

Phylogenetic Analysis of Human Immunodeficiency Virus Type 1 in Distinct Brain Compartments Provides a Model for the Neuropathogenesis of AIDS

Marco Salemi,^{1,2*} Susanna L. Lamers,^{3,4} Stephanie Yu,⁴ T. de Oliveira,⁵ Walter M. Fitch,¹ and Michael S. McGrath^{4*}

Department of Ecology and Evolutionary Biology, University of California at Irvine, Irvine, California¹; Department of Pathology, Immunology, and Laboratory Medicine, University of Florida, Gainesville, Florida²; Gene Johnson Inc., St. Augustine, Florida³; Department of Laboratory Medicine, Positive Health Program, University of California at San Francisco, San Francisco, California⁴; and Evolutionary Biology Group, Zoology Department, University of Oxford, Oxford, United Kingdom⁵

Received 17 December 2004/Accepted 20 May 2005

“Phylogenetic” analysis combines various statistical procedures that can be used to correlate the epidemiological and evolutionary behavior of viral pathogens with the immune system of the host. We utilized this approach to examine human immunodeficiency virus type 1 (HIV-1) gp120 envelope DNA sequences (V1, V2, and V3) isolated from different brain compartments of a T-cell-depleted patient diagnosed with severe HIV-associated dementia at the time of death. In agreement with previous reports, phylogenetic analysis showed distinct virodemes but also revealed a significant amount of viral gene flow among different brain compartments. Local-molecular-clock analysis showed that HIV-1 meninges and temporal lobe subpopulations evolve about 30 and 100 times faster, respectively, than the other viral populations in the brain. However, maximum likelihood codon-based substitution models did not detect any site under significant positive selective pressure, and the main cause of HIV-1 genetic variation appeared to be random genetic drift. Therefore, the higher evolutionary rate in the meninges and temporal lobe could be due to an enhanced infection/expansion rate of macrophages as a consequence of the immune system failure. In conclusion, in this case study, viral infection in the brain progressed with a nonspecific genetic evolution, recurrent migration events, and an expansion of macrophage-tropic sequences. The data suggest that after immune failure newly produced viral variants, which would be rapidly cleared under normal conditions, begin to productively infect macrophages in a “self-amplifying” cycle of infection/inflammatory response that could be at the origin of HIV-associated dementia.

Approximately 20% of individuals with primary human immunodeficiency virus type 1 (HIV-1) infection syndrome have symptoms of viral meningitis (2, 5, 23, 38). HIV-1 infection of the brain occurs early, whereas HIV-associated dementia (HAD) generally develops only in late-stage HIV-1 disease (<50 CD4 cells/ μ l) (23). HIV-1 expansion within the brain differs from lymphoid organs because the virus primarily infects macrophages and utilizes the CCR5 coreceptor (22). This is in agreement with the finding that perivascular macrophages are the cells primarily infected in the central nervous system (CNS) of macaques that develop encephalitis (40). Presumably, early infection of the brain occurs via a series of events involving CNS damage and subsequent viral migration through a variety of tissues, including blood, cerebrospinal fluid (CSF), meninges (lining the outer layer of the brain parenchyma), and various brain tissues and/or compartments. The CSF is a dynamic pool of fluid that interfaces the meningeal layer. Averaging 125 ml, the CSF turns over five times per day, and in fact, a key symptom during primary HIV infection is meningitis

(38). Once the virus enters the brain parenchyma, spread is likely due to local extension of brain macrophages and microglia. In a recent paper, Williams and Hickey (39) have suggested that while during the early phase of the disease the immune system still controls the infection in the macrophages, later it becomes essentially unable to prevent the emergence of productively infected cells as the patient progresses to AIDS. In their view, the continuous accumulation in the brain of activated macrophages initiates a self-amplifying cycle eventually leading to the development of HAD. The model is attractive as an explanation for both the latency and the long-term brain damage observed in many patients. What is unknown are the stimulus and the mechanism responsible for the increased macrophage trafficking during the AIDS phase. A relevant factor could be the emergence of specific HIV-1 pathogenic sequences late in the disease, when immune control has failed, or viral subpopulations migrating through specific routes and/or viruses replicating in specific brain compartments at different rates.

Several reports on HIV-1 sequence heterogeneity in the brains of infected individuals, with or without HAD, have shown that brain-derived viral sequences are often monophyletic in the inferred phylogenetic tree, suggesting distinct virodemes (subdivided viral populations) in the CNS (13, 22, 26, 29, 42). In particular, the finding of independently evolving HIV-1 subpopulations in the frontal lobe, basal ganglia, medial temporal lobe, and nonmedial temporal lobe suggests the ex-

* Corresponding author. Mailing address for M. Salemi: Dept. of Pathology Immunology and Laboratory Medicine, 1600 S.W. Archer Road, University of Florida, Gainesville, FL 32610. Phone: (352) 392-3749. Fax: (352) 392-3053. E-mail: msalemi68@yahoo.com. Mailing address for M. S. McGrath: Dept. of Laboratory Medicine UCSF 0874, SFGH Bldg. 80, Wd 84, 1001 Potrero Ave., San Francisco, CA 94110. Phone: (415) 476-4082 ext. 415. Fax: (415) 476-6017. E-mail: mmcgrath@php.ucsf.edu.

istence of restricted pathways for HIV-1 evolution within the CNS (6, 29). Compartmentalization and independent evolution of primary and secondary drug resistance mutations to reverse transcriptase and protease inhibitors in diverse CNS regions of HIV-1 patients, with and without HAD, have also been shown (33). However, the role, if any, that HIV-1 sequence variation might play in the neuropathogenesis of HAD is currently unknown. This is due, in part, to complexities concerning viral migration across the blood-brain barrier and to a paucity of studies of compartmentalization of HIV-1 quasispecies in different brain-derived tissues (1).

The main goal of the present study was to identify the extent of tissue-specific HIV-1 evolutionary patterns in the CNS. In particular, we studied the viral subpopulations infecting tissues usually associated with initial infection (meninges and cortex) and disease progression (white matter and spinal cord) isolated from a T-cell-depleted patient diagnosed with severe HAD at the time of death. We used a so-called "phylogenetic" framework, which combines various phylogenetic and population genetic analyses to investigate the correlation between the epidemiological and evolutionary behavior of viral pathogens and the immune system of the host (10). Our findings not only suggest that the Williams and Hickey (39) model may be correct, but also offer a plausible mechanism that may explain the interplay between macrophage activation and viral evolution in the development of HAD.

MATERIALS AND METHODS

Characterization of patient specimens. Brain tissue samples from a T-cell-depleted patient with severe HAD at the time of death were obtained from the University of California at San Francisco AIDS and Cancer Specimen Resource (30). Five CNS tissues were sampled: meninges (lining the surface of the brain), subcortical white matter from the frontal lobe, occipital lobe cortex, temporal lobe (gray-white matter junction), and spinal cord. Control sequences were obtained from both a lymphocyte-depleted iliac lymph node and seminal vesicles. DNA was prepared, and the HIV copy number/genomic equivalent within specimens was quantitated as previously described (19). Immunohistochemistry on tissues to evaluate the sites of HIV expression was also performed as previously described (19).

PCR, sequencing, and data management. The envelope V1-V3 region was amplified by nested PCR using the following set of primers: EnvF1 (outer forward), AACATGTGGAAAAATAACATGGT; EnvR1 (outer reverse), TG WATTACAGTAGAAAAATTCCTCC; EnvF2 (inner forward), TGGTAGAAC AGATGCAKAGAGGA; and EnvR2 (inner reverse), CCTCCACAATTAATAA CTGTGY. Outer PCRs were carried out for 35 cycles of 15 s at 95°C, 30 s at 55°C, and 1 min at 68°C, followed by a final 8-min extension step at 68°C, in a 100- μ l volume using 100 ng of DNA template, 10 μ l of 10 \times buffer (Roche, Indianapolis, IN), 10 mM deoxynucleoside triphosphates, 20 μ M of each primer, and 0.45 units of Roche *Taq* polymerase. Inner PCRs were carried out under the same conditions using 1 μ l of the outer PCR product. Inner products were cloned into the pCRII-TOPO vector, and positive clones were detected by PCR, using the TOPO TA Cloning Kit Dual Promoter (Invitrogen, Carlsbad, CA) with conditions recommended by the manufacturer. Sequencing was carried out by ELIM Biopharmaceuticals Inc. (Hayward, CA) on an ABI 3730xl sequencer. About 20 clones (>120 total sequences) of PCR-amplified *env*-V1-V3 from each different brain region were obtained. V1-V2 and V3 domains were isolated using HIVbase software (<http://www.hivbase.com>) (14). Sequence integrity and possible contamination were tested according to the Los Alamos HIV Sequence Database guidelines (<http://hiv-web.lanl.gov>). Charge analysis was performed using HIVbase software. Multiple sequence alignments were generated using the Clustal algorithm (37) and manually edited in order to maximize positional homology in gap-rich regions following the Lamers et al. protocol (16). Alignments are available from the authors upon request.

Phylogenetic analysis. To exclude the possibility of intrapatient recombination in the envelope region studied, V1-V2 and V3 alignments were analyzed separately. The samples were all identified as HIV-1 subtype B by phylogenetic

TABLE 1. HIV copy number per genomic equivalent of tissue DNA

HAD patient brain region	HIV copy no. ^a
Meninges.....	93
Temporal subcortex.....	153
Occipital cortex.....	55
Frontal subcortex.....	95
Spinal cord.....	66
Seminal vesicle.....	486
Lymph node.....	17

^a HIV *gag* copy number versus *c-jun* copy number as genome equivalent control (21). The number of cells per HIV copy is given.

analysis with known reference strains, and no significant differences were detected between V1-V2 and V3 phylogenies (data not shown). Maximum likelihood (ML) trees were estimated as follows. First, the best-fitting nucleotide substitution model was tested with a hierarchical likelihood ratio test following the strategy described by Swofford and Sullivan (35) using a neighbor-joining tree with Jukes and Cantor corrected distances. ML trees were then inferred with the selected model (HKY + Γ , eight categories, for both data sets) and ML-estimated substitution parameters. The heuristic search for the best tree was performed using a neighbor-joining tree as the starting tree and the TBR branch-swapping algorithm with the collapse option off to avoid polytomies in the final tree. Calculations were performed with PAUP* 4.0b10 (D. L. Swofford, Sinauer Associates, Sunderland, MA). Statistical support for internal branches in the tree was obtained by bootstrapping (1,000 replicates) and the ML-based zero-branch-length test (35). Trees were rooted by ML rooting by selecting the rooted tree with the best likelihood under the molecular-clock constraint.

Gene flow tests and migration counts. The hypothesis of compartmentalization, i.e., the existence of distinct HIV-1 subpopulations in the brain, was tested by the Slatkin and Maddison test for gene flow (31) using the MacClade program (20). A one-character data matrix is obtained from the original data set by assigning to each taxon in the tree a one-letter code indicating its tissue of origin. Then, the putative origin of each ancestral sequence (i.e., internal node) in the tree is inferred by the Fitch algorithm (9) by finding the most parsimonious reconstruction (MPR) of the ancestral character. The final tree length, i.e., the number of observed migrations in the genealogy, can easily be computed and compared to the tree length distribution of 10,000 trees obtained by random joining-splitting. Observed genealogies significantly shorter than random trees indicate the presence of subdivided populations (virodemes). The presence of structure in the trees was also tested with the Finkelstein test (8), which gives the probability of any particular clade being the result of random association due to runs of identical events. Finally, specific migrations among different compartments (states) were traced with the state changes and stasis tool (MacClade), which counts the number of changes in a tree for each pairwise state. When multiple MPRs are present (as they are in our data sets), the algorithm calculates the average migration count over all possible MPRs for each pair. The resulting pairwise migration matrix was normalized, and a migration network was sketched representing the proportion of observed migrations in the tree from/to each tissue sampled.

Parametric and nonparametric estimates of demographic history. A generalized skyline plot is a nonparametric estimate of demographic history based on an inferred ultrametric genealogy (34). A genealogy reconstructed from randomly sampled HIV sequences contains information about population level processes, such as change in population size and growth rate (11, 25, 36). A skyline plot can be inferred from a genealogy with clock-like branch lengths, and it represents an estimate of how the population size changes over time (25, 34). Given a viral phylogeny, P , and a vector, ϕ , representing the parameters of the demographic model $N(t)$, parametric estimates can also be obtained by calculating the log of the conditional probability, $\ln[\phi|P]$. ML estimates of ϕ can be found by numerical optimization of $\ln[\phi|P]$, and 95% confidence intervals for the estimates are obtained with the likelihood ratio statistic (25, 34). For example, in a model of exponential population growth, the estimated parameters will be $N(0)\mu$ and r/μ , where μ is the evolutionary rate in nucleotide substitutions per site per year and $N(0)$ is the effective number of infections at present. Notice that if the evolutionary rate cannot be inferred from the data, as with sequences collected at a single time point without any known divergence time in the tree, time is expressed in terms of nucleotide substitutions per site, and it runs backward into the past, so that $N(0)$ is the effective number of infections at present, and $N(t)$ is the effective number of infections at time t in the past. Parametric estimates for

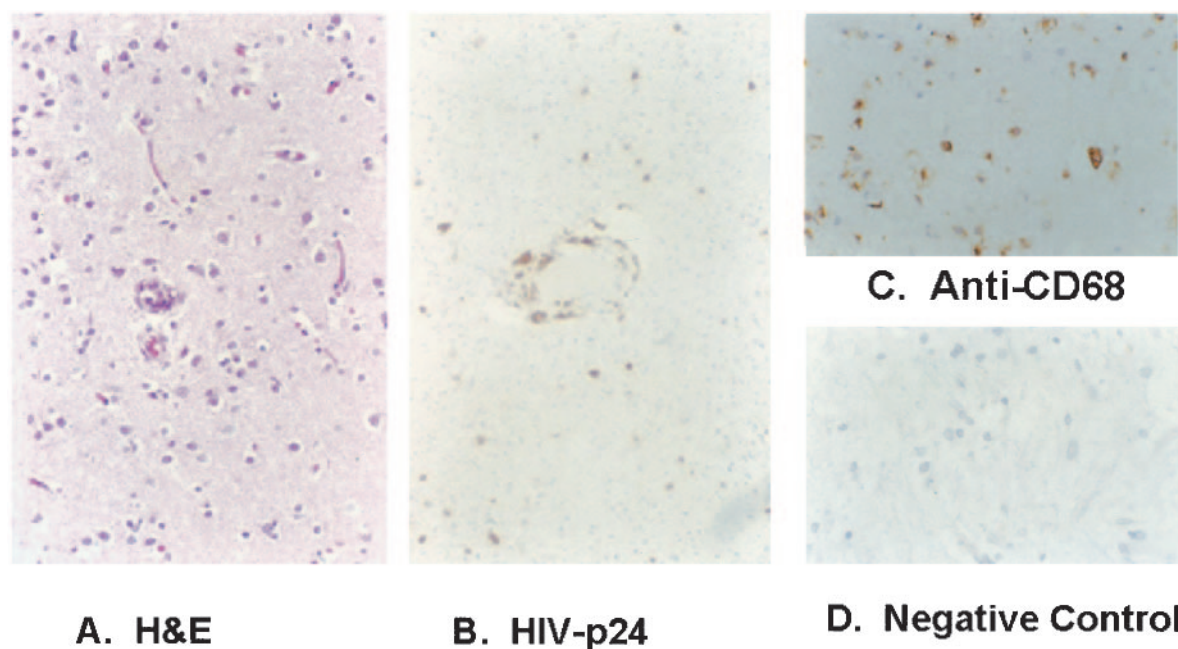


FIG. 1. Histology of HAD brain. (A) Hematoxylin and eosin (H&E) staining of the frontal lobe subcortical white matter. (B) Anti-HIV p24 staining. (C) Anti-CD68 (macrophage) staining. (D) Control antibody staining.

the constant-size, exponential, and logistic-growth models were obtained with the GENIE v3.0 software package (24). The tested models are nested; therefore, we could use the likelihood ratio test to assess which one best fits the data set studied. Differential equations defining the models can be found in the GENIE manual distributed with the package. Generalized skyline plots were obtained with the APE package implemented in R (<http://cran.r-project.org>). Because the molecular clock for the brain data set is rejected (see Results), the tree used as the input file for the skyline plot calculation is the tree with local-clock-corrected branch lengths assuming a different evolutionary rate for the major tissue-specific clades in the tree.

Molecular-clock analysis. Ultrametric trees were obtained by enforcing a molecular clock on the inferred genealogy and reestimating the branch lengths and substitution parameters with maximum likelihood using the codon substitution model M0 described by Yang et al. (44). The local-molecular-clock model was implemented by defining a baseline evolutionary rate, equal to 1 by default, for the tree containing HIV-1 sequences isolated from different brain tissues, plus five different local clocks for the major tissue-specific clades in the tree employing the same codon substitution model used for the global clock. Calculations were performed with the CODEML program of the PAML package (43). Different clock hypotheses were tested with the likelihood ratio test. Degrees of freedom for the test between global- and local-clock models were calculated by considering that for a binary tree a local-clock model has $n - 1 + r$ free parameters, where r is the number of unconstrained relative evolutionary rates (43). Branch lengths proportional to divergence time were also inferred with the nonparametric-smoothing algorithm implemented in the program r8s (<http://ginger.ucdavis.edu/r8s/>). Nonparametric smoothing relaxes the assumption of a perfect molecular clock, allowing evolutionary rates to vary smoothly along a tree (28).

Positive-selection analysis. Positive-selection analysis was performed by comparing the maximum likelihood codon substitution models developed by Yang et al. (44) and implemented in CODEML. Two categories of models were tested. (i) Branch site models (43) tested for different dN/dS (nonsynonymous/synonymous substitution) ratios along given branches of the tree. Three different models were compared: model 0, assuming a single dN/dS for the whole tree; model 1, assuming a different dN/dS for each branch in the tree; and model 2, assuming a baseline dN/dS equal to 1 for the entire tree and five different dN/dS along the branches of the five major tissue-specific clades. Model 0, model 1, and model 2 are nested, and a hierarchical likelihood ratio test can be used to check which one of them fits the data significantly better (43). (ii) Nnsites models (43) tested for specific sites under positive selection. The models compared were M7 and M8. M7 is a neutral model assuming a β distribution with 10 classes of dN/dS across codons, where each class is constrained to have a dN/dS of <1 . M8 is a

positive-selection model assuming a β distribution with 11 classes, 10 with a dN/dS of <1 and one class with a dN/dS of >1 . Since the two models are nested, the likelihood ratio test can be used to check whether assuming an extra class of codons under positive selection fits the data significantly better.

All analyses were carried out with the latest available version of the PAML (v3.14) software package (43).

V3 loop cross-pattern tissue analysis. The ancestral sequence of the node that linked the meninges, temporal lobe, and lymph node-semen main clusters was reconstructed by maximum likelihood, using the BASEML program implemented in the PAML package (43). Sequences were also divided into 10 groups based on the major supported monophyletic clades present in the inferred genealogy, and the most recent common ancestor (MRCA) for each group was reconstructed. Signature pattern analysis was carried out with the VESPA program by comparing each sequence group to its inferred ancestor (12).

Epitope and motif searches. All published cytopathic-T-lymphocyte (CTL) epitopes and associated HLA serotypes can be found in table format at the HIV Molecular Immunology website (<http://www.hiv.lanl.gov/content/immunology>). All CTL epitopes specific for envelope were built into a search query within HIVbase software and applied to the isolated data sets in this study. We also built search queries for crest motifs previously associated with HIV-1 neurotropic variants (32). In order to screen both our sequences and sequences in the Los Alamos HIV databases (<http://www.hiv.lanl.gov>) for motifs, all the available sequences from the HIV databases were imported into HIVbase and the appropriate query was carried out with the HIVbase query tool. The saved search queries are available for download (<http://www.HIVbase.com>).

Nucleotide sequence accession numbers. Individual sequences have been deposited in GenBank under accession numbers DQ121207 to DQ121294.

RESULTS

HIV DNA copy number frequency in tissues studied. Table 1 shows the relative HIV-1 copy number per genomic equivalent of tissue DNA as determined using the Lightcycler-based quantitative-PCR technique. Note the very high level of viral DNA in the lymph node (one HIV-1 copy per 17 genomic-DNA equivalents) compared to the seminal vesicles (one HIV-1 copy per 456 genomic equivalents). The HIV-1 levels in the brain sites studied were between these two extremes.

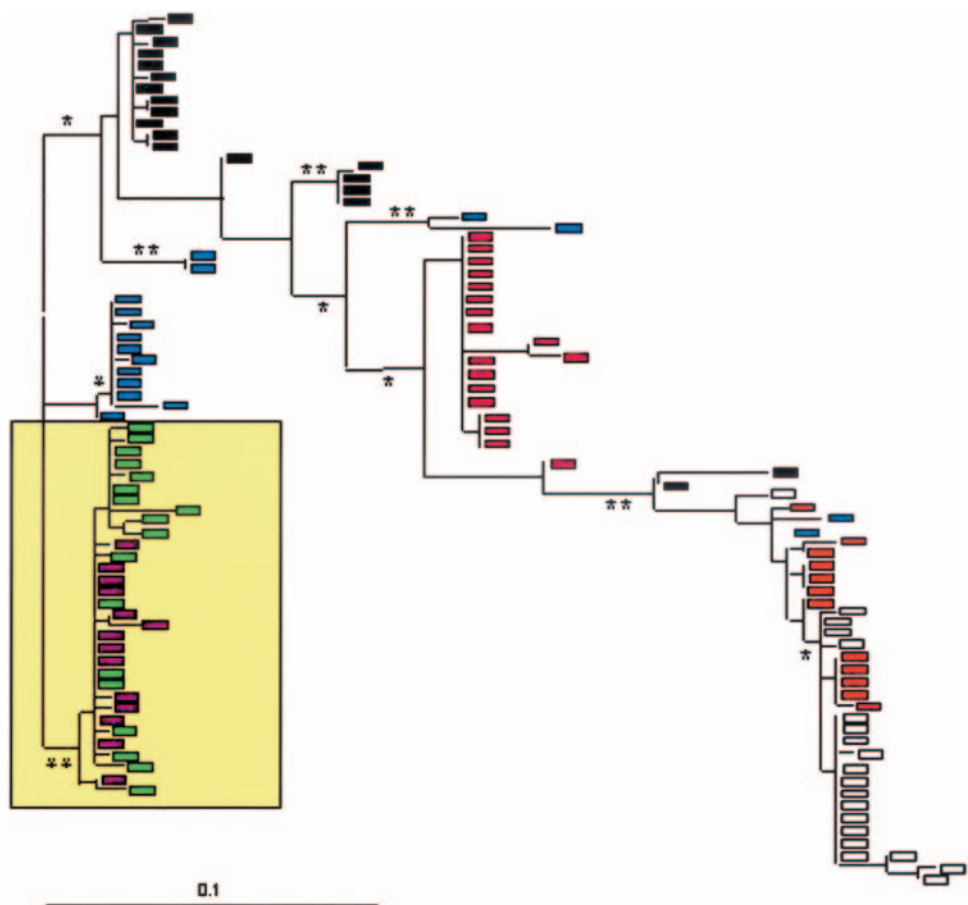


FIG. 2. Maximum likelihood unrooted genealogy of HIV-1 gp120 V3 isolated both from lymphoid (in yellow box) and site-specific brain tissue. The colored boxes represent different tissues: white, frontal lobe; blue, temporal lobe; red, occipital lobe; orange, spinal cord; black, meninges; green, iliac lymph node; magenta, seminal vesicles. Branch lengths were estimated with the HKY + Γ model, which was selected with a hierarchical likelihood ratio test (25). Branch lengths are drawn to scale, with the bar at the bottom indicating 0.1 nucleotide substitution per site. One asterisk along a branch represents support with a P value of < 0.005 in the zero-branch-length test (25) for the clade subtending that branch. Two asterisks indicate that the clade subtending the branch was also supported by a bootstrap value of $>75\%$.

Tissue localization of HIV. In order to test whether the current study would be an evaluation of HIV-1 within a primary macrophage compartment, immunohistochemical studies were performed on all of the tissue specimens used for DNA extraction and subsequent phylogenetic analysis. Figure 1 shows an example of the general findings present in all tissues evaluated. Within the brain, which showed classic changes as-

sociated with HAD, such as gliosis and perivascular cellular infiltration (Fig. 1A), the only cells found to express HIV p24 antigen were perivascular and parenchymal macrophages (Fig. 1B and C). As expected, CD68-expressing macrophages were found in abundance in both areas (Fig. 1C). Cells expressing HIV p24 in the lymphocyte-depleted lymph node and seminal vesicle were also predominantly macrophages (not shown). Forms of HIV-1 inserted into the brain macrophages were confirmed, using laser capture microdissection.

Phylogenetic analysis. As expected, in our study, the genealogies of HIV-1 sequences isolated both from lymphoid and nonlymphoid organs also support a clear separation of brain versus nonbrain sequences (Fig. 2). The maximum likelihood tree in Fig. 2 was inferred from the V3 alignment. The V1-V2 tree showed a strikingly similar topology (data not shown), which is interesting in light of the fact that previous studies have reported independent variation between the viral envelope variable domains early in infection within patients in their peripheral blood mononuclear cells (16). Within the monophyletic clade joining viral strains sampled from lymphoid organs, the largest monophyletic subclade consisted of 10 sequences

TABLE 2. Finkelstein test (10) for runs of identical events within biological sequences

Tissue of origin	No. of taxa ^a	Number of taxa in largest clade ^a	Proportion of sequences not in largest clade (%)	P value
Frontal lobe	20	15	25	7×10^{-7}
Temporal lobe	17	11	35.3	6×10^{-7}
Occipital lobe	20	19	5	3×10^{-7}
Meninges	19	12	36.8	9×10^{-7}
Spinal cord	12	5	58.3	5×10^{-4}
Lymph nodes	18	10	44.4	8×10^{-7}
Seminal vesicles	14	2	85.7	0.82

^a Refers to the number of distinct viral sequences present in the tree in Fig. 2.

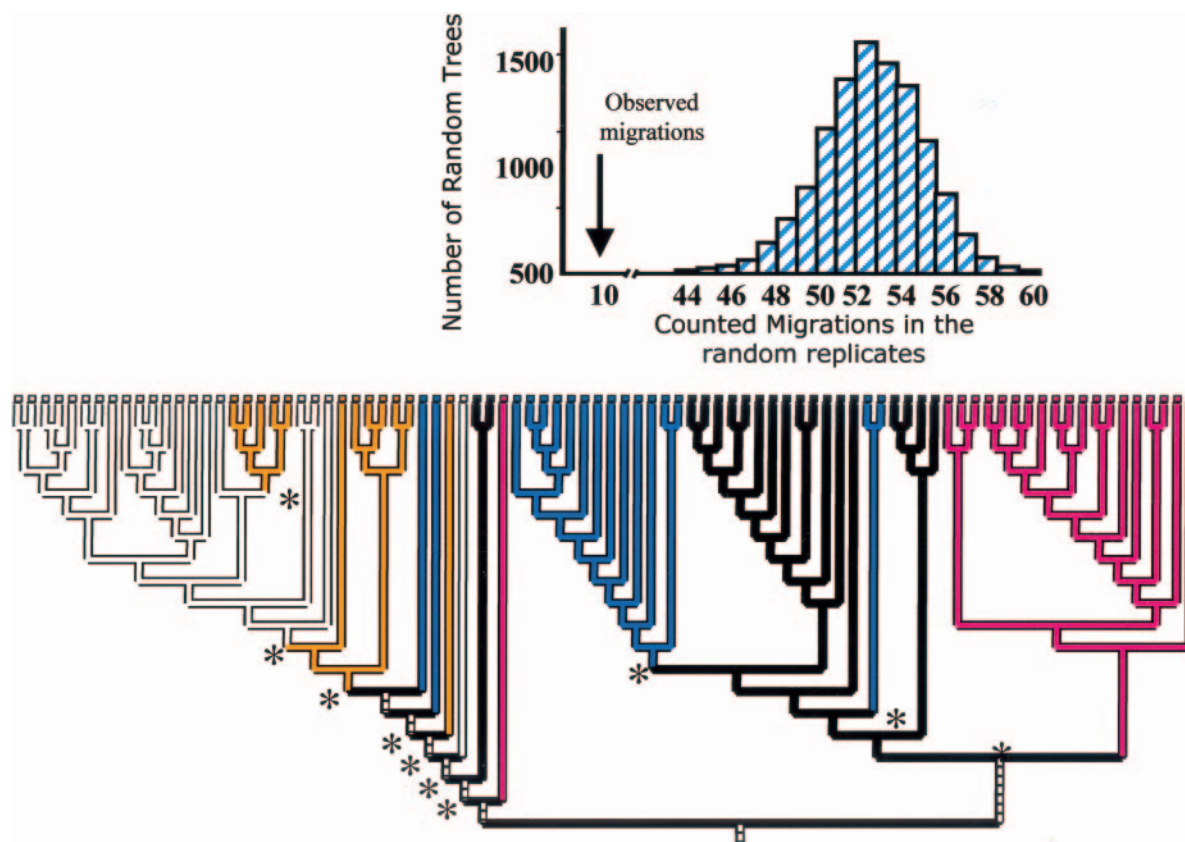


FIG. 3. Migration analysis of HIV-1 in the brain according to the Slatkin and Maddison test (31); maximum likelihood rooted cladogram of HIV-1 strains isolated from different brain regions. Colored boxes on the tips of the tree indicate the tissue of origin of the taxa according to the legend to Fig. 2. Internal nodes represent reconstructed ancestral sequences. The color of a branch indicates the tissue of origin of the top node (sequence) of that branch. Striped branches are those for which more than one parsimonious reconstruction is possible for the ancestral character. Different colors between two branches, indicated by an asterisk, imply that a migration must have occurred. Despite the uncertainty of the state assignments for some of the internal nodes, the total number of observed migrations, 10, can still be computed precisely with the Fitch algorithm (9). The distribution of tree lengths, i.e., the number of observed migrations in 10,000 trees generated by random joining-splitting from the original tree, is also shown.

isolated from lymph nodes with the remaining sequences, both from lymph nodes and seminal vesicles, intermingled in the clade. Similarly, within the nonlymphoid clade, viral strains isolated from meninges, spinal cord, and frontal, occipital, and temporal lobes appeared to be intermixed to various degrees. However, well-supported tissue-specific subclades of different sizes joining sequences isolated from each brain compartment were also present in the tree (Fig. 2). We evaluated the hypothesis that each of the largest monophyletic clades containing sequences isolated from a specific tissue resulted from runs of identical events; in other words, it was a random artifact due to sampling bias (8). Table 2 shows that, with the exception of seminal vesicles, each tissue sampled harbored a distinct HIV-1 subpopulation, but also that a significant proportion of sequences isolated from different tissues were intermixed, thus implying the trafficking of HIV-1 variants among different compartments.

Migration analysis. The Slatkin and Maddison test (31) was employed to investigate the hypothesis of gene flow among distinct subpopulations. The test supports the hypothesis of population subdivision within the brain. Figure 3 shows the cladogram, rooted by maximum likelihood, of the genealogy

given in Fig. 2 when only the brain sequences were considered. Although the tissue of origin for some of the hypothetical taxonomic units (internal nodes representing ancestral sequences) in the tree cannot be unambiguously assigned by the parsimony algorithm used (see Materials and Methods), the Fitch algorithm (9) ensures the exact computation of the tree length, i.e., in this case, the total number of observed migrations among different tissues. The number of migrations observed in the real data set, 10, was significantly less ($P < 0.0001$ in the Slatkin and Maddison test [Fig. 3]) than those expected from a random model where each virodeme is freely diffusible and equally likely to exchange virus with any other (Fig. 3). Moreover, the migration patterns of different HIV-1 subpopulations within the CNS, summarized by the viral migration network superimposed on the surface of the brain and portrayed in Fig. 4, appeared highly nonrandom. Three types of gene flow were observed: main flow out of a tissue (contributors), main flow into a tissue (receivers), and equal exchange of viruses between tissues (equilibrium). In this patient with HAD, the meninges and the occipital lobe cortex appeared to be major contributors, accounting for migrations to all other tissues. The temporal lobe (gray-white junction) and frontal

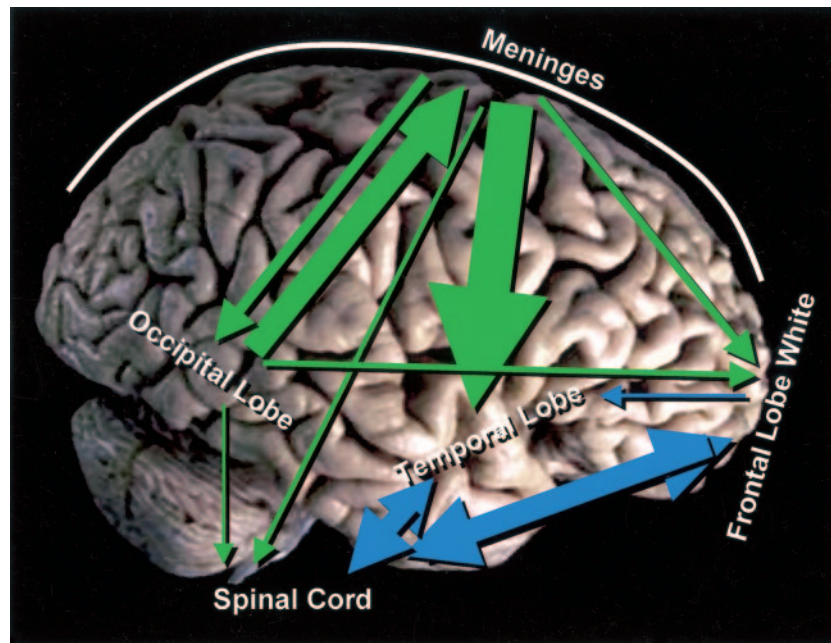


FIG. 4. Migration network of HIV-1 sequences in the CNS and tissues sampled superimposed on the surface of the brain. The meninges line on the surface of the brain is indicated by the white line. Arrow size corresponds to the percentage of total migrations observed in the tree from one tissue to another; the largest migration (21.5%) occurs from meninges to temporal lobe; the smallest migration shown (1.8%) occurs from occipital lobe to spinal cord. Migrations of less than 1% are not shown. Green arrows represent gene outflow from the meninges and occipital lobe, whereas blue arrows indicate gene flow associated with the temporal lobe, frontal lobe, and spinal cord. Double-headed arrows indicate the almost equal exchange of viruses between tissues. None of the percent changes shown are significantly different from zero. (The brain image is copyright-protected material used with permission of the authors and the University of Iowa's Virtual Hospital [<http://www.vh.org>].)

lobe (white matter) exchanged virus in approximately equal proportions with the spinal cord but also received, especially in the case of the temporal lobe, considerable gene flow from the meninges and occipital lobe.

HIV demographic modeling in the brain. Both nonparametric and parametric estimates of the effective HIV-1 population size over time in the brain are given in Fig. 5. The constant population size model cannot be rejected when compared with the exponential or logistic model of population growth in a likelihood ratio test (24, 34) ($P > 0.05$). Skyline plots and maximum likelihood estimates of demographic-model parameters were calculated using the genealogy of HIV-1 brain sequences shown in Fig. 3 and branch lengths inferred by enforcing local molecular clocks for each major tissue-specific clade (see Materials and Methods). Although the estimate of the demographic history may be biased by the presence of significant population subdivision in the CNS, both skyline plots and parametric estimates indicate constant population size, i.e., the effective number of HIV-1 infectious particles expressed as $N(t)\mu$ (see Materials and Methods), in the brain over time. In the present study, all sequences were from T-cell-depleted tissues, and based on charge analysis of the envelope V3 loop, were predicted to be macrophage tropic (+2 to +3 charge). Therefore, the network in Fig. 4 implies that macrophages are the cells responsible for HIV-1 replication, sequence evolution, and migration to different CNS anatomic compartments. The constant population size over time may indicate that the overall HIV-1 quasispecies infecting the CNS is at equilibrium and relatively isolated from other virodemes outside the brain.

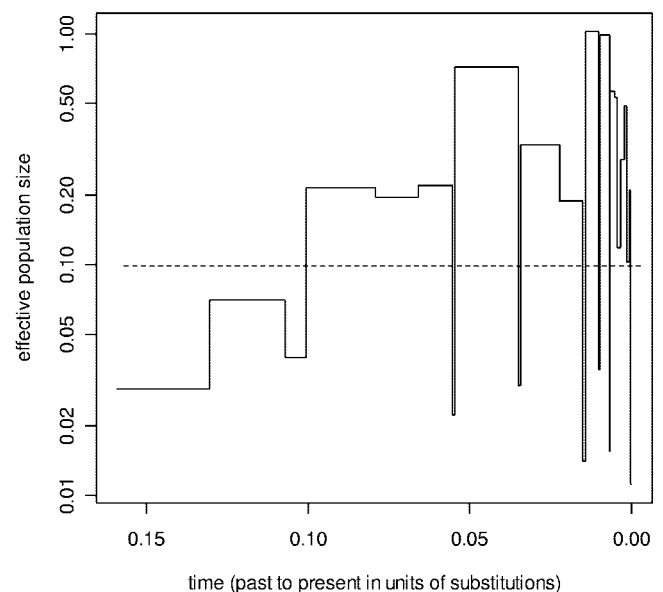


FIG. 5. Generalized Skyline plot of HIV-1 quasispecies in the brain. Nonparametric (solid line) and parametric (broken line) estimates of $N(t)\mu$ (15), the effective HIV-1 population size over time rescaled by the virus evolutionary rate, in the brain. The plot was inferred using the rooted cladogram in Fig. 3 with ultrametric branch lengths estimated by maximum likelihood (see Materials and Methods). Time, on the x axis, runs backward from the present ($t = 0$), and it is given as the number of nucleotide substitutions per site accumulated since the time of sampling.

TABLE 3. Molecular-clock analysis of HIV-1 brain sequences

Model ^a	log _e (Lk)	Likelihood ratio test performed ^b	P value	Relative rate estimated for local-clock model ^c
Global molecular clock (np = 87)	−1,060.2	No clock vs. global clock (dof = 86)	0.003	
Local molecular clocks (np = 91)	−1,053.1	Local clock vs. global clock (dof = 4)	0.007	Occipital lobe, 0.6 Meninges, 30.6 Temporal lobe, 99.2 Frontal lobe, 3.93
No molecular clock (np = 173)	−9,97.4	No clock vs. local clock (dof = 82)	0.02	

^a Different molecular-clock models were tested using the tree in Fig. 3 and the Yang et al. (44) codon substitution model M0 (29). The number of parameters (np) for each model is given in parentheses.
^b dof, degrees of freedom between the general and the null hypothesis.
^c The local-clock model assumes a base evolutionary rate of 1 along each branch of the tree in Fig. 3 but a different relative rate (estimated via maximum likelihood) for each of the largest tissue-specific monophyletic clades indicated in the table.

Local molecular clocks. The rate of evolution of HIV-1 in the sampled brain tissues was investigated by enforcing on the tree in Fig. 3 a local molecular clock for each of the largest monophyletic clades joining viral sequences isolated from meninges and occipital, temporal, and frontal lobes (Table 2), plus a baseline clock for the remaining part of the tree. Table 3 shows that while a global molecular clock for the HIV-1 brain sequences was significantly rejected, the hypothesis of five different local clocks could not be rejected at the 2% level. Even more interesting, the relative rate of evolution of the HIV-1 temporal lobe and meninges subpopulations, estimated by maximum likelihood, appeared to be about 100 and 30 times higher than the baseline rate of the tree (1 by default) (Table 3), while subpopulations in the other compartments seemed to evolve within the same order of magnitude. To investigate whether synonymous or nonsynonymous substitutions were responsible for the observed distortion of the global clock, we performed additional molecular-clock likelihood ratio tests using only the first plus second codon position and only the third codon position, respectively. In both cases, the global-clock model was rejected with highly significant *P* values (*P* < 0.001), suggesting that both synonymous and nonsynonymous replacement may be involved in the distortion of the molecular clock.

In this patient, the temporal lobe not only contained viruses originating from the meninges, it also appeared to be an important viral reservoir. Considering that damage to the temporal lobe is associated with symptoms typically reported in HAD, such as cognitive/motor disorders, information retrieval difficulties, and behavioral changes (3), the HIV-1 evolutionary dynamics observed in the current study could indeed be related to the evolution of dementia.

Positive-selection analysis. Maximum likelihood analysis did not detect any branch in the brain tree or any site under positive selection in the V3 or V1-V2 alignment. Two nested codon-based substitution models (see Materials and Methods), one assuming neutrality (null hypothesis) and the other assuming a proportion of sites under positive selection (alternative hypothesis), gave almost identical log likelihoods, −981.28 and −981.02, respectively, indicating that the hypothesis of sites under positive selection does not fit the data significantly better. We then checked whether significant selective pressure was present in the internal branches that separate different clusters. Model 1, assuming a different dN/dS ratio for each branch in the tree, has, as expected, the highest log likelihood score (−957.72), but its likelihood is not significantly different

than the one calculated for model 0 (−997.41; *P* > 0.05) or for model 2 (−994.59; *P* > 0.05). The comparison of model 0 and model 2 by the likelihood ratio test indicates that the null model (model 0) cannot be rejected at the 5% level. In conclusion, no branches in the tree seem to be under positive selection. We also compared average dN/dS ratios among maximum likelihood inferred ancestral sequences at the internal nodes of the significantly supported monophyletic clades of the tree in Fig. 3; dN/dS values were always significantly lower than 1 (data not shown), suggesting that the observed HIV-1 sequence diversity, both within and between subpopulations, is most probably the effect of random genetic drift after bottleneck events and of migration pathways in the brain.

V3 loop cross-pattern tissue analysis. Figure 6A again shows the V3 data set genealogy with sequences assigned to 11 distinct subgroups: 10 based on the 10 monophyletic clades within the brain and 1 including the sequences within the semen-lymph node cluster. Figure 6B is the alignment of the consensus sequences for each subgroup with the ancestral sequence at the midpoint root of the tree. Meninges (MENg1), spinal cord (SCg1 and SCg2), and frontal lobe (FLg1 and FLg2) subgroups share exclusive mutations at positions 16 (E16K), 32 (S32G), 61 (T61E), and 62 (K62N/E). The same five subgroups also show the loss of the third N-linked glycosylation site indicated in Fig. 6B, which is commonly conserved among different subtypes (13a). In addition, an extra phosphorylation site is present at positions 43 to 46 in the meninges (MENg3) and the semen-lymph node subgroups.

CTL epitope analysis. Figure 6B shows the HIV-1 sequence peptides encoding targets for T-cell reactivity (CTL epitopes) (4, 17, 21). Only one known epitope motif was identified in the sequences isolated from meninges, cortex, and peripheral organ, whereas no presenting CTL epitopes were found in the exposed part of the V3 loop in sequences from other brain regions. The single epitope in our case study (RPNNTTRKSI) is associated with HLA type B7 individuals and is commonly identified in both macrophage and T-cell-tropic viruses (data not shown). Finally, we screened our data sets and 229 other V3 loop sequences isolated from brain-derived tissues available at the Los Alamos Database for crest motifs. Such motifs have previously been found in sequences isolated from brains of HAD patients (6). Only a minority of the available database sequences (HIGPGRAF, 8/229; RIGPGRAF, 0/229; HIGPGSAI, 0/229; HLGPGSAF, 0/229; HIGPESAI, 0/229; HLGPE

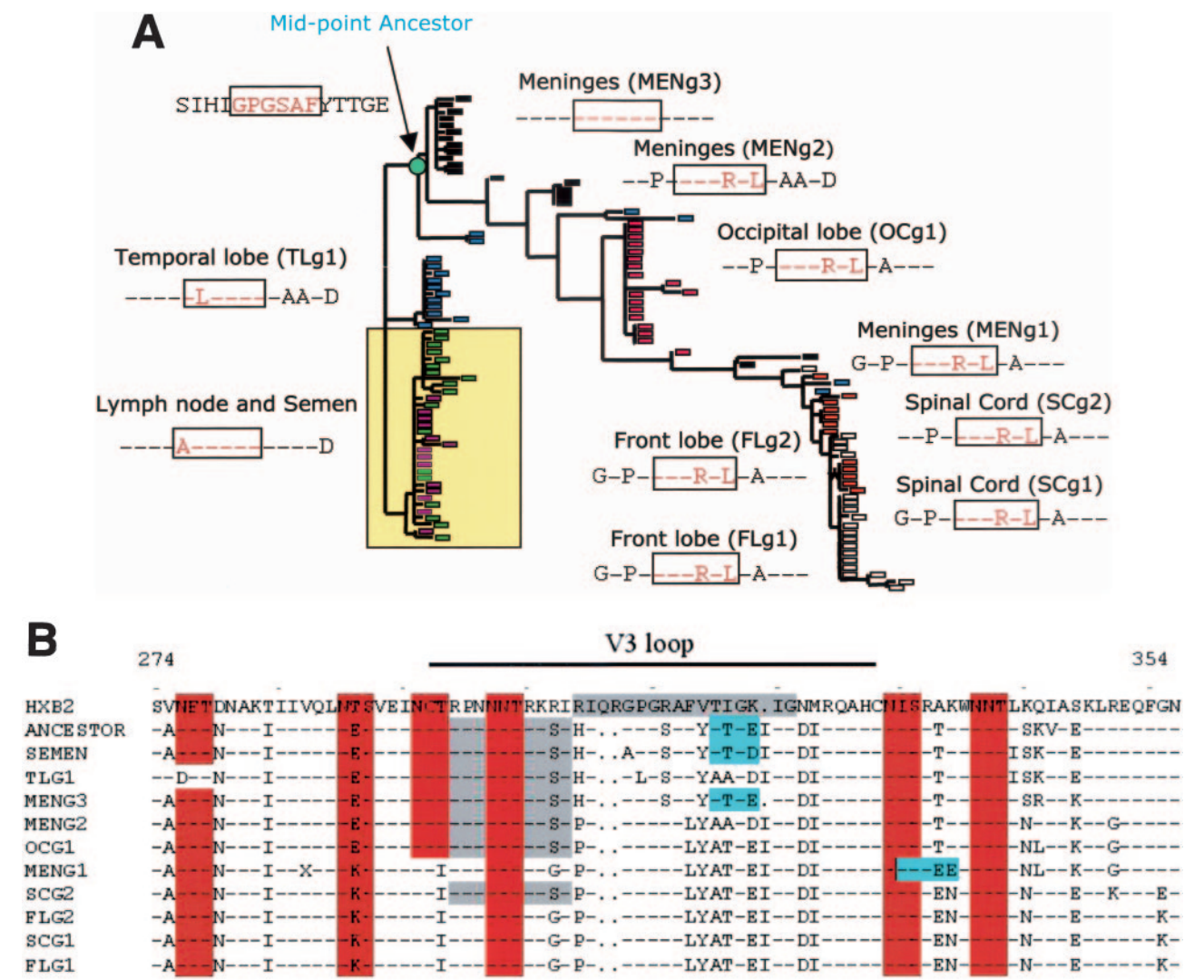


FIG. 6. Signature analysis of ancestral sequences; inferred HIV-1 V3 loop ancestral sequences from both lymphoid and nonlymphoid tissues. (A) HIV-1 V3 loop ancestral sequence of the midpoint ancestor (indicated by the green circle) of the tree from Fig. 2 (same color code) and ancestral sequences inferred for the MRCA of each tissue-specific monophyletic clade. Inferred amino acid sequences are shown next to the clade they belong to, with dashes indicating identical residues with respect to the midpoint ancestral sequence. The amino acids on the tip of the V3 loop are given in orange within the box. (B) V3 alignment of the midpoint ancestral sequence (ANCESTOR) and the consensus sequences of the strains belonging to each monophyletic clade indicated in panel A. The sequences are aligned to HXB2. Dots represent identical amino acid residues with respect to the midpoint ancestral sequence. The alignment is numbered according to the beginning of the HXB2 envelope protein by using the Sequence Locator Tool (<http://www.hiv.lanl.gov>). Putative N-glycosylation and protein kinase C phosphorylation sites along the sequences are highlighted in red and blue, respectively. Gray highlights indicate encoding targets for T-cell reactivity (CTL epitopes). In the brain sequences, only one CTL epitope motif that is associated with HLA type B7 is present. Within the gray highlighted region of HXB2, six CTL epitope motifs are found which are associated with a wide range of HLA types (A*0201, A11, A2, A3, D, Dd, H2, H2D, and B27).

SAI, 0/229; and YLRPGSAF, 0/229) and none of the sequences in our data set contained crest peptides.

DISCUSSION

A variety of studies have attempted to correlate the degree of HIV-1 infection in the central nervous system with HAD severity, and it is clear that an association between macrophage infiltration and HAD exists. Recent reports have suggested that blood-derived macrophages from HIV-infected patients with dementia are highly activated and express inflammatory mediators, as well as HIV gene products, and through some

unknown mechanism these inflammatory cells have the ability to break down the blood-brain barrier and contribute to neurologic disease (38). All the HIV-1 sequences in the present study were from T-cell-depleted tissues of a patient who died with severe HAD: an ideal data set to investigate the population dynamics of different macrophage-tropic HIV-1 subpopulations infecting different regions of the brain. Considering that 50 to 100 ng (15,000 to 30,000 cell equivalents) of genomic DNA were included in each PCR amplification run and that 1/100 cells on average had an HIV-1 copy (see Materials and Methods), the frequency of viral DNA genomes in each sample

was sufficient to allow the study of the HIV-1 genetic population dynamics.

In a study like the present one where brain biopsies from living patients are impractical, if not unethical, it is obviously impossible to obtain serial samples over time. The samples used in the analysis were collected at the time of death. The inferred trees represent a "snapshot" of HIV-1 genealogy since the time of the last viral bottleneck in those tissues (25, 34). Therefore, the MRCA at the root of the tree in Fig. 2 is unlikely to be the original common ancestor of the brain sequences. However, the inferred genealogy still contains information about the demographic history in the brain of this patient. The neutral evolution of the V3 loop, known to be usually under strong immune selection, could be explained by the T-cell depletion in the patient. On the other hand, the striking presence of quite different viral evolutionary clocks in the meninges and in the temporal lobe, instead of a uniform molecular clock as predicted by the neutral theory, may offer an explanation for the continuous activation/accumulation of macrophages in the brain during the AIDS phase. We suggest that after the failure of the immune system, newly produced viral variants, which would be rapidly cleared under normal conditions, begin to productively infect macrophages. To use Williams and Hickey's words, "in what becomes a self-amplifying cycle, infected macrophages activate other macrophages in the vicinity," and so on. In turn, more macrophages become infected and even more HIV-1 variants are produced and fixed in the viral population over time, giving rise to the increased evolutionary rate apparent in our data. "Macrophage dysregulation" (38), i.e., the accumulation, activation, and infection of macrophages in the temporal lobe that can eventually lead to neurodegeneration, is initiated by the production of new viral variants that the immune system is no longer able to control. At the same time, the cycle of continuous macrophage accumulation/infection is responsible for the increased evolutionary rate of the HIV-1 subpopulation in the temporal lobe. The scenario is loosely analogous to the finding of a 200-times-faster evolutionary rate in a human retrovirus evolving under neutrality that was observed as a consequence of a higher transmission rate within a network of injecting drug users (27).

As shown in Fig. 6B, the variation in the amino acid signature of the V3 loop leads to a loss or acquisition of different glycosylation sites. However, the phylodynamic patterns observed suggest that the main cause of the HIV-1 genetic variation in the brain is random genetic drift rather than positive selection. Neuronal injury is believed to result from both the direct effects of viral proteins and the indirect effects mediated by macrophage activation and resulting secretion of neurotoxic products (1). Carbohydrate binding proteins on the cell surface or macrophage endocytosis receptors interact with oligosaccharides on HIV-1 and can affect its infectivity. Therefore, the nonselective or random loss of glycosylation sites may also promote the macrophage activation cycle, thereby increasing the inflammatory response and the secretion of cytotoxic products, which are implicated in the onset of HAD (17).

The HLA haplotype of this patient is unknown; however, it is interesting that when screened for all known CTL epitope motifs, the sequences contained only a single epitope, which is associated with HLA type B7. In contrast, a known T-cell-

tropic isolate, HXB2, contains multiple CTL epitope motifs that correspond to a variety of HLA haplotypes and spans the tip of the V3 loop (amino acid positions 308 to 325). If the patient was indeed HLA B7, then the epitope has nonselectively disappeared during the evolution of the viruses. Alternatively, if the patient was not HLA B7, then no known CTL epitopes exist in this population of brain sequences. The sequences isolated from brain regions typically involved in HAD, spinal cord and frontal lobe white matter, do not contain any known epitopes and are less glycosylated. Additional glycosylation or reglycosylation has been shown to be related to a more efficient virus in macrophages (18), and the altering of an epitope has a quantitative as well as qualitative impact on CTL recognition (41). A macrophage-tropic variant of HIV-1 not containing dominant, cross-reactive T-cell epitopes might contribute to enhanced viral spread *in vivo*, especially in the context of primary infection. The recently described immune activation/viral-load "set point" study (7) demonstrated that individuals with primary infection develop an activation/viral-load set point that is essentially predictive of their long-term survival. Higher levels of activation predict worse long-term outcomes. One possibility, given the finding of V3 epitope-free HIV-1 strains in a HAD patient's brain, is that individuals infected by strains of virus missing T-cell epitopes would have a greater likelihood of developing a greater body burden of HIV-1. This initial spread, not regulated properly by T-cell function, would yield a significant load of long-lived HIV-infected macrophages capable of seeding HIV-1 throughout the body and relatively resistant to highly active antiretroviral therapy.

Finally, it is important to emphasize that generalizing from the present work should be done with caution because only one patient has been examined. However, our findings point out that a description of the viral phylodynamic (10) processes in the brain may be invaluable in helping to address the details of HIV-1 inpatient evolution and crucial for understanding the pathogenesis of HAD. Most importantly, the study suggests a plausible mechanism for macrophage dysregulation as a cause of neurodegeneration, underlining the need for effective antiretroviral treatments targeting not only T cells but also macrophages to successfully treat patients suffering from dementia and ultimately to eradicate HIV-1 infection.

ACKNOWLEDGMENTS

The authors are thankful to Ziheng Yang and the two anonymous reviewers for their valuable suggestions.

Marco Salemi is supported by grants AI065265 and HD32259 and the Department of Pediatrics of the University of Florida, Gainesville. Susanna Lamers is supported by grants NCI-U01-CA66529-12 and NCI U01-CA96230-04, grants for the AIDS and Cancer Specimen Resource, and NIH-R01 MH73510-01.

REFERENCES

1. Annunziata, P. 2003. Blood-brain barrier changes during invasion of the central nervous system by HIV-1. Old and new insights into the mechanism. *J. Neurol.* **250**:901-906.
2. Berger, J. R., L. Moskowitz, M. Fischl, and R. E. Kelley. 1987. Neurologic disease as the presenting manifestation of acquired immunodeficiency syndrome. *South Med. J.* **80**:683-686.
3. Blumer, D., and D. Benson. 1975. Personality changes with frontal and temporal lesions. In D. F. Benson and F. Blumer (ed.), *Psychiatric aspects of neurologic diseases*. Grune and Stratton, New York, N.Y.
4. Borrow, P., H. Lewicki, X. Wei, M. S. Horwitz, N. Pfeffer, H. Meyers, J. A. Nelson, J. E. Gairin, B. H. Hahn, M. B. Oldstone, and G. M. Shaw. 1997.

- Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat. Med.* **3**:205–211.
5. Boufassa, F., C. Bachmeyer, N. Carre, C. Deveau, A. Persoz, C. Jadand, D. Sereni, and D. Buchet. 1995. Influence of neurologic manifestations of primary human-immunodeficiency-virus infection disease progression. *J. Infect. Dis.* **171**:1190–1195.
 6. Chang, J., R. Jozwiak, B. Wang, T. Ng, Y. C. Ge, W. Bolton, D. E. Dwyer, C. Randle, R. Osborn, A. L. Cunningham, and N. K. Saksena. 1998. Unique HIV type 1 V3 region sequences derived from six different regions of brain: region-specific evolution within host determined quasispecies. *AIDS Res. Hum. Retrovir.* **14**:25–30.
 7. Deeks, S. G., C. M. Kitchen, L. Liu, H. Guo, R. Gascon, A. B. Narvaez, P. Hunt, J. N. Martin, J. O. Kahn, J. Levy, M. S. McGrath, and F. M. Hecht. 2004. Immune activation set-point during early HIV infection predicts subsequent CD4⁺ T cell changes independent of viral load. *Blood* **104**:942–947.
 8. Finkelstein, M., W. M. Fitch, C. A. Lanciani, and M. M. Miyamoto. 1998. Estimating probabilities of runs of identical events within biological sequences. *Mol. Biol. Evol.* **15**:470–472.
 9. Fitch, W. M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* **20**:401–416.
 10. Grenfell, B. T., O. G. Pybu, J. R. Gog, J. L. Wood, J. M. Daly, J. A. Mumford, and E. C. Holmes. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**:327–332.
 11. Kingman, J. F. C. 1982. The coalescent. *Stoch. Proc. Appl.* **13**:235–248.
 12. Korber, B., and G. Myers. 1992. Signature pattern analysis: a method for assessing viral sequence relatedness. *AIDS Res. Hum. Retrovir.* **8**:1549–1560.
 13. Korber, B. T., K. J. Kunstman, B. K. Patterson, M. Furtado, M. M. McEvilly, R. Levy, and S. M. Wolinsky. 1994. Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. *J. Virol.* **68**:7467–7481.
 - 13a. National Institutes of Health, U.S. Department of Health and Human Services. 2002. HIV Sequence Compendium. [Online.] <http://hiv-web.lanl.gov/content/hiv-db/HTML/2002compendium.html>.
 14. Lamers, S., S. Beason, L. Dunlap, R. Compton, and M. Salemi. 2002. HIV-base: a PC/Windows-based software offering storage and querying power for locally held HIV-1 genetic, experimental and clinical data. *Bioinformatics* **20**:436–438.
 15. Lamers, S. L., J. W. Sleasman, and M. Goodenow. 1996. A model for alignment of env V1 and V2 hypervariable domains from human and simian immunodeficiency viruses. *AIDS Res. Hum. Retrovir.* **12**:1169–1178.
 16. Lamers, S. L., J. W. Sleasman, K. A. Barrie, S. M. Pomeroy, D. J. Barrett, and M. M. Goodenow. 1993. Independent variation and positive selection in env V1 and V2 domains within maternal-infant strains of human immunodeficiency virus type 1 in vivo. *J. Virol.* **67**:3951–3960.
 17. Luo, L., Y. Li, J. S. Chang, S. Y. Cho, T. Y. Kim, M. J. Choi, H. S. Cheong, H. J. Kim, H. J. Ahn, M. K. Min, B. H. Chun, S. M. Jung, S. G. Woo, S. Y. Park, and C. Y. Kang. 1998. Induction of V3-specific cytotoxic T lymphocyte responses by HIV gag particles carrying multiple immunodominant V3 epitopes of gp120. *Virology* **240**:316–325.
 18. Ly, A., and L. Stammatos. 2000. V2 loop glycosylation of the human immunodeficiency virus type 1 SF162 envelope facilitates interaction of this protein with CD4 and CCR5 receptors and protects the virus from neutralization by anti-V3 loop and anti-CD4 binding site antibodies. *J. Virol.* **74**:6769–6776.
 19. Mack, K. D., X. Jin, S. Yu, R. Wei, L. Kapp, C. Green, B. Herndier, N. W. Abbey, A. Elbaggari, Y. Liu, and M. S. McGrath. 2003. HIV insertions within and proximal to host cell genes are a common finding in tissues containing high levels of HIV DNA and macrophage-associated p24 antigen. *J. Acquir. Immune Defic. Syndr.* **33**:308–320.
 20. Maddison, W. P., and D. R. Maddison. 2003. *MacClade version 4: analysis of phylogeny and character evolution*. Sinauer Associates, Sunderland, Mass.
 21. Novitsky, V., N. Tybak, M. F. McLane, P. Gilbert, P. Chigwedere, I. Klein, S. Gaolekwe, S. Y. Chanbe, T. Peter, I. Thior, T. Ndung'u, F. Vannberg, B. T. Foley, R. Marlink, T. H. Lee, and M. Essex. 2001. Identification of human immunodeficiency virus type 1 subtype C Gag-, Tat-, Rev-, and Nef-specific ELISPOT-based cytotoxic T-lymphocyte responses for AIDS vaccine design. *J. Virol.* **75**:9210–9228.
 22. Ohagen, A., A. Devitt, K. J. Kunstman, P. R. Gorry, P. P. Rose, B. Korber, J. Taylor, R. Levy, R. L. Murphy, S. M. Wolinsky, and D. Gabuzda. 2003. Genetic and functional analysis of full-length human immunodeficiency virus type 1 env genes derived from brain and blood of patients with AIDS. *J. Virol.* **77**:12336–12345.
 23. Paton, P., H. Poly, P. M. Gonnard, J. C. Tardy, J. Fontana, K. Kindbeiter, R. Tete, and J. J. Madjar. 1990. Acute meningoradiculitis concomitant with seroconversion to human immunodeficiency virus type 1. *Res. Virol.* **141**:427–433.
 24. Pybus, O. G., and A. Rambaut. 2002. GENIE: estimating demographic history from molecular phylogenies. *Bioinformatics* **18**:1404–1405.
 25. Pybus, O. G., A. Rambaut, and P. H. Harvey. 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* **155**:1429–1437.
 26. Saksena, N. K., B. Wang, Y. C. Ge, J. Chang, D. E. Dwyer, S. H. Xiang, D. R. Packham, C. Randle, and A. L. Cunningham. 1997. Region-specific changes, gene duplications, and random deletions in the nef gene from HIV type 1-infected brain tissues and blood of a demented patient. *AIDS Res. Hum. Retrovir.* **13**:111–116.
 27. Salemi, M., M. Lewis, J. F. Egan, W. W. Hall, J. Desmyter, and A.-M. Vandamme. 1999. Different population dynamics of human T cell lymphotropic virus type II in intravenous drug users compared with endemically infected tribes. *Proc. Natl. Acad. Sci. USA* **96**:13253–13258.
 28. Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* **14**:1218–1231.
 29. Shapshak, P., D. M. Segal, K. A. Crandall, R. K. Fujimura, B. T. Zhang, K. Q. Xin, K. Okuda, C. K. Petito, C. Eisdorfer, and K. Goodkin. 1999. Independent evolution of HIV type 1 in different brain regions. *AIDS Res. Hum. Retrovir.* **15**:811–820.
 30. Silver, S., L. Ayers, J. Orenstein, and M. S. McGrath. 2003. AIDS and CancerSpecimen Resource: a resource for your research. Presented at the ISBER Annual Meeting, Biodiversity and International Regulations, Philadelphia, Pa.
 31. Slatkin, M., and W. P. Maddison. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* **123**:401–416.
 32. Smit, T. K., B. J. Brew, W. Tourtellotte, S. Morgello, B. B. Gelman, and N. K. Saksena. 2004. Independent evolution of human immunodeficiency virus (HIV) drug resistance mutations in diverse areas of the brain in HIV-infected patients, with and without dementia, on antiretroviral treatment. *J. Virol.* **78**:10133–10148.
 33. Smit, T. K., B. Wang, T. Ng, R. Osborne, B. Brew, and N. K. Saksena. 2001. Varied tropism of HIV-1 isolates derived from different regions of adult brain cortex discriminate between patients with and without AIDS dementia complex (ADC): evidence for neurotropic HIV variants. *Virology* **279**:509–526.
 34. Strimmer, K., and O. G. Pybus. 2001. Exploring the demographic history of DNA sequences using the generalized skyline plot. *Mol. Biol. Evol.* **18**:2298–2305.
 35. Swofford, L., and J. Sullivan. 2003. Phylogeny inference based on parsimony and other methods with PAUP*. p. 160–206. *In* The phylogenetic handbook—a practical approach to DNA and protein phylogeny, 1st ed. Cambridge University Press, New York, N.Y.
 36. Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly. 1997. Inferring coalescence times from DNA sequence data. *Genetics* **145**:505–518.
 37. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
 38. Vanhems, P., C. Dassa, J. Lambert, D. A. Cooper, L. Perrin, J. Vizzard, B. Hirschel, S. Kinloch-de Loes, A. Carr, and R. Allard. 1999. Comprehensive classification of symptoms and signs reported among 218 patients with acute HIV-1 infection. *J. Acquir. Immune Defic. Syndr.* **21**:99–106.
 39. Williams, K. C., and W. F. Hickey. 2002. Central nervous system damage, monocytes and macrophages, and neurological disorders in AIDS. *Annu. Rev. Neurosci.* **25**:537–556.
 40. Williams, K. C., S. Corey, S. V. Westmoreland, D. Pauley, H. Knight, C. deBakker, X. Alvarez, and A. A. Lackner. 2001. Perivascular macrophages are the primary cell type productively infected by the simian immunodeficiency virus in the brain of macaques: implication for the neuropathogenesis of AIDS. *J. Exp. Med.* **193**:905–915.
 41. Wolinsky, S. M., B. T. Korber, A. U. Neumann, M. Daniels, K. J. Kunstman, A. J. Whetsell, M. R. Furtado, Y. Cao, D. D. Ho, and J. T. Safrin. 1996. Evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* **272**:537–542.
 42. Wong, J. K., C. C. Ignacio, F. Torriani, D. Havlir, N. J. Fitch, and D. D. Richman. 1997. In vivo compartmentalization of human immunodeficiency virus: evidence from the examination of pol sequences from autopsy tissues. *Virology* **71**:2059–2071.
 43. Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biol. Sci.* **13**:555–556.
 44. Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen. 2000. Codon substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**:431–449.