections.

The relationship between infection history, pathogenicity, and the hierarchy of the inferred trees indicates that env plays a role in the progression to AIDS during SIV infection. The sequence divergence and its correlation with enhanced virulence suggest that our recorded amino acid substitutions are important to the shortening of the asymptomatic period of progression to AIDS in the Asian rhesus macaques.

Viral diversity decreased during passaging (Fig. 6, 8, and 9) (53), in particular in the constant regions, following the initial adaptation to the new host. The phenomenon of viral diversity is pivotal to the production of vaccines. We have shown here that this diversity is related to the level of adaptation of the virus to its new environment and that, after several infections, an equilibrium is reached in amino acid substitution (unpub-
lished data), $K_r/K_a$ ratio (53), and branch length variation at which the virus is optimally tailored to the surrounding conditions. Variation in the amino acids of the env product was greatest in the V1 and V4 regions followed by V2 and V3. The difference in receptor sequence between the CD4 of sooty mangabeys and Asian rhesus macaques is not known, nor is the difference in relatedness of the accessory receptors for this virus in these two hosts (1, 13, 15). The adaptive process may hinge on the viral need for the most efficient env conformation to bind the CD4 T cells and begin the infective process. Increased pathogenicity would imply increased viral adaptation and increased ability to bind and enter the target cells of the new host. The most variable, or most adaptable, regions were V1 and the V4. The precise role of V1 (in HIV or SIV infections) is not yet known. V3 and C4 are known to play a role in the binding of CD4 (23); although not binding directly, they play a conformational role. Because V3 and C4 span V4, V4 (along with other regions in the env gene) may encode amino acids that are part of the CD4 binding domain. If the binding of the CD4 and accessory receptor are relevant to the development of lentivirus pathogenicity, it would follow that the fastest-adapting areas of the env gene would be those need-

FIG. 8. Phylogenetic inference for the variable regions of the SIVsmB670 passage alignment. The tree was constructed by ML determination (PUZZLE 3.0) with the Tamura-Nei distance approximation and 1,000 tree-puzzling steps. The QP percentages of the important branches are shown by boldface numbers. The scale bar is in the lower left-hand corner. The pattern of evolution of the variable regions is less structured than that of the constant regions (Fig. 9). The P1 to P6 sequences cluster together on one branch, except for the root (P1-1) and four P1 clones (P1-2, P1-3, P1-6, and P1-10).
ed to contact the CD4 and accessory receptor for SIVsm (CCKR-5 is the known accessory receptor for HIV-1 and SIV (15). If so, V1 and V4 are in conformational contact and are the major env structures involved in binding and entry.

The relative clustering of the env sequences by virulence and by sequence divergence shows the impact of genetic variation of env upon pathogenicity. The difficulty of nonrobust phylogenetic separation of the P3 to P6 sequences shows that env
evolution is not a simple linear process. Apparently, convergence can occur during multiple serial infections after adaptation has occurred in relation to the env gene of SIVsm. The decreasing heterogeneity seen in the P3 to P6 taxa may reflect the reduced adaptation and convergence or sequence stability of the new consensus around which these sequences evolve in a decreasing area or may indicate that selection has effectively removed unsuccessful variants.

ACKNOWLEDGMENTS

We thank Korbinian Strimmer and Arndt von Haeseler (Universität München) for providing a prerelease version of PUZZLE 3.0 and helpful advice, Carla Kuiken (Los Alamos) and David Mindell (University of Michigan) for critical reading and revision of the manuscript, Nick Goldman (University of Cambridge) and David Hillis (University of Texas) for explaining the basics of quartet resolution, and Lucy Phillips for editorial wisdom. An anonymous reviewer was exceedingly helpful in coaxing the manuscript into a more informative and readable form. Thank you.

REFERENCES


53. Valli, P. J. S., V. V. Lukashov, and J. Goudsmit. Shortening of the symptom-free period in rhesus macaques related to decreasing non-synonymous variation in the env variable regions of SIVsm during passage. Submitted for publication.

53a. Valli, P. J. S., J. L. Heeney, and J. Goudsmit. Selection for macrophage tropism during serial passage of SIVsm in rhesus macaques concomitant with increased virulence. Submitted for publication.

