Genomic Evolution, Patterns of Global Dissemination, and Interspecies Transmission of Human and Simian T-cell Leukemia/Lymphotropic Viruses

Jill Pecon Slattery, Genoveffa Franchini and Antoine Gessain

Genome Res. 1999 9: 525-540
Access the most recent version at doi:10.1101/gr.9.6.525

Review

Genomic Evolution, Patterns of Global Dissemination, and Interspecies Transmission of Human and Simian T-cell Leukemia/Lymphotropic Viruses

Jill Pecon Slattery,[1,4] Genoveffa Franchini,[2] and Antoine Gessain[3]

[1]*Laboratory of Genomic Diversity, National Cancer Institute–Frederick Research Development Center (NCI-FCRDC) Frederick, Maryland 21702 USA;* [2] *Bethesda Research Laboratories (BRL) NCI, National Institutes of Health (NIH), Bethesda, Maryland 20892 USA;* [3]*Unite d'Oncologie Virale, Institut Pasteur, 5724 Paris CEDEX 15, France*

Using both *env* and long terminal repeat (LTR) sequences, with maximal representation of genetic diversity within primate strains, we revise and expand the unique evolutionary history of human and simian T-cell leukemia/lymphotropic viruses (HTLV/STLV). Based on the robust application of three different phylogenetic algorithms of minimum evolution–neighbor joining, maximum parsimony, and maximum likelihood, we address overall levels of genetic diversity, specific rates of mutation within and between different regions of the viral genome, relatedness among viral strains from geographically diverse regions, and estimation of the pattern of divergence of the virus into extant lineages. Despite broad genomic similarities, type I and type II viruses do not share concordant evolutionary histories. HTLV-I/STLV-I are united through distinct phylogeographic patterns, infection of 20 primate species, multiple episodes of interspecies transmission, and exhibition of a range in levels of genetic divergence. In contrast, type II viruses are isolated from only two species (*Homo sapiens* and *Pan paniscus*) and are paradoxically endemic to both Amerindian tribes of the New World and human Pygmy villagers in Africa. Furthermore, HTLV-II is spreading rapidly through new host populations of intravenous drug users. Despite such clearly disparate host populations, the resultant HTLV-II/STLV-II phylogeny exhibits little phylogeographic concordance and indicates low levels of transcontinental genetic differentiation. Together, these patterns generate a model of HTLV/STLV emergence marked by an ancient ancestry, differential rates of divergence, and continued global expansion.

Emerging viral pathogens are those that have either invaded a new host species or expanded into new geographic populations of host species. As represented by the global prevalence of human immunodeficiency virus (HIV) that has occurred in <20 years, or the massive Spanish influenza outbreak of the 1920s, viral pathogens can be highly transmissible and virulent. At first glance, these episodes appear to be unpredictable. Yet, whether viral, bacterial, or parasitic, a close examination reveals a common trend whereby a pre-existing pathogen becomes selectively activated by changing environmental conditions (for review, see Morse 1995). At this point, the pathogen propagates within a host and may increase in prevalence via interspecies transmission. Thus, a viral pathogen may be benign while residing within a "reservoir" species, yet on entering a new host, increase in virulence. Efforts to control and regulate outbreaks rely on epidemiological research of each event and involve defining patterns of dissemination, virulence, and mode of transmission between individuals. Such information provides the basis for subsequent interdisciplinary considerations encompassing virology, cell biology, immunology, and pharmacology in devising effective treatment strategies. Here, using genetic variation of the human T-cell leukemia/lymphotropic viruses and related simian retroviruses, we present an overview of the application of a powerful tool in countering emergent pathogens–molecular phylogenetics.

Forming a link between evolutionary history and epidemiology, molecular phylogenetics addresses five major aspects integral to viral emergence. First, the genetic diversity present within the virus is estimated by comparing among all known viral strains. Second, estimates of the pattern and rate of mutation of each gene within the virus can be examined. Third, the identity of the causative viral strain of each new outbreak can be ascertained and compared with previously described viruses. By linking viral strains, the corresponding host species or population is also identified, forming the basis of determining mode of transmission. Fourth, using such a comparative approach, the geographic as well as evolutionary origin of different viral strains can be inferred through phylogenetic associations. Fifth, depiction of the mutation rate

[4]Corresponding author.
E-MAIL Slattery@mail.ncifcrf.gov; (FAX) 301-846-6327.

of each gene, and the extent of genetic similarity uniting viral strains, is essential to devising an effective drug treatment and vaccine development.

Successful application of phylogenetic analysis is best represented in the research of human and simian T-cell leukemia/lymphotropic virus (HTLV/STLV). Transmission of HTLV/STLV occurs by sexual contact (Murphy et al. 1989; Vitek et al. 1995), from mother to child by breast feeding (Hino et al. 1985; Vitek et al. 1995), and through blood transfusion or contact (Okochi and Sato 1984). These viruses possesses unique pathogenicity, patterns of dissemination, global patterns of endemicity, and mutation processes unlike any known retrovirus.

As each new viral strain is identified, a remarkable pattern of distribution demarcates type I and type II viruses (Fig. 1). HTLV-I is distributed worldwide [with 15–20 million people estimated to be infected with the virus (Gessain 1996)] with local regions of high prevalence including southern Japan, intertropical Africa, the Caribbean, and some areas within South America, the Middle East, and Melanesia (Fig. 1). HTLV-I is now recognized as the causative agent of adult T-cell leukemia/lymphoma (ATLL) (Poiesz et al. 1980; Hinuma et al. 1982), a malignant lymphoma of CD4 cells causing high mortality. Another disease caused by HTLV-I infection is the chronic, debilitating neurological disorder tropical spastic paraparesis/HTLV-1-associated myelopathy (TSP/HAM) (Gessain et al. 1985). HTLV-I seemingly is linked with additional diseases such as the rare infective dermatitis (La Grenade et al. 1990) and to a lesser extent with some cases of polymyositis, both in Jamaica (Morgan et al. 1989), and cases of uveitis in young adults (Mochizuki et al. 1996) and rheumatoid arthritis (Nishioka 1996) in Japan.

Less clear are disease associations of HTLV-II. Originally identified from a patient with a variant form of hairy T-cell leukemia (Kalyanaraman et al. 1982), HTLV-II is as yet only loosely correlated with rare neurological diseases resembling TSP/HAM (Jacobson et al. 1993; Murphy et al. 1997a) or other opportunistic infections attributable to a compromised immune system of patients harboring HTLV-II (Modahl et al. 1997; Murphy et al. 1997b). In sharp contrast to HTLV-I, type II viruses exhibit a markedly different pattern of distribution. Originally, the virus was thought to be a New World pathogen restricted to isolated Amerindian tribes throughout North and South America (Fig. 1). High prevalence in isolated ethnic groups, such as Guaymi in Panama (Lairmore et al. 1990; Pardi et al. 1995); Cayapo, Kraho, and Kaxuyana of Brazil (Maloney et al. 1992; Biggar et al. 1996); Toba, Mataco-Mataguayo, and Mapuche of Argentina (Biglione et al. 1993, 1999; Ferrer et al. 1993, 1996); Pume, Guahibo, and Yaruro of Venezuela (Echeverria de Perez et al. 1993; Leon-Ponte et al. 1998); Wayuu of Colombia (Switzer et al. 1995a); and other less isolated tribes in New Mexico (Hjelle et al. 1990) and Florida (Levine et al. 1993), suggests the virus was brought into the New World by ancient human migrations of 10,000–20,000

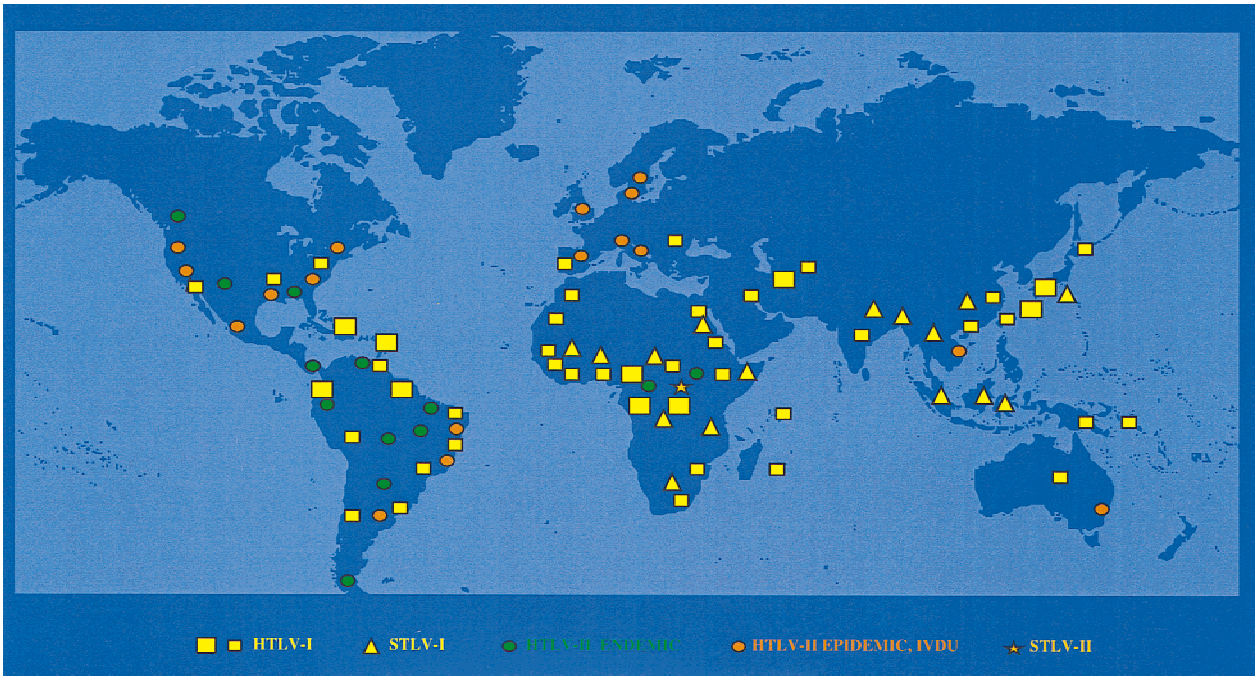

Figure 1 Map of distribution of HTLV/STLV viral isolates of known geographic origin. (Small rectangles) HTLV-I; (large rectangles) HTLV-I high endemicity; (triangles) STLV-I; (green ovals) HTLV-II endemic ethnic groups; (red ovals) HTLV-II IVDU; (star) STLV-II.

years ago (Maloney et al. 1992; Switzer et al. 1996) and is maintained and transmitted between generations by heterosexual contact and cultural practices such as communal breast-feeding (Black et al. 1994; Vitek et al. 1995). The discovery of diverse strains of the virus in different tribes of the oldest African human ethnic group, the Pygmy of Cameroon (Gessain et al. 1995), Central African Republic (Giri et al. 1997), and Democratic Republic of Congo (Zaire) (Goubau et al. 1992, 1993; Vandamme et al. 1998a), and from isolated families in Gabon (Tuppin et al. 1996) contradicts the view that type II viruses are exclusive to the New World. In addition, the very recent invasion of new host populations of intravenous drug users (IVDU) of Europe (Salemi et al. 1996, 1998a) and North America (Biggar et al. 1991) indicates a changing epidemiology for HTLV-II from highly localized to potentially global distribution. Moreover, molecular epidemiological characterization of North American IVDU indicates at least two episodes of invasion, with correspondingly different patterns of subsequent dissemination (Murphy et al. 1998).

Subsequent investigations of nonhuman primates revealed simian forms of the viruses identified as STLV-I and STLV-II. STLV-I is confined to Africa and Asia among at least 19 species of Old World primates, but little is known of the seroprevalence of the virus in natural populations within each simian species (Fig. 1). ATLL-like disease has been described in some STLV-I infected animals (summarized in International Agency for Research on Cancer 1996). The distribution of STLV-II is virtually unknown. A notable exception is STLV-II, isolated independently by Giri et al. (1994) and Liu et al. (1994) from captive bonobo chimps (*Pan paniscus*) taken from Zaire in central Africa (Fig. 1).

A third viral type, a highly divergent strain isolated from *Papio hamadryas* in Eritrea (STLV-PH969), is not affiliated with type I or II, and remains the sole representative of the monotypic PTLV-L (Goubau et al. 1994; Van Brussel et al. 1996, 1998).

Methods in Phylogenetic Analysis of Nucleotide Sequences

Nearly all published phylogenetic analyses of HTLV/STLV are derived from nucleotide data. As such, molecular phylogenetic analysis typically employs three algorithms: distance-based minimum evolution (ME), maximum likelihood (ML), and maximum parsimony (MP) (for review, see Swofford et al. 1996). Each method is derived from a suite of evolutionary assumptions that are not necessarily compatible. As such, the methods vary in performance, accuracy, and precision, and the relative strengths and weaknesses of each have been discussed elsewhere (Sourdis and Nei 1988; Hasegawa and Fujiwara 1993; Huelsenbeck and Hillis 1993; Kuhner and Felsenstein 1994; Tateno et al. 1994). However, all methods are alike in that they reconstruct phylogenetic associations into a tree and then test the tree under an explicit optimality criterion. Concordance among the phylogenetic trees derived from each method is interpreted as evidence that the particular genetic marker is consistent in estimation of the true phylogeny. Distance-based methods compute a genetic measure between each pair of taxa (e.g., viral strains). The resulting matrix is used as input for analysis by ME–neighbor-joining (NJ) method. The final tree is selected after a heuristic search and branch rearrangement results in minimization of the error (or fit) between the pairwise distance estimates and the final tree. ML is the most statistically robust and computer-intensive method and searches for a tree of the greatest probability of occurrence given the data and an explicit model of substitution. Lastly, MP transforms the data into character states and searches for the tree topology that invokes the least number of changes (or steps) under the optimality criterion that the shortest tree is the best estimate of the true phylogeny.

Additional a posteriori resampling methods of bootstrap and jackknife are used to test the robustness of the phylogeny. Each method indicates the degree to which the phylogenetic signal is consistent, reliable, and randomly distributed throughout the genetic data. A bootstrap analysis is an iterative process that creates multiple, randomized artificial data sets from the original input and repeats the phylogenetic reconstruction for each. In general, 100 iterations are a sufficient representation of the consistency of the data to repeat the same topology (Hillis and Bull 1993). Jackknife analyses remove randomly a specified proportion of the data (either sites or taxa) from the original input and retest the phylogenetic relationships with the remaining subset.

Application of molecular phylogenetic methods in viral emergence imposes an evolutionary context to observed viral associations. Consequently, selection of the appropriate genetic marker of evolutionary divergence among viral strains under consideration is of paramount importance. A basic implicit assumption is that mutations accumulate in a manner roughly proportional with time and at an equivalent rate among viral strains analyzed. Consequently, evolving gene segments become uninformative once viral divergence times exceed the point at which the nucleotide sequences are completely randomized with respect to each other. In the simplest case, based on equilibrium frequencies of A, C, G, and T, sequences are considered uninformative if divergence exceeds 25% (Jukes and Cantor 1969).

Other models of nucleotide substitution increase the above estimation of sequence divergence limitations by incorporating among-site rate variation, tran-

sition/tranversion ratio, and insertion and deletions events, in addition to nucleotide frequencies, (for review, see Swofford et al. 1996; Li 1997). These parameters are empirically derived based on the diversity present in the sampling of viral gene regions analyzed. Prior investigation and estimation of these parameters is essential for selection of the appropriate model of substitution for phylogenetic reconstruction. Among-site rate variation is typified in nucleotide sequences marked by highly conserved motifs interspersed with more variable regions. With coding genomic regions, additional substitution rate differences occur among the three codon positions. In particular, third position changes are synonymous and are first to achieve saturation and loss of phylogenetic signal. Variable substitution rates among sites are a consequence of differential selective constraints and may result in underestimation of sequence divergence (Gillespie 1986; Takahata 1991) leading to errors in phylogenetic reconstruction (Yang and Kumar 1996). Another consideration is that nucleotide substitution patterns vary between transition and transversion changes. In general, closely related sequences are characterized by a high transition/tranversion ratio that subsequently declines with increased divergence times (Adkins and Hunicutt 1994), an effect that may be biased by among-site rate variation (Wakeley 1994). The relative importance of insertion/deletion events, represented by gaps among sequences, may be either ignored or incorporated in the phylogenetic analysis depending on the model employed.

Phylogenetic inferences in viral emergence are strongly influenced by the specific viral strains included in the analysis. With HTLV/STLV, few viral strains are sequenced in entirety and most are identified by partial sequences from either the LTR, *env*, *pol*, or *pX* region. The viral composition of published sequences is not consistent, with few viral strains sequenced across more than one gene segment. Thus, discrepancies in evolutionary analyses of HTLV/STLV arise between studies merely because of choice of genetic marker, selection of viral strains for analyses, and identification of the appropriate model of substitution for phylogenetic reconstruction. In the present analysis, we opted for maximal representation of available sequences and selected a portion of *env* (452 bp) for an analysis of type I, type II, and PTLV-L strains; LTR (519 bp) for analyses of HTLV-I/STLV-I; and LTR (417 bp) for HTLV-II/STLV-II.

HTLV/STLV Possesses Distinctive Genetic Characteristics

Unlike other retroviruses, which have high mutation rates leading to quasi-speciation because of high replication levels and lack of a proofreading mechanism of the viral polymerase (Katz and Skulka 1990; Williams and Loeb 1992), HTLV/STLV exhibits unusually low levels of diversity within individuals (Gessain et al. 1992). The observed paradox of long periods of latency in conjunction with high proviral load (Watell et al. 1996) yet low levels of intraindividual genetic variation is attributed to clonal expansion of HTLV-harboring cells for both type I and type II (Wattel et al. 1995; Cimarelli et al. 1996; Etoh et al. 1997). It is postulated that on infection, the virus undergoes a period of replication via reverse transcription but all subsequent proliferation occurs via clonal expansion of infected T-cells. Consequently, viral substitution rates are speculated to be regulated in part by cell division of the host species (Wattel et al. 1996).

Analyses of different genes across diverse viral strains indicate varying levels of nucleotide substitution. Assuming a rough molecular clock (Zuckerkandl and Pauling 1965), in which the number of accumulated substitutions is roughly proportional with the time since two viral strains last shared a common ancestor, differential substitution rates among genomic regions are instrumental in defining hierarchical levels within viral evolution. For example, a simple comparison between the most diverse strains of type I and type II indicate LTR as the most variable genomic region, followed by *env*, with the *tax/rex* genes as the most conserved (Table 1). Most likely, high values between type I, type II, and PTLV-L ranging from 44.3% to 70.1% reflect saturation and loss of phylogenetic signal with LTR. However, *env* and *tax* appear to be more useful for between-type comparisons ranging in value between 28%–45% and 28%–35%, respectively.

At present, estimation of nucleotide substitution rates indicates a rate less than other retroviruses. Estimates from the LTR, 1.08×10^{-4}–2.7×10^{-5}, were based on the introduction of HTLV-II into IVDU in Europe 25 years ago (Salemi et al. 1998a). A lower value for HTLV-I was determined from a consensus of *gag*, *pol*, *env*, and *pX* sequences as $0.4\% \times 10^{-7}$ – $6.8\% \times 10^{-7}$ (Yanagihara et al. 1995) compared with an LTR value of 1.25×10^{-5} to 5×10^{-5} derived from the introduction of Japanese strains into Peru ~400–100 years ago (Van Dooren et al. 1998). These values are two to four orders of magnitude less than that for HIV (Suzuki and Gojobori 1998) and are consistent with the hypothesis of reduced mutation rate linked with clonal replication of host T-cells (Wattel et al. 1995).

Testing the Hypothesis of Host Specificity with HTLV-I/STLV-I Strains

One of the initial concerns upon the discovery of related strains from nonhuman primates was the concept of host–virus coevolution. Under this hypothesis,

Table 1. Genetic Distance Estimates Among Representative Viral Strains from Type I, Type II, and PTLV-L

	Type I					
	Atk			Mto-te4		
	LTR	*env*	*tax/rex*	LTR	*env*	*tax/rex*
Atk	0	0	0			
Mto-te4	24.9	12.4	8.9	0	0	0
Ppa-7qB	60.4	28.3	32.3	61.2	42.2	34.0
Efe-2	48.3	39.4	29.5	42.7	40.1	31.5
Nra	48.0	40.9	28.7	38.7	40.8	29.5
MoT	44.3	42.1	29.8	45.3	40.3	30.5
Pha-969	46.6	44.9	29.8	63.6	45.0	30.8

	Type II												PTLV-L		
	Ppa-7qB			Efe-2			Nra			MoT			Pha-696		
	LTR	*env*	*tax/rex*	LTR	*env*	*tax/rex*	LTR	*env*	*tax/rex*	LTR	*env*	*tax/rex*	LTR	*env*	*tax/rex*
Atk															
Mto-te4															
Ppa-7qB	0	0	0												
Efe-2	24.9	28.5	20.4	0	0	0									
Nra	29.4	28.8	20.8	8.6	7.6	4.8	0	0	0						
MoT	29.3	28.3	20.4	10.1	6.7	5.1	9.4	6.7	3.9	0	0	0			
Pha-969	70.1	43.1	35.7	59.8	37.9	31.1	63.0	39.6	30.8	56.5	38.1	32.7	0	0	0

Models of substitution: Tamura–Nei (1993) (LTR: 858 bp), Kimura (1980) two-parameter (*env:* 1430 bp), and Jukes–Cantor (1969) (*tax/rex:* 1001 bp).

viral relationships would mimic host species' evolutionary associations. This hypothesis was rejected by a seminal, comprehensive phylogenetic analysis of all available HTLV-I/STLV-I *env* sequences (Koralnik et al. 1994) and corroborated subsequently by others (Ureta Vidal et al. 1994a; Ibrahim et al. 1995; Liu et al. 1996; Gessain et al. 1996; Mahieux et al. 1997a,b, 1998a,b; Suzuki and Gojobori 1998; Vandamme et al. 1998a,b). In the present analyses, based on a portion of the *env* gene (Fig. 2), four distinct human clades within type I viruses support previously established subtypes A–D and include two additional subtypes, E and F, identified subsequently (Salemi et al. 1998b). However, the host–pathogen hypothesis predicts the closest relatives of the human strains would be those from other humans. As the interleaved positioning of the human clades demonstrates, the closest relatives of HTLV-I are those from STLV-I. Similarly, the LTR analysis of type I viruses (Fig. 3), comprised of a different sampling of viral strains, corroborates the absence of a monophyletic clade uniting all HTLV-I. Both *env* and LTR derive the same terminal clades corresponding to recognized HTLV-I subtypes and STLV-I groups, but differ in the internal branching uniting these groups.

The present phylogenetic analysis corroborate previously established evolutionary groups as well as discovering unique affiliations between newly described sequences. Each analysis clearly depicts HTLV-I/STLV-I crossing species barriers by recapitulation of monophyletic clusters composed of multiple host species. One of the best examples, corroborated by LTR and *env*, is the previously described close affiliation between HTLV subtype D and newly identified viral sequences from a mandrill (*Mandrillus sphinx*; Msp) colony whose founders were captured in Gabon (Mahieux et al. 1998a) (Figs. 2 and 3). With bootstrap values of 52% (NJ: *env*), 93% (NJ: LTR) and 64% (MP: *env*), 96% (MP: LTR) and significant ($P < 0.001$) support with ML analyses, the STLV-Imsp strains unequivocally share a common ancestry with human subtype D viruses isolated from individuals in Gabon and Pygmy villagers living both in southwest Cameroon and the Central African Republic. However, another mandrill sequence from the same captive colony (Msp-mnd9) was unrelated but instead was closely affiliated with a newly identified subtype HTLV-If from a patient (LIB2) also from Gabon, a result not found in earlier analyses (Salemi et al. 1998b; Mahieux et al. 1998a). Together, these two sequences clustered within a clade comprised of STLV from both captive wild-caught baboon from Kenya (*Papio anubis*) and captive Hamadryas baboons (*P. hamadryas*). This HTLV–I/STLV-I group was repeated in ME, MP, and ML analyses and was supported in part by the LTR phylogeny forming a cluster

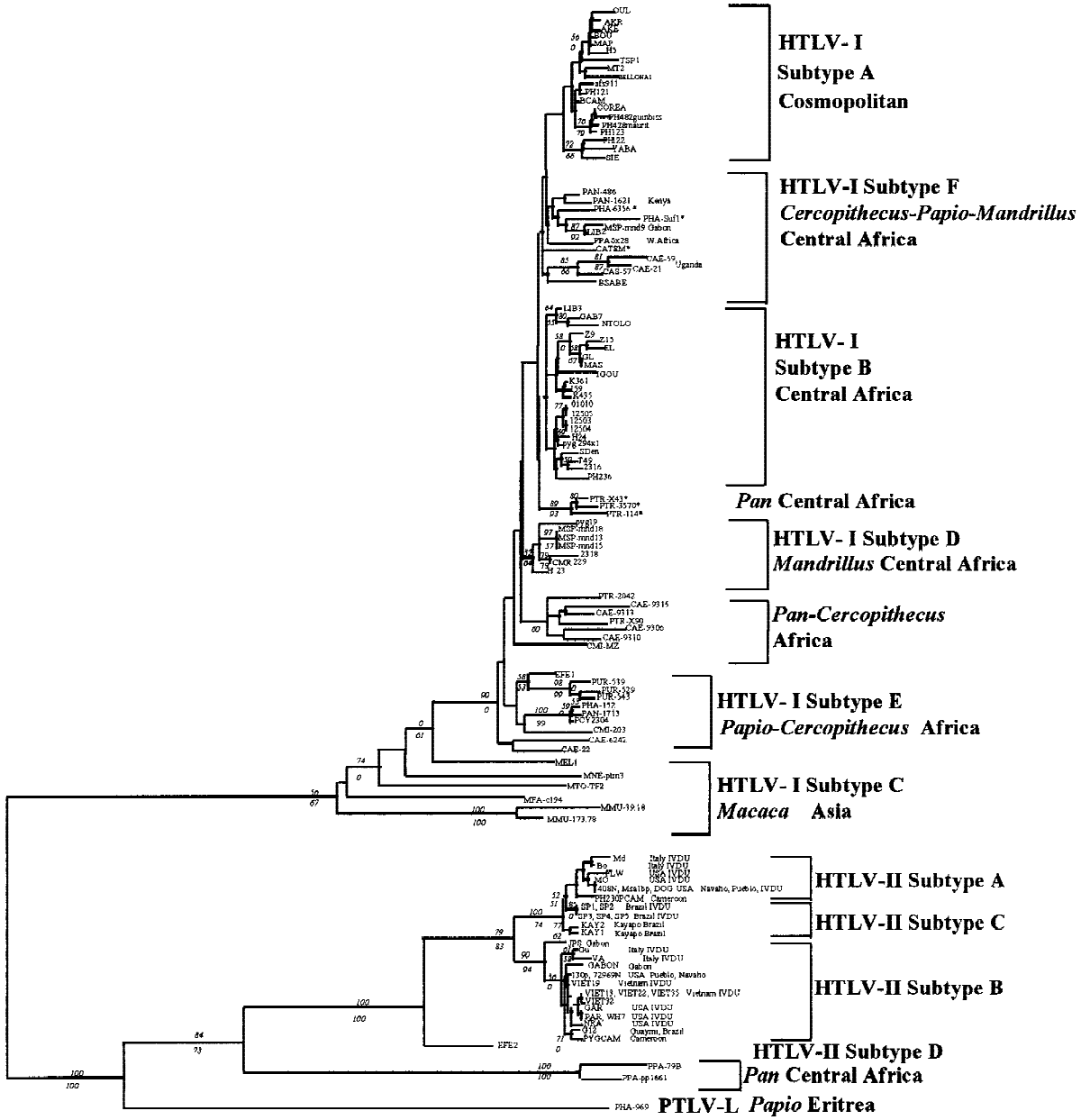

Figure 2 Phylogenetic analysis of 452 bp of the *env* gene across 113 representative HTLV/STLV. Shown is the 50% majority rule consensus of six trees derived from ME estimated by NJ. A 50% majority rule consensus of 5000 trees (length = 824; C.I. = 0.481) and a ML tree (ln likelihood = −4997.59; 21909 trees examined) corroborate the topology depicted. Numbers in italics denote bootstrap proportions in support of adjacent node (NJ/MP). Both NJ and MP trees were reconstructed using PAUP* (by permission of D. Swofford, Smithsonian Institution, Washington, DC). Unrooted NJ trees were based on the Kimura two-parameter model of substitution (Kimura 1980); negative branch lengths allowed with tree-bisection-reconnection swapping algorithm. Conditions for MP heuristic search used unordered characters of equal weight, gaps were treated as a fifth base, and trees were rooted by the midpoint method. ML tree reconstructed by PHYLIP35 (Felsenstein 1993) subroutine DNAML. Asterisks (*) denote simian virus strains of unknown geographic origin. Accession numbers for all 113 strains used in this analysis are listed at http://rex.nci.nih.gov/RESEARCH/basic/lgd/front_page.htm.

with PAN-486 (58%NJ: 82%MP) (Fig. 3) but not in the *env* phylogeny. Likewise, the current analyses indicate another newly identified subtype HTLV-Ie (Salemi et al. 1998b), previously linked with *Cercopithecus aethiops* from South Africa and Kenya and baboons (*Papio cynocephalus*) from Tanzania, is equally affiliated with STLV from wild-caught chacma baboons (*Papio ursinus*) from South Africa (Mahieux et al. 1998b).

The shared evolutionary history between HTLV-Id and wild-born mandrills, HTLV-Ie and wild-caught chacma baboons, and HTLV-If and both wild-caught olive baboons and mandrills is compelling evidence in

support of the hypothesis that HTLV-I subtypes arose from interspecific transmission between natural populations of simian taxa and humans (Koralnik et al. 1994; Liu et al. 1996; Vandamme et al. 1996; Mahieux et al. 1997, 1998a). Under the criterion of consistent formation in trees from ME–NJ, MP, and ML analyses and corroborated independently between LTR and *env*, the other (albeit less rigorous) example supporting this hypothesis includes HTLV-I subtype B from central Africa and STLV isolated from captive descendants of chimpanzee from Sierra Leone. (PTR-114.1, 3570, x43) (Figs. 2 and 3). Although part of a polyphyletic cluster with divergent strains from Asian macaques and orangutan (LTR only), subtype C (Melanesia) forms no clear affiliation in either analysis (Figs. 2 and 3). Lastly, as no putative simian origin for has been discovered as yet, subtype A may have arisen from a pre-existing HTLV-I.

Phylogeographical Patterns Support Common Ancestry Due to Location and not Host Species

Superimposition of the geographic origin of each HTLV-I/STLV-I strain against its phylogenetic position verifies the hypothesis that the basis for shared evolutionary history is geographic proximity (Figs. 3 and 4). The exception is subtype A, which is an assemblage of closely related viral strains from throughout the world. Possible explanations for this transcontinental clade include viral dispersion facilitated by the slave trade from Africa (Koralnik et al. 1994) and the extensive maritime explorations of European countries ~500 years ago (Yanagihara et al. 1995). Furthermore, an earlier episode of dissemination is possible given the high prevalence of cosmopolitan variants in ancient ethnic peoples of Japan (Hinuma 1986; Ishida and Hinuma 1986; Ureta Vidal et al. 1994b).

With a greater number of sequences available for STLV-I in the *env* gene, strains within each simian clade were linked by a common geographic region. Yet, multiple groups exist within the same geographic region as well. For example, the viral Kenya/Tanzanian clade of *P. hamadryus*, *P. anubis*, *C. aethiops*, is more similar to South African chacma baboon and HTLV-Ie strains than to other Kenyan STLV interspersed throughout the phylogeny (Fig. 2). Likewise, the common chimpanzee strains isolated from animals from Sierra Leone (Ptr x90 and Ptr 114.1) are not closely affiliated, with the former more affiliated with *C. aethiops* strains from Senegal and the latter with HTLV subtype B.

Asian STLV, located apart from African STLV, is more similar to HTLV-I from Melanesia. Marked by long branch lengths, these viral strains appear as the most divergent members of type I. Other analyses based on genomic regions encompassing *env/tax* and a portion of the *tax* gene corroborate the genetic uniqueness of the Asian STLV. Thus, STLV-I from a stump-tail macaque (*Macaca arctoides*) is recognized as the earliest divergence within type I viruses (Mahieux et al. 1997b). Likewise, novel *tax* gene sequences from three species of macaques from Indonesia and India consistently aligned with previously determined STLV-I from Asia (Giri et al. 1997).

Distinct, but different, patterns in viral evolution are suggested by the composition and placement of the derived groupings of the type I phylogenetic analyses. For the HTLV-I sequences, subtype A represents many sequences with no obvious STLV association. Marked by short branch lengths (low genetic diversity among strains) the cosmopolitan group is likely to have a recent origin from either HTLV from Africa or an as-yet-unidentified STLV from Asia or Africa, and may have been disseminated worldwide by sixteenth century traders. In contrast, the basal position of the *Pan troglodytes* STLV clade relative to HTLV-I subtype B in both *env* and LTR suggests a transmission from chimpanzee into humans in a common region of western Africa. The inclusion of the mandrill sequences with HTLV-I subtype D, but not in a basal position, prevents any interpretation of the direction of the viral transmission. Further, the results suggest the ancestral virus of mandrill STLV–HTLV-Id clade is either extinct or not yet discovered. The identical interpretation is possible for the inclusion HTLV-Ie within an established clade of South African/Tanzanian STLV-I. Another evolutionary event, suggestive of a recent interspecies transmission, is indicated by the close affiliation between HTLV-If and the mandrill (Msp-mnd9) STLV. Lastly, the high genetic diversity of Asian HTLV-STLV and the divergent STLV-I of *M. arctoides* (not shown) may reflect the retention of unique ancestral lineages within Asia. Thus, all other extant HTLV-I/STLV-I would be more recent, and the common ancestor to type I viruses may have emerged in Asia.

Therefore, the phylogenetic pattern suggests an evolutionary history marked by repeated interspecies transmission events within the same geographic region, even among the same suite of species, but with divergent viral strains. Each evolutionary clade represents the successful outcome of the origin of a novel strain, its subsequent introduction into new host populations, and the successful vertical transmission of the virus from generation to generation.

Unique Phylogenetic and Genetic Divergence of Type II Viruses

HTLV-II is characterized by an intriguing pattern of distribution and levels of genetic diversity unlike type I viruses. Both the *env* (Fig. 2) and LTR (Fig. 4) phylogenetic structures are marked by a basal bifurcation leading to the two sequences isolated from bonobo chimps followed by a unique HTLV-II (subtype D) identified from a Efe pygmy villager in Democratic Re-

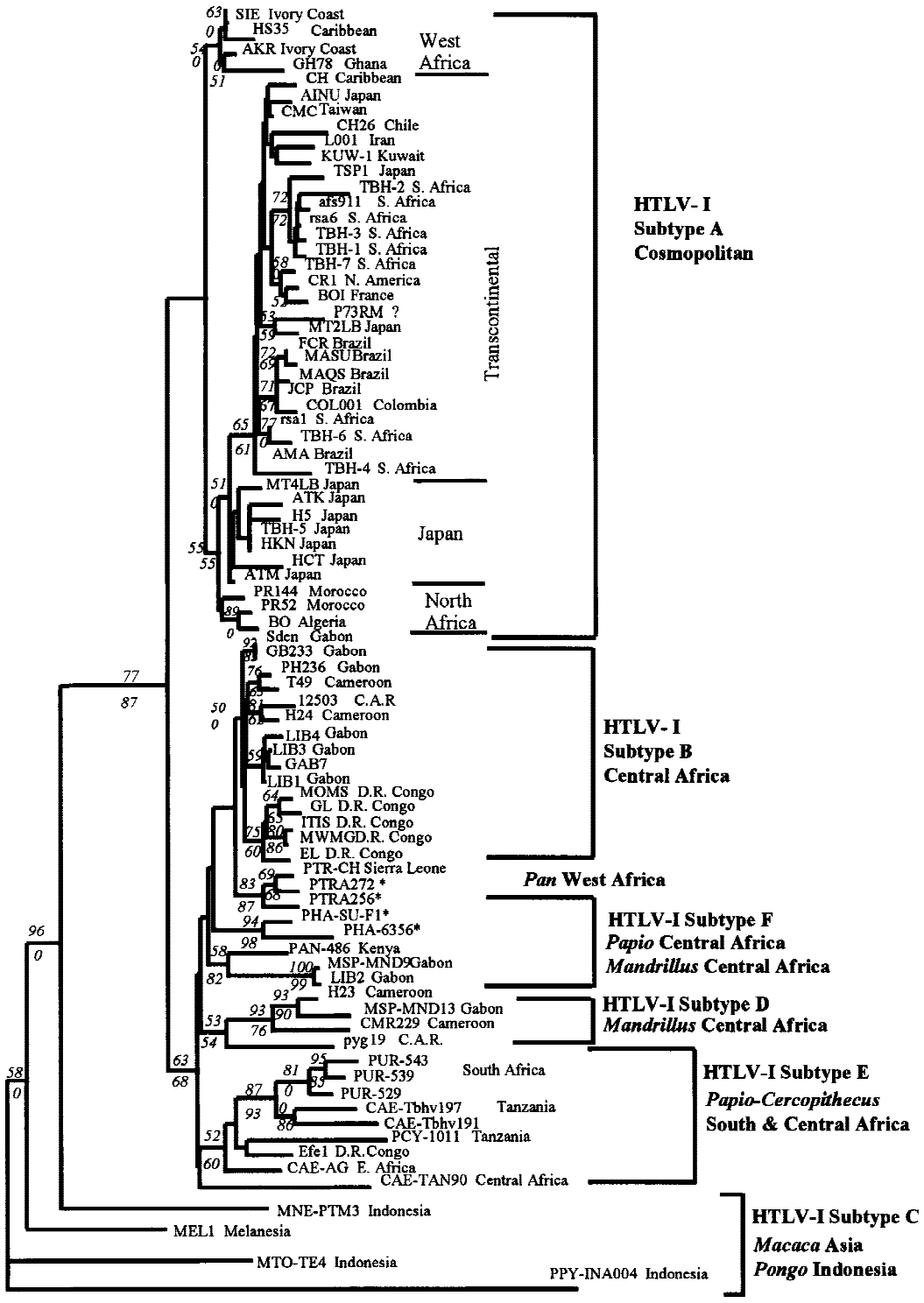

Figure 3 Phylogenetic analysis of 519 bp of LTR region sequenced from 81 strains of HTLV-I/STLV-I. Shown is one of two equivalent trees obtained via ME estimated by neighbor-joining using an heuristic search with the Tamura-Nei (1993) distance measure, negative branch lengths allowed, and the tree-bisection-reconnection swapping algorithm as implemented by PAUP*. Concordant phylogenetic trees were obtained both with MP (50% majority rule of 5000 trees; length = 624; C.I. = 0.543) and ML (ln likelihood = −3443.2; 24,327 trees examined). Conditions for MP heuristic search used unordered characters of equal weight, gaps were treated as a fifth base, and trees were rooted by the midpoint method. ML tree reconstructed by PHYLIP35 (Felsenstein 1993) subroutine DNAML. Asterisks (*) denote simian virus strains of unknown geographic origin. Accession numbers for all 81 strains used in this analysis are listed at http://rex.nci.nih.gov/RESEARCH/basic/lgd/front_page.htm.

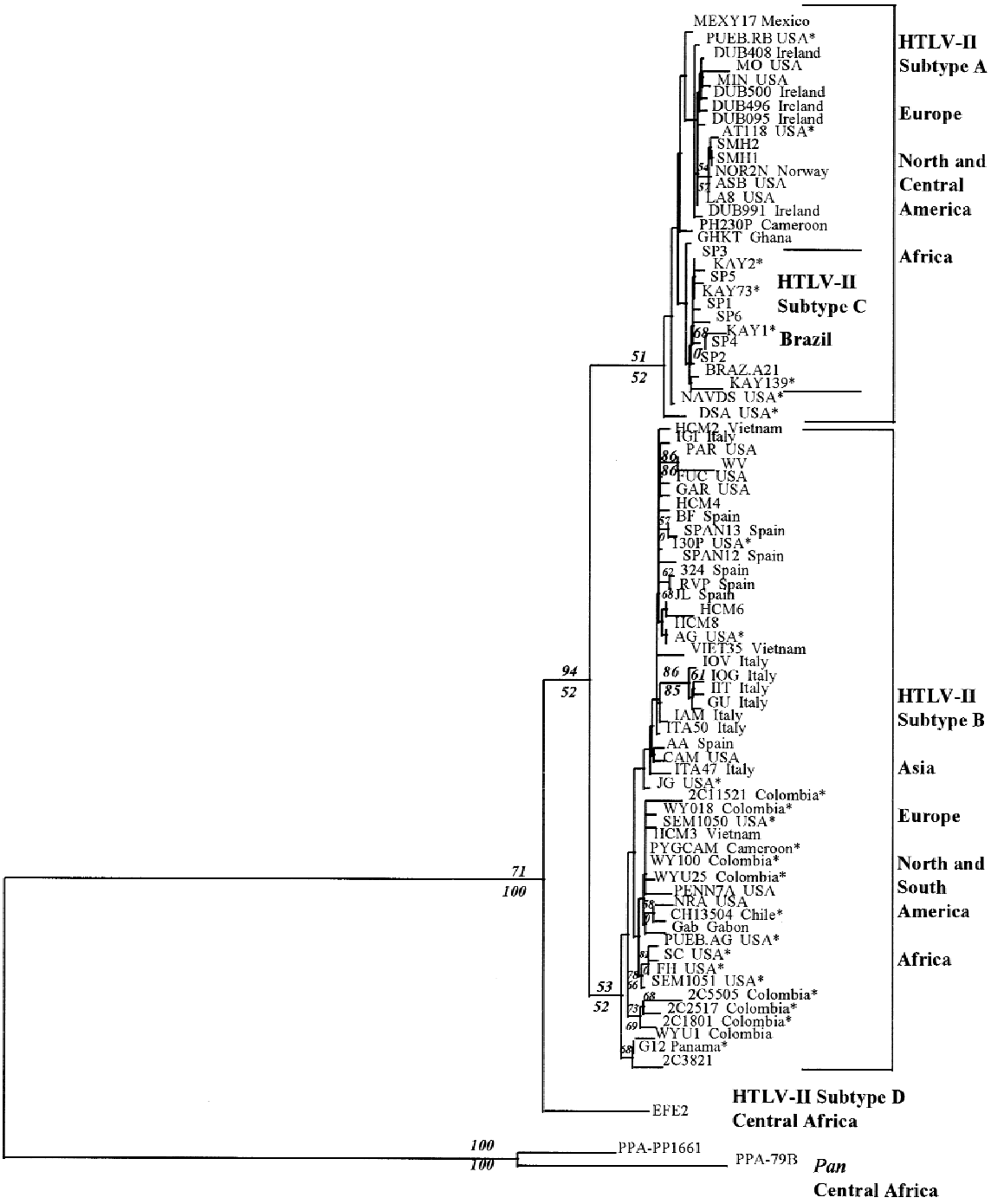

Figure 4 Phylogenetic tree based on 417 bp of LTR region sequenced from 82 viral strains of HTLV-II/STLV-II. Shown is one of two equivalent trees obtained by ME estimated by NJ. A 50% majority rule consensus of 8610 trees (length = 577; C.I. = 0.797) and a ML tree (ln likelihood = −1987; 26,882 trees examined) corroborate the topology depicted. Numbers in italics denote bootstrap proportions in support of adjacent node (NJ/MP). Both NJ and MP trees were reconstructed using PAUP* (by permission of D. Swofford). Unrooted NJ trees were based on the Tamura-Nei (1993) model of substitution and negative branch lengths were allowed using the tree-bisection-reconnection swapping algorithm. Conditions for MP heuristic search used unordered characters of equal weight, gaps were treated as a fifth base, and trees were rooted by the midpoint method. ML tree reconstructed by PHYLIP35 (Felsenstein 1993) subroutine DNAML. Asterisks (*) indicate Amerindian strains. Accession numbers for all 82 strains used in this analysis are listed at http://rex.nci.nih.gov/RESEARCH/basic/lgd/front_page.htm.

public of Congo. The remaining strains form a complex assemblage isolated from diverse ethnic groups and IVDU worldwide and are clustered into subtypes IIa, IIb (Hall et al. 1992; Dube et al. 1993), and IIc (Eiraku et al. 1996).

Although the two bonobo chimpanzee STLV, Ppa-79B and Ppa-pp1661, and human subtype D, originated from Central Africa, the remaining strains are united by factors in addition to geographic proximity. The genetic divergence (8.9% LTR, 3.2% *env*) between STLV-II strongly suggests the virus has resided within *P. paniscus* for a long time. However, the divergence of these sequences relative to known HTLV-II (Table 1; Figs. 2 and 4) is uninformative as to whether the virus originated in the bonobo chimps and then infected humans, or if a common ancestor infected both humans and *P. paniscus* early within type II evolution. In contrast, defined HTLV-II clades, with the exception of subtype C and monotypic D, exhibit little geographic concordance. Resembling the cosmopolitan group of type I viruses, subtype IIa, characterized by the prototype strain Mo (Shimotohno et al. 1985), and subtype IIb, characterized by G12 (Pardi et al. 1993) and Nra (Lee et al. 1993), are each composed of sequences from Africa, Asia, Europe, and the Americas. These subtypes are well supported by both *env* and LTR analyses with high bootstrap values of 100% NJ, 74% MP (*env*) and 51% NJ, 52% MP (LTR) for subtype IIa compared with 90% NJ, 94% MP (*env*) and 53% NJ, 52% MP (LTR) for subtype IIb. With the greater number of strains available in the LTR analysis, most Amerindian HTLV-II appear as subtype IIb.

The phylogenetic distinctiveness of subtype IIc, composed mainly of strains from Brazilian Kayapo Indians and IVDU from Sao Paulo, is less clear. The *env* analysis does not create a monophyletic cluster, but rather places the Kayapo strains together (77% NJ, 62% MP) apart from the SP strains (85% NJ, 0% MP) but both within the subtype IIa lineage. With LTR sequences, the Brazilian strains form a monophyletic cluster (with no bootstrap support) and include additional sequences from Kayapo Indians (Kay73, Kay139), IVDU (Braz.a21), and a strain from a prostitute from Ghana (Switzer et al. 1995b) (Fig. 4). Therefore, phylogenetic evidence in support of subtype IIc is not as strong as for the other three subtypes. Yet, the most unusual feature shared by some IIc strains (Kay1-2 and SP1-6) is that the protein encoded by the *tax* gene resembles subtype IIb and is 25 amino acids longer than IIa (331 amino acids) (Eiraku et al 1996). Thus, the IIc paraphyletic position within subtype IIa with *env* and LTR is contradicted by the *tax* gene homology with subtype IIb. A possible evolutionary interpretation of these discordant results is that a longer Tax protein is more ancestral and retained by a progenitor of IIc but not IIa. The longer Tax proteins of the divergent strains STLV-II (400 amino acids) and IId (344 amino acids) (Vandamme et al. 1998a) offer some support for this hypothesis.

Phylogenetic Conflict Depicts a Paradox in HTLV-II Evolution

The general lack of phylogenetic concordance with geographic location of strains within subtype IIa and IIb forms an evolutionary puzzle. First, the discovery of the virus in different, isolated pygmy tribes in Central Africa (Gessain et al. 1995; Vandamme et al. 1998a) suggests HTLV-II has resided within these peoples for long periods of time, a result substantiated by the HTLV-IId sequence from an Efe tribesman but not from pygmy sequences from Bakola villagers (pygcam) that are within the subtype IIb lineage. Second, culturally and geographically isolated ethnic groups dwelling on different continents share similar forms of the virus. The best example of this phenomena is the identical LTR (456 bp) between pygcam and a Wayuu Indian from Colombia (Wy100) (Figs. 4 and 5). Third, genetic drift associated with long viral residence times within isolated ethnic groups should generate higher levels of divergence among Amerindian and African pygmy sequences than within IVDU. This assumption is not supported; the tree branch length of nearly all viral strains within the major groups are short, indicating only a few genetic differences are unique to each strain irrespective of origin.

Possible explanations for this paradox include (1) selection, (2) a recent origin of modern day HTLV-II with repeated episodes of intercontinental dissemination, or (3) retention of ancestral lineages among disparate populations. Under the hypothesis of selection, the viral genome is constrained so that stochastic accumulation of nucleotide substitution over time is limited. Although possible for coding genes within the virus, it is less likely for the LTR. Alternatively, if extant type II viruses are newly derived, then few substitutions would be phylogenetically informative. The presence of shared types between continents would reflect a random, panmictic assemblage of viral strains analogous to the cosmopolitan subtype A of type I. However, the unique genetic diversity of subtype IId, considered together with bonobo STLV-II, implies a more ancient origin for HTLV-II.

A more plausible scenario suggests type II viruses diverged from a common ancestor with other HTLV/STLV in Africa, and HTLV-II subsequently formed a minimum of three major lineages (IIa, IIb, IId) within Africa. With ancestral human migration events, subtypes IIa and IIb were carried into the New World and segregated among ethnic Amerindian tribes. Subtype IIc, exclusive to Brazil, represents either a more recent divergence within the IIa grouping or is an ancestral lineage (based on the retention of the longer *tax* gene

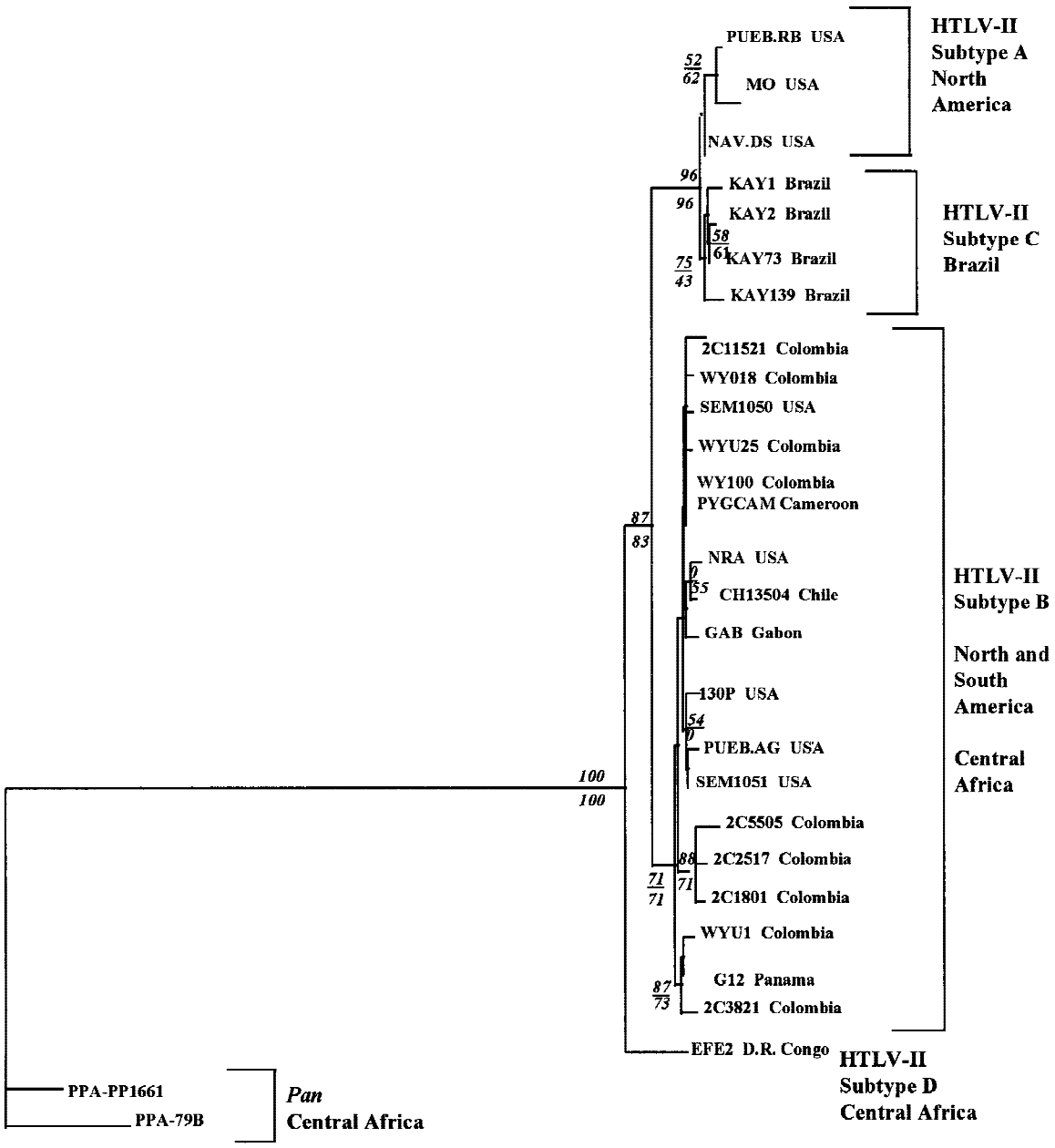

Figure 5 Phylogenetic tree based on 417 bp alignment of LTR region sequenced from 28 viral strains of HTLV-II/STLV-II excluding IVDU. Shown is one of three equivalent trees obtained by ME estimated by NJ. Corroborating trees were obtained by MP that yielded a 50% majority rule of 15 trees of equivalent length = 209; C.I. = 0.856 and ML (ln likelihood = −1595.1; 2776 trees examined). Numbers in italics denote bootstrap proportions in support of adjacent node (NJ/MP). Identical conditions presented in phylogenetic analyses from Fig. 4 were used.

phenotype) that has been extirpated elsewhere. Superimposed against this "backbone" ethnic phylogeny (Fig. 5) are viral sequences from North American, Asian, and European IVDUs (Fig. 4). The interspersed positions denote recent, multiple invasions of new host populations with no informative new mutations to characterize the IVDU as a distinct group.

Thus, the identical strain between Wy100 and pygcam for the LTR may represent the retention of an ancestral polymorphism of initially high frequency within the two host populations. The only other transcontinental comparison within a subtype, not involving intravenous drug use or prostitution as mode of transmission, is a viral sequence from a Mapuche Amerindian in Chile (Ch13504) (Miura et al. 1997) and a sequence from a villager living in an isolated region in Gabon (Gab) (Letourneur et al. 1998). These two sequences differ by five substitutions over the identical

456-bp LTR region. Imposition of a molecular clock, based on the rates of change from IVDU (Salemi et al. 1998a), yields a recent divergence time of 100–400 years ago for the transcontinental strains.

The contradiction between IVDU-based estimates and the presumed ancient origin of New World HTLV-II is resolved by assuming unequal rates of nucleotide substitution or, alternatively, convergent evolution among LTR lineages. Host–pathogen coevolution may result in variation among viral lineages due to genetic consequences of cultural practices (communal breast-feeding) and social structure (e.g., inbreeding or polygamy) over multiple generations within the isolated host population. In contrast, a recent introduction into a new host population, such as IVDU, with diverse immunological backgrounds may facilitate increased rates of mutation. These coevolutionary models stipulate that estimates of mutation rate may be orders of magnitude different than observed in the IVDU population.

Multiple genetic studies offer different views as to the timing of ancestral human migration into the New World. A consensus corroborates the Asian origin of ancient New World peoples, the number of migration events differs from one (Bonatto and Salzano 1997) to three (Greenberg et al. 1986) or four (Horai et al. 1993). Compelling evidence from mitochondrial data isolated from pre-Colombian Oneata individuals suggests a signature expansion of 23,000–37,000 years ago (Stone and Stoneking 1998). Thus, if these four transcontinental alleles (pygcam, Ch13504, Gab, and Wy100) were ancestral and widely distributed within Asian populations in historical times, then the resultant mutation rate estimates would then vary between one and five changes over 456 sites within 23,000–37,000 years (4.7×10^{-8}–1.4×10^{-7}/site per year).

Estimation of the Origin of HTLV/STLV Using Phylogenetic Analyses

At present, the cumulative research of all forms of HTLV/STLV remains inconclusive concerning both the origin and the age of the virus. However, with the isolation of more divergent type I and type II strains from simian species, it can be established that the progenitor virus originated within nonhuman primates. The co-occurrence of highly divergent strains of type II virus within bonobo chimpanzee and human pygmy villagers in the same area of Central Africa supports an ancestral African origin of the type II virus, which subsequently moved into the New World with historical episodes of human migration (Gessain and de The 1996). Considered together with the third and anomalous strain isolated from a baboon *P. hamadryas* in Eritrea (Goubau et al. 1994), the broad genetic differences among the three viral types offer the intriguing hypothesis that African simian species harbor as-yet-unknown but equally diverse representatives of the T-cell leukemia/lymphotropic virus.

A strict interpretation of the present pattern of phylogeny suggests that the progenitor of extant type I virus diverged in Asia. The genetically diverse assemblage of strains, both human and simian, is indicative of a longer residence within Asia than other areas of the world. Thus, unless novel African strains more diverse than those from Asia are discovered, the optimal interpretation of this phylogenetic pattern is that present-day type I viruses throughout the world arose after a period of time from a common ancestor with Asian viruses. Although this interpretation does not preclude that the progenitor of all HTLV/STLV arose in Africa, further insights await discovery of additional, diverse strains.

Considerations in the Applications of Phylogenetic Analyses to HTLV/STLV Epidemiology

The power of phylogenetic analyses in viral epidemiology is apparent in the evolutionary history of HTLV/STLV. The extreme dichotomy in evolutionary patterns between type I and type II viruses reconstructed by phylogenetic analyses reveal disparate associations with primate natural history. In type I viruses, the close affiliation between geographic location and genetic similarity across primates demonstrates the facility of interspecies transmission. These viruses have become globally distributed by a stochastic combination of multiple episodes of interspecies transmission and successful invasions of new host populations.

In contrast, the lack of STLV-II within species other than *Homo sapiens* and *P. paniscus* implies type II is less likely to jump between species. Further, the low levels of genetic differences between HTLV-II isolates from ancient ethnic groups generate the hypothesis that selection may be more of a factor in the mutation process for type II relative to type I viruses. However, the continued future expansion of type II viruses is presaged by the increased prevalence of HTLV-II within the IVDU populations throughout the world and epitomizes viral emergence into a new host population.

Distinctive patterns in the emergence of other RNA viruses have demonstrated the broad utility of molecular phylogenetic analyses. The remarkable structure of "trunk lineages" within human influenza A phylogenetic trees (i.e., a continuum in the viral lineage is preserved at the internal nodes, with the tips of the trees representing isolates that caused epidemics, but then died out) provides strong indicators of the genetic composition of candidate strains in future epidemics (Fitch 1996; Fitch et al. 1997). The virulent canine distemper outbreak of the Serengeti lions of 1994 is linked via phylogenetic analyses with local populations of domestic dog (Roelke-Parker et al. 1996). Feline

immunodeficiency virus, ubiquitous in wild populations of exotic felids, is likely benign and marked by a phylogeny indicative of long residence time within these species (Brown et al. 1994; Carpenter et al. 1996). In contrast, evolution of HIV, recently introduced into humans, is difficult to reconstruct, because of quasispeciation and high mutation rates leading to saturation of sites and loss of phylogenetic signal. However, recent evidence based on a small number of viral samples indicates a subspecies of chimpanzee (*P. troglodytes troglodytes*) may be the source of HIV-1 strains M, N, and O (Gao et al. 1999). These studies, along with the evolutionary history of HTLV/STLV, confirm molecular phylogenetic analyses as a critical component in devising strategies of treatment and management of viral pathogens.

ACKNOWLEDGMENTS

We sincerely thank Drs. S.J. O'Brien and J. Claiborne Stephens for helpful comments and support of this review article. We thank A. Robert for technical assistance. We acknowledge the NCI for allocation of computer time and assistance at the Frederick Supercomputing Center.

REFERENCES

Adkins, R.M. and R.L. Honeycutt. 1994. Evolution of the primate cytochrome c oxidase subunit II gene. *J. Mol. Evol.* **38:** 215–231.

Biggar, R.J., Z. Buskall-Bales, P.N. Yakshe, D. Caussey, G. Gridley, and L. Seeff. 1991. Antibody to human retroviruses among drug users in three east coast American cities, 1972–1976. *J. Infect. Dis.* **163:** 57–63.

Biggar, R., M. Taylor, J. Neel, B. Hjelle, P. Levine, F. Black, G. Shaw, P. Sharp, and B. Hahn. 1996. Genetic variants of human T-lymphotropic virus type II in American Indian groups. *Virology* **216:** 165–173.

Biglione, M., A. Gessain, S. Quiruelas, O. Fay, M.A. Tabords, E. Fernanadez, S. Lupo, A. Panzita, and G. De The. 1993. Endemic HTLV-II infection among Tobas and Matacos Amerindians from North Argentina. *J. Acquir. Immune Defic. Synd.* **6:** 631–633.

Biglione, M., O. Vidan, R. Mahieux, M. de Colombo, M. de Los Angeles de Basualdo, M. Bonnet, G. Pankow, M. Avila de Efron, A. Zorrilla, F. Tekaia et al.. 1999. Seroepidemiological and molecular studies of human T cell lymphotropic virus type 2, subtype B, in isolated groups of Mataco, and Toba Indians of northern Argentina. *AIDS Res. Hum. Retrovir.* **15:** 407–417.

Black, F.L., R.J. Biggar, J.V. Neel, E.M. Maloney, and D.J. Waters. 1994. Endemic transmission of HTLV type II among Kayapo Indians of Brazil. *AIDS Res. Hum. Retrovir.* **109:** 1165–1171.

Bonatto, S.L. and F.M. Salzano. 1997. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc. Natl. Acad. Sci.* **94:** 1866–1871.

Brown, E., N. Yuhki, C. Packer, and S.J. O'Brien. 1994. A lion lentivirus related to feline immunodeficiency virus: Epidemiologic and phylogenetic aspects. *J. Virol.* **68:** 5953–5968.

Carpenter, M.A., E.W. Brown, M. Culver, W. Johnson, J. Pecon-Slattery, D. Brousset, and S.J. O'Brien. 1996. Genetic and phylogenetic divergence of feline immunodeficiency virus in the puma *Puma concolor*. *J. Virol.* **70:** 6682–6693.

Cimarelli, A., C. Duclos, A. Gessain, C. Casoli, and U. Bertazzoni. 1996. Clonal expansion of human T-cell leukemia virus type II in patients with high proviral load. *Virology* **223:** 362–364.

Dube, D.K., M.P. Sherman, N.K. Saksena, V. Bryz-Gornia, J. Mendelson, J. Love, C.B. Arnold, T. Spicer, S. Dube, J.B. Glaser et al. 1993. Genetic heterogeneity in human T-cell leukemia/lymphoma virus type II. *J. Virol.* **67:** 1175–1184.

Echeverria de Perez, G., M. Leon-Ponte, O. Noya, C. Botto, D. Gallo, and N. Bianco. 1993. First description of endemic HTLV-II infection among Venezuelan Amerindians. *J. Acquir. Immune Defic. Synd.* **6:** 1368–1372.

Eiraku, N., P. Novoa, M. da Costa Ferreira, C. Monken, R. Ishak, O. da Costa Ferreira, S.W. Zhu, R. Lorenco, M. Ishak, V. Azvedo et al. 1996. Identification and characterization of a new and distinct molecular subtype of human T-cell lymphotropic virus type 2. *J. Virol.* **70:** 1481–1492.

Etoh, K., S. Tamiya, K. Yamaguchi, A. Okayama, H. Tsubouchi, T. Ideta, N. Mueller, K. Takatsuki, and M. Matsuoka. 1997. Persistent clonal proliferation of human T-lymphotropic virus type I-infected cells in vivo. *Cancer Res.* **57:** 4862–4867.

Felsenstein, J. 1993. *PHYLIP (Phylogeny Inference Package), version 3.5c*. University of Washington, Seattle, WA.

Ferrer, J.F., N. del Pino, E. Esteban, M.P. Sherman, S. Dube, D.K. Dube, M.A. Basombrio, E. Pimentel, A.M. Segovia, S. Quirulas, and B. Poiesz. 1993. High rate of infection with the human T-cell leukemia retrovirus type II on four Indian populations of Argentina. *Virology* **197:** 576–584.

Ferrer, J.F., E. Esteban, S. Dube, M.A. Basombrio, A. Segovia, M. Peralta-Ramos, D.K. Dube, K. Sayre, N. Aguayo, J. Hengst, and B.J. Poiesz. 1996. Endemic infection with human T cell leukemia/lymphoma virus type IIB in Argentinean and Paraguayan Indians: Epidemiology and molecular characterization. *J. Infect. Dis.* **174:** 944–953.

Fitch, W.M. 1996. The variety of human virus evolution. *Mol. Phyl. Evol.* **5:** 247–258.

Fitch, W.M., R.M. Bush, C.A. Bender, and N.J. Cox. 1997. Long term trends in the evolution of H(3)HA1 human influenza type A. *Proc. Natl. Acad. Sci.* **94:** 7712–7718.

Gao, F., E. Bailes, D.L. Robertson, Y. Chen, C.M. Rodenburg, S.F. Michael, L.B. Cummin, L.O. Arthur, M. Peeters, G.M. Shaw, P.M. Sharp, and B.H. Hahn. 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* **397:** 436–441.

Gessain, A. 1996. Epidemiology of HTLV-I and associated diseases. In *Human T-cell lymphotropic virus I* (ed. P. Hollsberg and D.A. Hafler), pp. 34–64. John Wiley and Sons, New York, NY.

Gessain, A. and G. de The. 1996. What is the situation of human T-cell lymphotropic virus type II in Africa? Origin and dissemination of genomic subtypes. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **13:** s228–s235.

Gessain, A., F. Barin, J.C. Vernant, O. Gout, L. Maurs, A. Calender, and G. de The. 1985. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. *Lancet* **2:** 407–410.

Gessain, A., R. Gallo, and G. Franchini. 1992. Low degree of human T-cell leukemia/lymphoma virus type I genetic drift *in vivo* as a means of monitoring viral transmission and movement of ancient human populations. *J. Virol.* **664:** 2288–2295.

Gessain, A., P. Mauclere, A. Froment, M. Biglione, J.Y. Le Hesran, F. Tekaia, J. Millan, and G. De The. 1995. Isolation and molecular characterization of a human T-cell lymphotropic virus type II, subtype B, from a healthy Pygmy living in a remote area of Cameroon: An ancient origin for HTLV-II in Africa. *Proc. Natl. Acad. Sci.* **92:** 4041–4045.

Gessain, A., R. Mahieux, and G. de The. 1996. Genetic variability and molecular epidemiology of human and simian T cell leukemia/lymphoma virus type I. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **13:** s132–s145.

Gillespie, J.H. 1986. Rates of molecular evolution. *Annu. Rev. Ecol. Syst.* **17:** 637–665.

Giri, A., P. Markham, L. Digilio, G. Hurteau, R. Gallo, and G. Franchini. 1994. Isolation of a novel simian T-cell lymphotropic virus from *Pan paniscus* that is distantly related to the human T-cell leukemia/lymphotropic virus types I and II. *J. Virol.* **68:** 8392–8395.

Giri, A., J. Pecon Slattery, W. Heneine, A. Gessain, E. Rivadeneira, R. Desrosiers, L. Rosen, R. Anthony, J. Pamungkas, D. Iskandriate et al. 1997. The *tax* gene sequences from two divergent monophyletic lineages correspond to types I and II of simian

and human T-cell leukemia/lymphotropic viruses. *Virology* **231:** 96–104.
Goubau, P., J. Desmyter, J. Ghesquiere, and B. Kasereka. 1992. HTLV-II among pygmies. *Nature* **359:** 201.
Goubau, P., H.F. Liu, G.G. De Lange, A.-M. Vandamme, and J. Desmyter. 1993. HTLV-II seroprevalence in pygmies across Africa since 1970. *AIDS Res. Hum. Retrovir.* **98:** 709–713.
Goubau, P., M. Van Brussel, A-M. Vandamme, H.F. Liu, and J. Desmyter. 1994. A primate T-lymphotropic virus, PTLV-L, different from human T-lymphotropic viruses types I and II, in a wild-caught baboon. *Proc. Natl. Acad. Sci.* **91:** 2848–2852.
Greenberg, J.H., C.G. Turner II, and S.L. Zegura. 1986. The settlement of the Americas: A comparison of the linguistic, dental and genetic evidence. *Curr. Anthropol.* **27:** 477–497.
Hall, W.W., T. Kubo, S. Ijichi, H. Takahashi, and S.W. Zhu. 1992. Multiple isolates and characteristics of human T-cell leukemia virus type II. *J. Virol.* **66:** 2456–2463.
Hasegawa, M. and M. Fujiwara. 1993. Relative efficiencies of the maximum likelihood, maximum parsimony, and neighbor-joining methods in estimating protein phylogeny. *Mol. Phyl. Evol.* **2:** 1–5.
Hillis, D.M. and J.J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analyses. *Syst. Biol.* **42:** 182–192.
Hino, S., K. Yamaguchi, S. Katamine, H. Sugiyama, T. Amagsaki, K. Kinoshita, Y. Yoshida, H. Doi, Y. Tsuji, and T. Miyamoto. 1985. Mother-to-child transmission of human T-cell leukemia virus type I. *Jpn J. Cancer Res.* **76:** 474–480.
Hinuma, Y. 1986. Seroepidemiology of adult T-cell leukemia virus HTLV-ATLV: Origin of virus carriers in Japan. *AIDS Res.* **2:** s17–s22.
Hinuma, Y., H. Komoda, T. Chosa, T. Kondo, M. Kohakura, T. Takenada, M. Kikuchi, M. Ichimaru, K. Yunoki, I. Sato. et al. 1982. Antibodies to adult T-cell leukemia-virus-associated antigen (ATLA) in sera from patients with ATL and controls in Japan: A nationwide sero-epidemiologic study. *Int. J. Cancer* **15:** 631–635.
Hjelle, B., R. Scalf, and S. Swenson. 1990. High frequency of human T-cell leukemia-lymphoma virus type II infection in New Mexico blood donors: Determination by sequence-specific oligonucleotide hybridization. *Blood* **676:** 450–454.
Horai, S., R. Kondo, Y. Nakagawa-Hattori, S. Hayashi, S. Sonoda, and K. Tajima. 1993. Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol. Biol. Evol.* **10:** 23–47.
Huelsenbeck, J.P. and D.M. Hillis. 1993. Success of the phylogenetic methods in the four taxon case. *Syst. Biol.* **42:** 182–192.
Ibrahim, F., G. de The, and A. Gessain. 1995. Isolation and characterization of a new simian T-cell leukemia virus type I from naturally infected celebes macaques: Complete nucleotide sequence and phylogenetic relationship with the australo-melanesian human T-cell leukemia virus type I. *J. Virol.* **69:** 6980–6993.
International Agency for Research on Cancer. 1996. Human immunodeficiency viruses and human T-cell lymphotropic viruses. *IARC Monogr.* **67:** 308–311.
Ishida, T. and Y. Hinuma. 1986. The origin of Japanese HTLV-I. *Nature* **322:** 504.
Jacobson, S., T. Lehky, M. Nishimura, S. Robinson, D.E. McFarlin, and S. Dhib-Jalbut. 1993. Isolation of HTLV-II from a patient with chronic, progressive neurological disease clinically indistinguishable from HTLV-I-associated myelopathy/tropical spastic paraparesis. *Ann. Neurol.* **33:** 392–396.
Jukes, T.H. and C.R. Cantor. 1969. Evolution of protein molecules. In *Mammalian protein metabolism* (ed. H.N. Munro), pp. 21–132. Academic Press, New York, NY.
Kalyanaraman, V.S., M.G. Sarngadharan, M. Robert-Guroff, I. Miyoshi, D. Blayney, D.W. Golde, and R.C. Gallo. 1982. A new subtype of human T-cell leukemia virus HTLV-II associated with a T-cell variant of hairy cell leukemia. *Science* **218:** 571–573.
Katz, R.A. and A.M. Skalka. 1990. Generation of diversity in retroviruses. *Annu. Rev. Genet.* **24:** 409–445.
Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16:** 111–120.
Koralnik, I.J., E. Boeri, W.C. Saxinger, A. Monico, J. Fullen, A. Gessain, H.G. Guo, R. Gallo, P. Markham, V. Kalyanaraman et al. 1994. Phylogenetic associations of human and simian T-cell leukemia/lymphotropic virus type I strains: Evidence for interspecies transmission. *J. Virol.* **68:** 2693–2707.
Kuhner, M.K. and J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* **11:** 459–468.
La Grenade, L., B. Hanchard, V. Fletcher, B. Cranston, and W. Blattner. 1990. Infective dermatitis of Jamaican children: A marker for HTLV-I infection. *Lancet* **336:** 1345–1347.
Lairmore, M.D., S. Jacobson, F. Gracia, B.K. De, L. Castillo, M. Larreategui, B.D. Roberts, P.H. Levine, W.A. Blattner, and J.E. Kaplan. 1990. Isolation of human T-cell lymphotropic virus type 2 from Guaymi Indians in Panama. *Proc. Natl. Acad. Sci* **87:** 8840–8844.
Lee, H., K. Idler, P. Swanson, J. Aparicio, K. Chin, J. Lax, M. Nguyen, T. Mann, G. Leckie, A. Zanetti et al. 1993. Complete nucleotide sequence of HTLV-II isolate NRA: Comparison of envelope sequence variation of HTLV-II isolates from U.S. blood donors and U.S. and Italian IV drug users. *Virology* **196:** 57–69.
Leon-Ponte, M., G. Echeverria de Perez, N. Bianco, J. Hengst, S. Dube, J. Love, and B.J. Poiesz. 1998. Endemic infection with HTLV-IIB in Venezuelan Indians: Molecular characterization. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **17:** 458–464.
Letourneur, F., L. d'Auriol, M.C. Dazza, M. Peteers, I. Bedjabaga, P. Poit, E. Delaporte, A. Gessain, and N. Monplaisir. 1998. Complete nucleotide sequence of an African human T-lymphotropic virus type II subtype b isolate: Molecular phylogenetic analysis. *J. Gen. Virol.* **79:** 269–277.
Levine, P.H., S. Jacobson, R. Elliot, A. Cavallero, G. Colcough, C. Dorry, C. Stepehenson, R.M. Knigge, J. Drummond, M. Nishimura et al. 1993. HTLV-II infection in Florida Indians. *AIDS Res. Hum. Retrovir.* **9:** 123–127.
Li, W.-H. 1997. *Molecular phylogenetics*. pp. 99-146. Sinauer, Sunderland, MA.
Liu, H.-F., A.-M. Vandamme, M. Van Brussel, J. Desmyter, and P. Goubau. 1994. New retroviruses in human and simian T-lymphotropic viruses. *Lancet* **344:** 265–266.
Liu, H.-F., P. Goubau, M. Van Brussel, K. Van Laethem, Y.C. Chen, J. Desmyter, and A.-M.Vandamme. 1996. The three human T-lymphotropic virus type I subtypes arose from three geographically distinct simian reservoirs. *J. Gen. Virol.* **77:** 359–368.
Liu, H.-F., P. Goubau, M. Van Brussel, J. Desmyter, and A.-M. Vandamme. 1997. Phylogenetic analysis of a simian T lymphotropic virus Type I from a hamadryas baboon. *AIDS Res. Hum. Retrovir.* **13:** 1545–1548.
Mahieux, R., F. Ibrahim, P. Mauclere, V. Herve, P. Michel, F. Tekaia, C. Chappey, B. Garin, E. Van Der Ryst, B. Guillemain et al. 1997a. Molecular epidemiology of 58 new African human T-cell leukemia virus type I strains: Identification of new and distinct HTLV-I molecular subtype in Central Africa and in pygmies. *J. Virol.* **71:** 1317–1333.
Mahieux, R., J. Pecon Slattery, and A. Gessain. 1997b. Molecular characterization and phylogenetic analyses of a new highly divergent simian T-cell lymphotropic virus type I STLV-1marc1 in *Macaca arctoides*. *J. Virol.* **71:** 6253–6258.
Mahieux, R., C. Chappey, M.-C. Georges-Courbot, G. Dubreuil, P. Mauclere, A. Georges, and A. Gessain. 1998a. Simian lymphotropic virus type 1 from *Mandrillus sphinx* as a simian counterpart of human T-cell lymphotropic virus type 1 subtype D. *J. Virol.* **72:** 10316–10322.
Mahieux, R., J. Pecon Slattery, G.M. Chen, and A. Gessain. 1998b. Evolutionary inferences of novel simian T lymphotropic virus type 1 from wild-caught chamca *Papio ursinus* and olive baboons *Papio anubis*. *Virology* **251:** 71–84.
Maloney, E., R. Biggar, J. Neel, M. Taylor, B. Hahn, G. Shaw, and W.

Blattner. 1992. Endemic human T cell lymphotropic virus Type II infection among isolated Brazilian Amerindians. *J. Infec. Dis.* **166:** 100–107.

Miura, T., M. Yamashita, V. Zaninovic, L. Cartier, J. Takehisa, T. Igarashi, E. Ido, T. Fujiyoshi, S. Sonoda, K. Tajima, and M. Hayami. 1997. Molecular phylogeny of human T-cell leukemia virus type I and II of Amerindians in Colombia and Chile. *J. Mol. Evol.* **44:** s76–s82.

Mochizuki, M., A. Ono, E. Ikeda, N. Hikita, T. Watanabe, K. Yamaguchi, K. Sagawa, and K. Ito. 1996. HTLV-I uveitis. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **13:** s50–s56.

Modahl, L.E., K.C. Young, K.F. Varney, H. Khayam-Bashi, and E.L. Murphy. 1997. Are HTLV-II-seropositive injection drug users at increased risk of bacterial pneumonia, abscess, and lymphadenopathy? *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **16:** 169–175.

Morgan, O.S., P. Rodgers-Johnson, C. Mora, and G. Char. 1989. HTLV-I and polymyositis in Jamaica. *Lancet* **2:** 1184–1187.

Morse, S.S. 1995. Factors in the emergence of infectious disease. *Emerg. Infect. Dis.* **1:** 7–15.

Murphy, E.L., J.P. Figueroroa, W.N. Gibbs, A. Brathwaite, M. Holding-Cobham, D. Waters, B. Cranston, B. Hanchard, and W.A. Blattner. 1989. Sexual transmission of human t-lymphotropic virus type I (HTLV-I). *Ann. Inter. Med.* **111:** 555–560.

Murphy, E.L., J. Friday, J.W. Smith, J. Engstrom, R.A. Sacher, K. Miller, J. Gibble, J. Stevens, R. Thomson, D. Hansma et al. 1997a. HTLV-associated myelopathy in a cohort of HTLV-I and HTLV-II infected blood donors. The REDS investigators. *Neurology* **48:** 315–320.

Murphy, E.L., S.A. Glynn, J. Friday, R.A. Sacher, J.W. Smith, D.J. Wright, B. Newman, J.W. Gibble, D.I. Ameti, C.C. Nass, G.B. Schreiber, and G.J. Nemo. 1997b. Increased prevalence of infectious diseases and other adverse outcomes in human T lymphotropic virus types I-and II-infected blood donors. Retrovirus Epidemiology Donor Study Group. *J. Infect. Dis.* **176:** 1468–1475.

Murphy, E.L., R. Mahieux, G. de The, F. Tekaia, D. Ameti, and A. Gessain. 1998. Molecular epidemiology of HTLV-II among United States blood donors and intravenous drug users: An age-cohort effect for HTLV-II RFLP type aO. *Virology* **242:** 425–434.

Nishioka, K. 1996. HTLV-I arthropathy and Sjogren syndrome. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **13:** s57–s62.

Okochi, K. and H. Sato. 1984. Adult T-cell leukemia virus, blood donors, and transfusion: Experience in Japan. *Prog. Clin. Biol. Res.* **182:** 245–256.

Pardi, D., W.M. Switzer, K.G. Hadlock, J.E. Kaplan, R.B. Lal, and T.M. Folks. 1993. Complete nucleotide sequence of an Amerindian human T-cell lymphotropic virus type II HTLV-II. isolate: Identification of a variant HTLV-II subtype b from a Guyami Indian. *J. Virol.* **67:** 4659–4664.

Pardi, D., B. Hjelle, T. Folks, and R. Lal. 1995. Genotypic characteristics of HTLV- II isolates from Amerindian and non-Indian populations. *Virus Genes* **10:** 27–35.

Poiesz, B., F.W. Ruscetti, A.F. Gazdar, P.A. Bunn, J.D. Minna, and R.C. Gallo. 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc. Natl. Acad. Sci.* **80:** 5402–5406.

Roelke-Parker, M.E., L. Munson C. Packer, R. Kock, S. Cleaveland, M. Carpenter, S.J. O'Brien, A. Pospischil, R. Hofmann-Lehmann, H. Lutz et al. 1996. A canine distemper virus epidemic in Serengeti lions *Panthera leo*. *Nature* **379:** 441–445.

Salemi, M., A.-M. Vandamme, F. Guano, C. Gradozzi, E. Cattaneo, C. Casoli, and U. Bertazzoni. 1996. Complete nucleotide sequence of the Italian human T-cell lymphotropic virus type II isolate Gu and phylogenetic identification of a possible origin of South European epidemics. *J. Gen. Virol.* **77:** 1193–1201.

Salemi, M., A.-M. Vandamme, C. Gradozzi, K. Van Laethem, E. Cattaneo, G. Taylor, C. Casoli, P. Goubau, J. Desmyter, and U. Bertazzoni. 1998a. Evolutionary rate and genetic heterogeneity of human t-cell lymphotropic virus type II using isolates from European injecting drug users. *J. Mol. Evol.* **46:** 602–611.

Salemi, M., S. Van Dooren, E. Audenaert, E. Delaporte, P. Goubau, J. Desmyter, and A.M. Vandamme. 1998b. Two new human T-lymphotropic virus type I phylogenetic subtypes in seroindeterminates, a Mbuti pygmy and a Gabonese, have closest relatives among African STLV-I strains. *Virology* **246:** 277–287.

Shimotohno, K., Y. Takahashi, N. Shimizu, T. Gojobori, D. Golde, I. Chen, M. Miwa, and T. Sugimura. 1985. Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: An open reading frame for the protease gene. *Proc. Natl. Acad. Sci.* **82:** 3101–3105.

Sourdis, J. and M. Nei. 1988. Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct tree. *Mol. Biol. Evol.* **5:** 298–311.

Stone, A.C. and M. Stoneking. 1998. MtDNA analysis of a prehistoric Oneata population: Implications for the peopling of the New World. *Am. J. Hum. Genet.* **62:** 1153–1170.

Suzuki, Y. and T. Gojobori. 1998. The origin and evolution of human T-cell lymphotropic virus types I and II. *Virus Genes* **16:** 69–84.

Switzer, W., S. Owen, D. Pieniazek, V. Nerurkar, E. Duenas-Barajas, W. Heneine, and R. Lal. 1995a. Molecular analysis of human t-cell lymphotropic virus type II from Wayuu Indians of Colombia demonstrates two subtypes of HTLV-IIB. *Virus Genes* **10:** 153–162.

Switzer, W., D. Pieniazek, P. Swanson, H.H. Samdal, V. Soriano, R.F. Khabbaz, J. Kaplan, R. Lal, and W. Heneine. 1995b. Phylogenetic relationship and geographic distribution of multiple human t-cell lymphotropic virus type II subtypes. *J. Virol.* **69:** 621–632.

Switzer, W., F. Black, D. Pieniazek, G. Biggar, R. Lal, and W. Heneine. 1996. Endemicity and phylogeny of the human t cell lymphotropic virus type II subtype A from the Kayapo Indians of Brazil: Evidence for a limited regional dissemination. *AIDS Res. Hum. Retrovir.* **12:** 635–640.

Swofford, D.L., G.J. Olsen, P.J. Waddell, and D.M. Hillis. 1996. Phylogenetic Inference. In *Molecular systematics*, 2nd ed. (ed. D.M. Hillis, C. Mortiz, and B.K. Mable), pp. 407–514. Sinauer, Sunderland, MA.

Takahata, N. 1991. Overdispersed molecular clock at the major histocompatability complex loci. *Proc. R. Soc. Lond. B* **243:** 13–18.

Tamura, K. and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10:** 512–526.

Tateno, Y., N. Takezaki, and M. Nei. 1994. Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rates varies with site. *Mol. Biol. Evol.* **11:** 261–277.

Tuppin, P., A. Gessain, M. Kazanji, R. Mahieux, J.Y. Cosnefroy, F. Tekaia, M.C. Georges-Courbot, A. Georges, and G. de The. 1996. Evidence in Gabon for an intrafamilial clustering with mother-to-child and sexual transmission of a new molecular variant of human T-lymphotropic virus type II subtype b. *J. Med. Virol.* **48:** 22–32.

Ureta Vidal, A., A. Gessain, M. Yoshida, F. Tekaia, B. Garin, B. Guillemain, T. Schulz, R. Farid, and G. De The. 1994a. Phylogenetic classification of human T cell leukemia/lymphoma virus type I genotypes in five major molecular and geographical subtypes. *J. Gen. Virol.* **75:** 3655–3666.

Ureta Vidal, A., A. Gessain, M. Yoshida, M. Mahieux, K. Nishioka, F. Tekaia, L. Rosen, and G. de The. 1994b. Molecular epidemiology of HTLV type I in Japan: Evidence for two distinct ancestral lineages with a particular geographical distribution. *AIDS Res. Hum. Retrovir.* **10:** 1557–1566.

Van Brussel, M., P. Goubau, R. Rousseau, J. Desmyter, and A.M. Vandamme. 1996. The genomic structure of a new simian T-lymphotropic virus, STLV-ph969, differs from that of human T-lymphotropic virus types I and II. *J. Gen. Virol.* **77:** 347–358.

Van Brussel, M., M. Salemi, MH.-F. Liu, J. Gabriels, P. Goubau, J. Desmyter, and A.M. Vandamme. 1998. The simian T-lymphotropic virus STLV-PP1664 from *Pan paniscus* is

distinctly related to HTLV-2 but differs in genomic organization. *Virology* **243:** 366–379.

Van Dooren, W., E. Gotuzzo, M. Salemi, D. Watts, E. Audenaert, S. Duwe, H. Ellerbrok, R. Grassman, E. Hagelberg, J. Desmyter, and A.M. Vandamme. 1998. Evidence for a post-Columbian introduction of human T-cell lymphotropic virus in Latin America. *J. Gen. Virol.* **79:** 2695–2708.

Vandamme, A.-M., H.F. Liu, M. Van Brussel, W. De Meurichy, J. Desmyter, and P. Goubau. 1996. The presence of a divergent T-lymphotropic virus in a wild-caught pygmy chimpanzee supports an African origin for the human T-lymphotropic/simian T-lymphotropic group of viruses. *J. Gen. Virol.* **77:** 1089–1099.

Vandamme, A.-M., M. Salemi, M. Van Brussel, H.-F. Liu, K. Laethem, M. Van Ranst, L. Michels, J. Desmyter, and P. Goubau. 1998a. African origin of human t-lymphotropic virus type 2 HTLV-2. supported by a potential new HTLV-2d subtype in Congolese Bambuti Efe pygmies. *J. Virol.* **72:** 4327–4340.

Vandamme, A.-M., M. Salemi, and J. Desmyter. 1998b. The simian origins of the pathogenic human T-cell lymphotropic virus type I. *Trends Microbiol.* **6:** 477–483.

Vitek, C.R., F.L. Gracia, R. Giusti, K. Fukuda, D.B. Green, L.C. Castillo, B. Armien, R.F. Khabbaz, P.H. Levine, J.E. Kaplan, and W. Blattner. 1995. Evidence for sexual and mother-to-child transmission of human T lymphotropic virus type II among Guaymi Indians, Panama. *J. Infect. Dis.* **171:** 1022–1026.

Wakeley, J. 1994. Substitution-rate variation among sites and the estimation of transition bias. *Mol. Biol. Evol.* **11:** 436–442.

Wattel, E., J.P. Vartanian, C. Pannetier, and S. Wain-Hobson. 1995. Clonal expansion of human T-cell leukemia virus type I-infected cells in asymptomatic and symptomatic carriers without malignancy. *J. Virol.* **69:** 2863–2868.

Wattel, E., M. Cavrois, A. Gessain, and S. Wain-Hobson. 1996. Clonal expansion of infected cells: A way of life for HTLV-I. *J. Acquir. Imm. Def. Synd. Hum. Retrovirol.* **13:** s92–s99.

Williams, K.J. and L.A. Loeb. 1992. Retroviral reverse transcriptases: Error frequencies and mutagenesis. *Curr. Top. Microbiol. Immunol.* **176:** 165–180.

Yanagihara, R., N. Saitou, V.R. Nerurkar, K.J. Song, I. Bastian, G. Franchini, and D.C. Gajdusek. 1995. Molecular phylogeny and dissemination of human T-cell lymphotropic virus type I viewed within the context of primate evolution and human migration. *Cell. Mol. Biol.* **41:** s145–s161.

Yang, Z. and S. Kumar. 1996. Approximate methods for estimating the pattern of nucleotide substitution and the variation of substitution rates among sites. *Mol. Biol. Evol.* **13:** 650–659.

Zuckerkandl, E. and L. Pauling. 1965. Molecules as documents of evolutionary history. *J. Theor. Biol.* **8:** 357–366.