JOURNAL OF VIROLOGY, Apr. 1994, p. 2693–2707
0022-538X/94/$04.00+0

Vol. 68, No. 4

Phylogenetic Associations of Human and Simian T-Cell Leukemia/Lymphotropic Virus Type I Strains: Evidence for Interspecies Transmission

IGOR J. KORALNIK,[1]† ENZO BOERI,[1] W. CARL SAXINGER,[1] ANITA LO MONICO,[1] JAKE FULLEN,[1] ANTOINE GESSAIN,[1]‡ HONG-GUANG GUO,[1] ROBERT C. GALLO,[1] PHILLIP MARKHAM,[2] VANIAMBADI KALYANARAMAN,[2] VANESSA HIRSCH,[3] JONATHAN ALLAN,[4] KRISHNA MURTHY,[4] PATRICIA ALFORD,[5] JILL PECON SLATTERY,[6] STEPHEN J. O'BRIEN,[6] AND GENOVEFFA FRANCHINI[1]*

Laboratory of Tumor Cell Biology, National Cancer Institute, Bethesda, Maryland 20892[1]; Advanced BioScience Laboratories, Inc., Kensington, Maryland 20895[2]; Laboratory of Infectious Diseases, NIAID Twinbrook, Rockville, Maryland 20852[3]; Southwest Foundation for Biomedical Research, San Antonio, Texas 78228[4]; University of Texas M. D. Anderson Cancer Center, Bastrop, Texas 78602[5]; and Laboratory of Viral Carcinogenesis, Frederick Cancer Research and Development Center, National Cancer Institute, Frederick, Maryland 21702[6]

Received 28 April 1993/Accepted 5 January 1994

Homologous *env* sequences from 17 human T-leukemia/lymphotropic virus type I (HTLV-I) strains from throughout the world and from 25 simian T-leukemia/lymphotropic virus type I (STLV-I) strains from 12 simian species in Asia and Africa were analyzed in a phylogenetic context as an approach to resolving the natural history of these related retroviruses. STLV-I exhibited greater overall sequence variation between strains (1 to 18% compared with 0 to 9% for HTLV-I), supporting the simian origin of the modern viruses in all species. Three HTLV-I phylogenetic clusters or clades (cosmopolitan, Zaire, and Melanesia) were resolved with phenetic, parsimony, and likelihood analytical procedures. Seven phylogenetic clusters of STLV-I were resolved with the most primitive (deeply rooted) divergence involving several STLV-I strains from Asian primate species. Combined analysis of HTLV-I and STLV-I revealed that neither STLV-I clusters nor HTLV-I clusters recapitulated host species specificity; rather, multiple clades from the same species were closer to clades from other species than to each other. We interpret these evolutionary associations as support for the occurrence of multiple discrete interspecies transmissions of ancestral viruses between primate species (including human) that led to recognizable phylogenetic clades that persist in modern species. Geographic concordance of divergent host species that harbor closely related viruses reinforces that physical feasibility for hypothesized interspecies virus transmission in the past and in the present.

The presence in monkeys and great apes of retroviruses designated simian T-cell leukemia/lymphotropic viruses (STLVs), which are related to the human T-cell leukemia/lymphotropic virus type I (HTLV-I) (13, 23, 43, 52), has been demonstrated by seroepidemiological studies and by viral isolation and partial DNA sequencing of a few STLV type I (STLV-I) isolates (3–6, 9, 19, 20, 22, 24, 26–28, 32, 34, 36, 48, 49, 57–60). Pathological manifestations, mirroring HTLV-I-caused adult T-cell leukemia/lymphoma, in STLV-I-infected gorillas, baboons, macaques, and African green monkeys have been described (24, 33, 34, 47, 57, 58). However, no cases of progressive myelopathy, like tropical spastic paraparesis/HTLV-I-associated myelopathy in humans (15, 40, 44), in STLV-I-infected animals have been reported to date. Molecular epidemiology studies of HTLV-I suggest that the degree of HTLV-I variation over a period of centuries is very low (17), and one form of HTLV-I (so-called cosmopolitan HTLV-I) was discovered to occur in many cities of the world, most likely disseminated through the slave trade (14, 17). This interpretation is supported by the genetic homogeneity of HTLV-I in West Africa, the Americas, and Japan (7, 12, 17, 31, 41, 52). However, the identification of genetically divergent HTLV-I variants in remote populations of Melanesia and Australia (2, 16, 18, 53, 61, 62) has raised new questions about the origin and epidemiological natural history of HTLV-I.

In the present study we have revisited the evolutionary relationship among HTLV-I variants by phylogenetic analysis of a 522-bp segment of the *env* gene in 17 strains from throughout the world. In addition, we used the homologous STLV-I sequence to examine the divergence among 25 STLV-I strains from 12 simian species that carry STLV-I infection. In both human and simian analyses, recognizable phylogenetic clusters (or clades) were defined and confirmed by three analytical approaches (phenetic or distance-based trees, maximum parsimony, and maximum likelihood). When sequences from both STLV-I and HTLV-I sequences were combined, the human and simian clades were resolved but were phylogenetically interspersed and not separated as were human versus simian sequence groupings. Furthermore, there were several examples in which distinct clades from one species were more closely related to sequence clades from another species than they were to each other. The simplest interpretation of these patterns would involve the periodic occurrence of interspecies transmission of STLV-I and HTLV-I viruses between different primate species. Several examples of emerging pathological

* Corresponding author. Mailing address: Laboratory of Tumor Cell Biology, National Cancer Institute, Building 37, Room 6A01, Bethesda, MD 20892. Phone: (301) 496-2386. Fax: (301) 496-8394.

† Present address: Neurology Division, Brigham and Women's Hospital, Boston, MA 02115.

‡ Present address: Institut Pasteur, Laboratoire d'Epidemiologie des Virus Oncogenes, 28 Rue du Dr. Roux, 75724 Paris, Cedex 15, France.

TABLE 1. Origins of the human samples

Phylogenetic cluster[a] and HTLV-I strain code	Clinical diagnosis[b]	Laboratory designation	Geographical origin
I			
H5	TSP/HAM	HTLV-I_5	Guadeloupe (West Indies)
H10	TSP/HAM	HTLV-I_{10}	Peru (South America)
H14	ATL	HTLV-I_{14}	Mauritania (Africa)
HATK	ATL	HTLV-I_{ATK}	Japan (Asia)
HHS35	ATL	HTLV-I_{HS35}	Jamaica (West Indies)
H15	ATL	HTLV-I_{15}	Ivory Coast (West Africa)
H16	ATL	HTLV-I_{16}	Ivory Coast (West Africa)
H18	ATL	HTLV-I_{18}	Ivory Coast (West Africa)
H19	ATL	HTLV-I_{19}	Ivory Coast (West Africa)
H1711	Healthy carrier	HTLV-I_{1711}	Guinea Bissau (West Africa)
H7	Healthy carrier	HTLV-I_{H7}	Bellona (Melanesia)
II			
HZ69	Healthy carrier	HTLV-I_{Z69}	Zaire (central Africa)
HZ6	NSP	HTLV-I_{Z6}	Zaire (central Africa)
HZ17	TSP/HAM	HTLV-I_{Z17}	Zaire (central Africa)
HZ15	Healthy carrier	HTLV-I_{Z15}	Zaire (central Africa)
III			
HMEL1	Healthy carrier	HTLV-I_{Mel1}	Papua New Guinea (Melanesia)
HMEL5	Healthy carrier	HTLV-I_{Mel5}	Solomon Islands (Melanesia)

[a] Phylogenetic cluster (clade) based on analysis in this report (Fig. 1).
[b] NSP, nonspastic paraparesis; TSP/HAM, tropical spastic paraparesis/HTLV-I-associated myelopathy; ATL, adult T-cell leukemia.

viruses between species have been described in the epidemiologic literature (35, 42, 54).

MATERIALS AND METHODS

HTLV-I samples. DNA sequences of the *env* genes were obtained previously by PCR amplification of genomic DNA (see below) from previously described HTLV-I strains (17). A listing of HTLV-I strains, geographic locales of collection, and their clinical diagnoses is presented in Table 1.

STLV-I samples. Twenty-five monkeys with positive serology in an HTLV-I Western blot (immunoblot) were included in this study (Table 2). Samples 22 and C194 represent cell lines established from the peripheral blood mononuclear cells

TABLE 2. General features of the nonhuman primate isolates

Species	Common name	Phylogenetic cluster (clade)	Laboratory designation	Geographical origin	Species or STLV-I strain code
Cercopithecus aethiops	Green monkey	S6	22	Kenya	CAE-22
Cercopithecus aethiops	Grivet	S6	6242	Ethiopia	CAE-6242
Cercopithecus aethiops	Sabeus	S3	9315	Senegal	CAE-9315
Cercopithecus aethiops	Sabeus	S3	9310	Senegal	CAE-9310
Cercopithecus aethiops	Sabeus	S3	9306	Senegal	CAE-9306
Cercopithecus aethiops	Sabeus	S3	9313	Senegal	CAE-9313
Cercopithecus aethiops	Tantalus	S6	21	Uganda	CAE-21
Cercopithecus aethiops	Tantalus	S6	59	Uganda	CAE-59
Cercopithecus ascanius	Red tail monkey	S6	57	Uganda	CAS-57
Cercopithecus mitis albolugaris	Sykes monkey	S6	203	Kenya	CMI-203
Cercopithecus mitis	Mitis	S6	MZ	Zaire	CMI-MZ
Macaca mulatta	Rhesus macaque	S1	173.78	India	MMU-173.78
		S1	39.83	India	MMU-39.83
Macaca nemestrina	Pig tail macaque	S1	PTM3	Asia	MNE-PTM3
Macaca fascicularis	Crab-eating macaque	S1	C194	Indonesia	MFA-C194
Papio cynocephalus	Yellow baboon	S1	991.ICC	Russia (Sukhumi Colony)	PCY-991
Papio cynocephalus	Yellow baboon	S7	2304	East Africa (Kenya/Tanzania)	PCY-2304
Papio anubis	Olive baboon	S7	1713	Kenya/Tanzania	PAN-1713
Papio hamadryas	Hamadryas baboon	S7	152	Kenya/Tanzania	PHA-152
Papio papio	Red baboon	S4	5X28	West Africa	PPA-5X28
Pan troglodytes	Common chimpanzee	S5	114.1	Sierra Leone	PTR-114.1
Pan troglodytes	Common chimpanzee	S2	X90	Sierra Leone	PTR-X90
Pan troglodytes	Common chimpanzee	S5	X43	Unknown	PTR-X43
Pan troglodytes	Common chimpanzee	S5	3570	Unknown	PTR-3570[a]
Pan troglodytes	Common chimpanzee	S2	2042	Unknown	PTR-2042[a]

[a] From Bastrop breeding center, Bastrop, Tex., record of geographic origin is unknown; PTR3570 was captive born and PTR2042 was wild born.

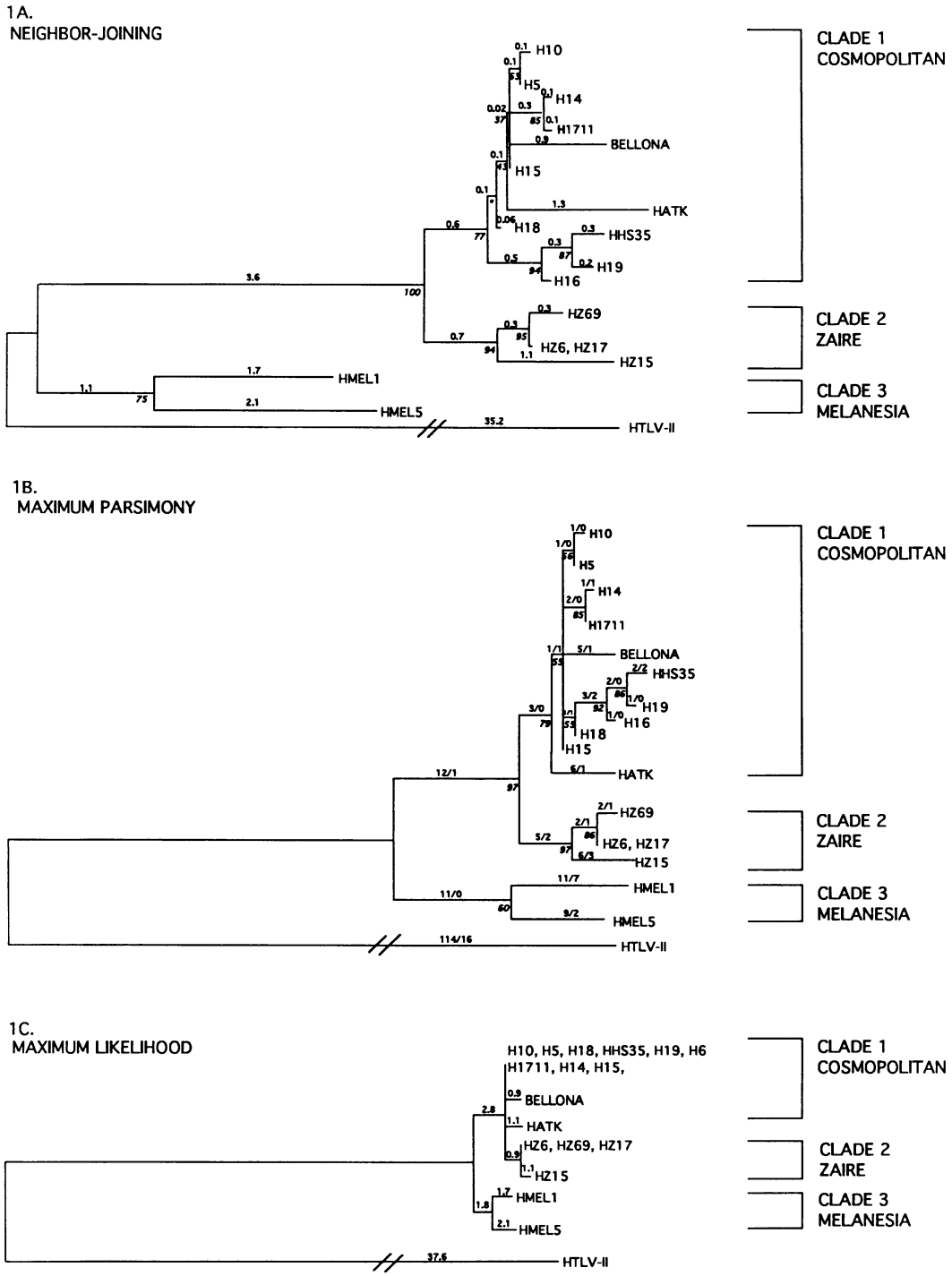


FIG. 1. Phylogenetic trees derived from analysis of 522 nucleotide bp from the envelope region of 17 HTLV-I strains. (A) Neighbor-joining tree determined by using the Kimura distance values given in Table 3. Branch lengths represent percent sequence divergence. Values in italics denote the number of trees generated by 100 bootstrap replications which support the adjacent node. (B) Tree derived by maximum parsimony (length, 232 steps; consistency index, 0.901) obtained by both heuristic search and bootstrap analyses. Numbers on branches represent limb length and homoplasy, in that order. Numbers in italics are bootstrap values (100 replications) in support of the adjacent node. (C) Maximum-likelihood tree (ln likelihood, −1709.54; 2,148 trees examined) generated with a transition/transversion ratio of 2. Limb lengths estimate the expected number of substitutions per site ×100. Branch lengths not significantly different from zero were collapsed into polytomies. For each method, alteration of sequence input did not alter phylogenetic associations.

TABLE 3. Genetic divergence among all pairs of viral sequences

HTLV-I or STLV-I strain code[a]	Genetic distance[b] and % sequence mismatch[c]																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. H10		0.1	0.9	0.7	1.3	0.3	0.5	1.9	1.7	1.3	1.7	2.9	2.5	2.5	3.3	8.0	7.6	4.1	3.9	3.9	4.1	3.7
2. H5	0.2		0.7	0.5	1.1	0.1	0.3	1.7	1.5	1.1	1.5	2.7	2.3	2.3	3.1	7.8	7.4	3.9	3.7	3.7	4.1	3.7
3. H14	1.0	0.8		0.1	1.5	0.5	0.7	2.1	1.9	1.5	1.9	2.7	2.3	2.3	3.5	8.2	7.8	4.3	4.1	4.1	4.5	3.7
4. H1711	0.8	0.6	0.2		1.3	0.3	0.5	1.9	1.7	1.3	1.7	2.9	2.5	2.5	3.3	8.0	7.6	4.1	3.9	3.9	4.3	3.9
5. H7-BELLONA	1.3	1.2	1.5	1.3		0.9	1.1	2.5	2.3	1.9	2.3	3.3	2.9	2.9	3.9	8.7	7.9	4.3	4.1	4.1	4.5	4.5
6. H15	0.4	0.2	0.6	0.4	1.0		0.1	1.5	1.3	0.9	1.3	2.5	2.1	2.1	2.9	7.6	7.2	3.7	3.5	3.5	3.9	3.5
7. H18	0.6	0.4	0.8	0.6	1.2	0.2		1.3	1.1	0.7	1.5	2.3	1.9	1.9	2.7	7.8	7.4	3.9	3.7	3.7	3.9	3.7
8. HHS35	1.9	1.7	2.1	1.9	2.5	1.5	1.3		0.5	0.9	2.9	3.7	3.3	3.3	4.1	8.9	8.1	5.3	5.1	5.1	5.3	5.1
9. H19	1.7	1.5	1.9	1.7	2.3	1.3	1.2	0.6		0.7	2.7	3.5	3.1	3.1	3.9	8.7	8.3	5.1	4.9	4.9	5.1	4.9
10. H16	1.3	1.2	1.5	1.3	1.9	1.0	0.8	1.0	0.8		2.3	3.1	2.7	2.7	3.5	8.2	7.8	4.3	4.1	4.1	4.7	4.5
11. HATK	1.7	1.5	1.9	1.7	2.3	1.3	1.5	2.9	2.7	2.3		3.5	3.1	3.1	3.5	8.7	8.3	4.7	4.5	4.5	4.9	4.5
12. HZ69	2.9	2.7	2.7	2.9	3.3	2.5	2.3	3.6	3.5	3.1	3.5		0.3	0.3	1.9	8.7	8.3	4.7	4.5	4.5	5.1	3.3
13. HZ6	2.7	2.5	2.5	2.7	3.1	2.3	2.1	3.4	3.3	2.9	3.3	0.6		0.0	1.5	8.3	7.8	4.3	4.1	4.1	4.7	2.9
14. HZ17	2.7	2.5	2.5	2.7	3.1	2.3	2.1	3.4	3.3	2.9	3.3	0.6	0.0		1.5	8.3	7.8	4.3	4.1	4.1	4.7	2.9
15. HZ15	3.3	3.1	3.5	3.3	3.8	2.9	2.7	4.0	3.8	3.5	3.5	1.9	1.7	1.7		9.1	8.7	5.1	4.9	4.9	5.5	4.1
16. HMEL1	7.7	7.5	7.9	7.7	8.3	7.3	7.5	8.4	8.3	7.9	8.3	8.3	8.0	8.0	8.6		3.9	8.7	8.5	8.5	8.5	8.9
17. HMEL5	7.3	7.1	7.5	7.3	7.5	6.9	7.1	7.7	7.9	7.5	7.9	7.9	7.7	7.7	8.3	3.8		8.7	8.5	8.5	8.9	8.5
18. PHA-152	4.0	3.8	4.2	4.0	4.2	3.6	3.8	5.2	5.0	4.2	4.6	4.6	4.4	4.4	5.0	8.3	8.3		0.1	0.5	3.7	4.5
19. PCY-2304	3.8	3.6	4.0	3.8	4.0	3.5	3.6	5.0	4.8	4.0	4.4	4.4	4.2	4.2	4.8	8.1	8.1	0.2		0.3	3.5	4.3
20. PAN-1713	3.8	3.6	4.0	3.8	4.0	3.5	3.6	5.0	4.8	4.0	4.4	4.4	4.2	4.2	4.8	8.1	8.1	0.6	0.4		3.5	4.3
21. CMI-203	4.0	4.0	4.4	4.2	4.4	3.8	3.8	5.2	5.0	4.6	4.8	5.0	4.8	4.8	5.4	8.1	8.4	3.6	3.5	3.5		4.7
22. CAE-22	3.6	3.6	3.6	3.8	4.4	3.5	3.6	5.0	4.8	4.4	4.4	3.3	3.1	3.1	4.0	8.4	8.1	4.4	4.2	4.2	4.6	
23. CAE-6242	5.0	5.0	5.4	5.2	5.8	4.8	5.0	6.3	6.1	5.8	5.8	5.8	5.6	5.6	6.1	9.0	9.4	4.2	4.0	4.0	4.8	3.6
24. CMI-MZ	3.5	3.3	3.6	3.5	4.0	3.1	3.3	4.6	4.4	4.0	4.4	4.0	3.8	3.8	4.4	8.4	8.3	4.4	4.2	4.2	5.0	4.6
25. CAE-21	4.2	4.0	4.4	4.2	4.8	3.8	3.8	5.2	5.0	4.6	4.8	4.2	4.0	4.0	4.6	9.8	9.4	5.4	5.2	5.4	5.8	5.2
26. CAE-59	4.8	4.6	5.0	4.8	5.4	4.4	4.4	5.8	5.6	5.2	5.4	4.4	4.2	4.2	4.8	9.8	9.4	6.0	5.8	6.0	6.3	5.2
27. CAS-57	3.3	3.1	3.5	3.3	3.8	2.9	3.1	4.4	4.2	3.8	3.8	3.1	2.9	2.9	3.5	7.9	7.9	4.8	4.6	4.8	5.6	3.5
28. PTR-114.1	4.0	3.8	4.2	4.0	4.6	3.6	3.5	4.4	4.2	3.8	4.6	3.1	2.9	2.9	3.1	8.6	8.3	5.8	5.6	5.6	5.8	4.8
29. PTR-3570	3.5	3.3	3.6	3.5	4.0	3.1	2.9	3.8	3.6	3.3	4.0	2.5	2.3	2.3	2.9	8.8	8.1	5.2	5.0	5.0	5.2	4.2
30. PTR-X43	3.1	2.9	2.9	2.7	3.6	2.7	2.5	3.5	3.3	2.9	3.6	2.1	1.9	1.9	2.5	8.4	7.7	4.8	4.6	4.6	4.8	3.8
31. PPA-5X28	3.3	3.1	3.4	3.3	3.8	2.9	3.1	3.6	3.8	3.4	3.4	4.2	3.6	3.6	4.2	8.6	8.2	5.4	5.2	5.2	5.6	5.2
32. PTR-2042	4.4	4.2	4.6	4.4	4.8	4.0	3.8	4.8	4.6	4.6	5.0	3.8	3.6	3.6	4.6	9.0	9.4	5.8	5.6	5.6	6.5	5.6
33. PTR-X90	4.8	4.6	5.0	4.8	5.2	4.4	4.2	4.8	5.0	4.6	5.4	4.2	4.0	4.0	5.0	8.8	9.6	5.8	5.6	5.6	6.5	6.0
34. CAE-9315	5.0	4.8	5.2	5.0	5.6	4.6	4.4	5.4	5.6	5.2	5.6	4.8	4.6	4.6	4.8	9.8	10.2	6.3	6.1	6.1	7.1	6.1
35. CAE-9313	4.8	4.6	5.0	4.8	5.2	4.4	4.2	4.8	5.0	4.6	5.4	3.8	4.0	4.0	5.0	9.0	9.4	6.1	6.0	6.0	6.9	5.6
36. CAE-9306	5.6	5.6	6.0	5.8	5.8	5.4	5.2	5.8	6.0	5.6	6.3	4.0	4.2	4.2	5.2	10.2	10.2	6.1	6.0	6.0	6.9	6.1
37. CAE-9310	4.2	4.0	4.4	4.2	4.2	3.8	3.6	4.6	4.8	4.4	4.8	3.3	3.1	3.1	3.6	9.8	9.4	5.6	5.4	5.4	6.3	5.4
38. MNE-PTM3	9.8	9.6	9.6	9.4	10.3	9.4	9.2	9.8	10.0	9.6	10.7	10.9	10.3	10.3	11.3	11.9	10.7	11.1	10.9	10.9	10.7	11.7
39. MFA-C194	11.7	11.7	11.3	11.5	12.5	11.5	11.3	11.9	12.1	11.9	12.5	11.1	11.3	11.3	11.5	13.1	12.9	12.1	11.9	12.1	12.3	11.9
40. MMU-39.83	14.4	14.4	14.0	14.2	15.0	14.2	14.4	15.0	15.2	15.0	15.0	13.8	14.0	14.0	14.2	15.2	15.2	15.4	15.2	15.2	15.2	14.8
41. PCY-991.ICC	13.6	13.6	13.2	13.4	14.2	13.4	13.6	14.2	14.4	14.2	14.2	13.1	13.2	13.2	13.4	14.4	14.4	14.6	14.4	14.4	14.4	14.0
42. MMU-173.78	13.2	13.2	12.9	13.1	13.8	13.1	13.2	13.8	14.0	13.8	13.8	12.7	12.8	12.8	13.1	15.0	15.0	14.2	14.0	13.6	14.0	14.0
43. HTLV-II	29.7	29.7	30.1	30.3	30.5	29.9	29.9	29.7	29.9	29.9	30.8	29.7	29.7	29.7	30.3	28.7	28.7	29.5	29.3	29.3	29.1	29.5

[a] Number given below with species or strain codes refers to the same species or strain when given over the field of data.
[b] Upper matrix is genetic distances estimated by using the two-parameter model of Kimura (30), and given in Kimura distance values.
[c] Lower matrix is the percent sequence mismatch between aligned sequences, with a gap scored as a single residue regardless of nucleotide length (39).

(PBMC) from an African green monkey and a crab-eating macaque, respectively. Sample 114.1 is a previously described STLV-I isolate from a Sierra Leone chimpanzee (28). Samples X90 and X43 represent cultured PBMC (28 weeks) from wild-caught chimpanzees held at the South West Foundation (San Antonio, Tex.). Samples 3570 and 2042 represent fresh PBMC from a captive-born chimpanzee and a wild-caught chimpanzee housed at the Bastrop Breeding Colony, respectively. Samples 57 and 203 are continuous cell lines from the PBMC of a red tail monkey housed at the Uganda Virus Research Institute and a blue monkey housed at the Yerkes Primate Center (Atlanta, Ga.), respectively. Samples 9310, 9315, 9306, 9313, and 6242 represent uncultured PBMC from four free-ranging sabeus monkeys and one grivet, respectively. Sample 1713 is a lymph node from an olive baboon, and samples 2304, 152, and 5x28 are PBMC from a yellow, a hamadryas, and a red baboon, respectively, held at the South West Foundation. Samples MMU-173.78 and MMU-39.83 are viral isolates (6) derived from wild-caught rhesus macaques generously provided by R. Desrosiers of the New England Regional Primate Research Center (Southborough, Mass.). Sample MZ represents fresh PBMC from a free-ranging *Cercopithecus mitis* generously provided by D. Messinger from the National Institute of Biomedical Research (Kinshasa, Zaire). Samples 21 and 59 are uncultured PBMC from free-ranging tantalus monkeys from Uganda. Sample 991.ICC represents a cell line obtained from a yellow baboon (19), and isolate PTM3 is from an Asian monkey (pig tail macaque) and was previously described (60).

DNA, PCR amplification, molecular cloning, and DNA sequencing. The PCR amplification of genomic DNA from antibody-positive animals was accomplished with the primer

TABLE 3—*Continued*

Genetic distance and % sequence mismatch[c]																				
23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
5.1	3.5	4.3	4.9	3.3	4.1	3.5	3.1	3.1	4.5	4.9	5.1	4.9	5.7	4.3	10.0	12.7	15.8	14.9	14.5	39.7
5.1	3.3	4.1	4.7	3.1	3.9	3.3	2.9	2.9	4.3	4.7	4.9	4.7	5.7	4.1	9.7	12.6	15.8	14.9	14.4	39.7
5.5	3.7	4.5	5.1	3.5	4.3	3.7	2.9	3.3	4.7	5.1	5.3	5.1	6.1	4.5	9.7	12.2	15.4	14.4	14.0	40.4
5.3	3.5	4.3	4.9	3.3	4.1	3.5	2.7	3.1	4.5	4.9	5.1	4.9	5.9	4.3	9.5	12.4	15.6	14.7	14.2	40.7
5.9	4.1	4.9	5.6	3.9	4.7	4.1	3.7	3.7	4.9	5.3	5.7	5.3	5.9	4.3	10.6	13.6	16.5	15.6	15.2	41.0
4.9	3.1	3.9	4.5	2.9	3.7	3.1	2.7	2.7	4.1	4.5	4.7	4.5	5.5	3.9	9.5	12.4	15.6	14.7	14.2	40.1
5.1	3.3	3.9	4.5	3.1	3.5	2.9	2.5	2.9	3.9	4.3	4.5	4.3	5.3	3.7	9.3	12.2	15.8	14.9	14.4	40.1
6.5	4.7	5.3	6.0	4.5	4.5	3.9	3.5	3.5	4.9	4.9	5.5	4.9	5.9	4.7	10.0	12.9	16.5	15.6	15.1	39.8
6.3	4.5	5.1	5.8	4.3	4.3	3.7	3.3	3.7	4.7	5.1	5.7	5.1	6.1	4.9	10.2	13.1	16.8	15.8	15.4	40.1
5.9	4.1	4.7	5.3	3.9	3.9	3.3	2.9	3.3	4.7	4.7	5.3	4.7	5.7	4.5	9.7	12.9	16.5	15.6	15.1	40.1
5.9	4.5	4.9	5.5	3.9	4.7	4.1	3.7	3.3	5.1	5.5	5.7	5.5	6.5	4.9	11.0	13.5	16.5	15.6	15.1	41.7
5.9	4.1	4.3	4.5	3.1	3.1	2.5	2.1	4.1	3.9	4.3	4.9	3.9	4.1	3.3	11.3	12.0	15.2	14.3	13.8	39.6
5.5	3.7	3.9	4.1	2.7	2.7	2.1	1.7	3.7	3.5	3.9	4.5	3.9	4.1	2.9	10.8	12.0	15.1	14.2	13.8	40.0
5.5	3.7	3.9	4.1	2.7	2.7	2.1	1.7	3.7	3.5	3.9	4.5	3.9	4.1	2.9	10.8	12.0	15.1	14.2	13.8	40.0
6.3	4.5	4.7	4.9	3.5	3.1	2.9	2.5	4.1	4.7	5.1	4.9	5.1	5.3	3.7	11.7	12.4	15.6	14.7	14.2	40.6
9.6	8.9	10.5	10.5	8.3	9.1	9.3	8.9	8.9	9.5	9.3	10.4	9.6	10.9	10.4	12.4	14.3	16.8	15.9	16.6	38.3
10.0	8.7	10.0	10.1	8.3	8.7	8.5	8.1	8.5	10.0	10.2	10.9	10.0	10.9	10.0	11.1	14.1	16.9	15.1	16.6	38.3
4.3	4.5	5.5	6.2	4.9	5.9	5.3	4.9	5.3	5.9	5.9	6.5	6.3	6.3	5.7	11.5	13.1	17.0	16.1	15.6	39.5
4.1	4.3	5.3	6.0	4.7	5.7	5.1	4.7	5.1	5.7	5.7	6.3	6.1	6.1	5.5	11.3	12.9	16.8	15.9	15.4	39.1
4.1	4.3	5.5	6.2	4.9	5.7	5.1	4.7	5.1	5.7	5.7	6.3	6.1	6.1	5.5	11.3	13.1	16.8	15.8	14.9	39.0
4.9	5.1	5.9	6.6	5.7	5.9	5.3	4.9	5.5	6.7	6.7	7.4	7.2	7.2	6.6	11.1	13.4	16.8	15.9	15.4	38.7
3.7	4.7	5.3	5.3	3.5	4.9	4.3	3.9	5.1	5.7	6.1	6.3	5.7	6.3	5.5	12.1	12.9	16.3	15.4	15.4	39.4
	5.7	6.6	7.2	5.9	7.0	6.3	6.1	6.1	6.5	7.2	7.4	6.8	7.2	7.4	11.9	13.8	16.3	15.2	14.9	37.9
5.6		5.5	6.2	4.5	5.3	4.7	4.3	5.1	5.7	6.1	5.9	5.9	6.5	5.3	11.0	13.1	15.6	14.7	14.9	39.0
6.3	5.4		1.7	2.7	5.1	4.5	4.1	5.1	5.9	5.5	5.7	6.0	7.0	5.3	11.8	14.3	18.1	17.6	17.1	41.5
6.9	6.0	1.7		2.1	5.8	5.1	4.7	6.2	6.6	6.6	7.2	7.0	8.1	6.4	12.9	15.0	19.3	18.3	17.8	42.1
5.8	4.4	2.7	2.1		3.9	3.7	3.3	4.5	4.7	5.1	5.7	5.5	6.5	4.9	11.7	13.3	17.2	16.3	15.8	41.3
6.7	5.2	5.0	5.6	3.8		1.7	1.3	5.3	3.9	5.9	5.3	5.9	6.6	4.9	11.7	13.8	17.5	16.5	16.1	39.7
6.1	4.6	4.4	5.0	3.6	1.7		0.7	4.7	4.5	5.3	5.1	5.3	5.9	4.7	11.1	13.1	17.3	16.8	15.8	40.3
6.0	4.2	4.0	4.6	3.3	1.3	0.8		4.3	4.1	4.9	4.7	4.9	5.5	4.3	10.2	13.1	12.2	6.3	15.4	40.3
6.1	5.2	5.2	6.1	4.6	5.4	4.8	4.4		4.9	4.1	5.1	4.5	5.9	4.7	10.6	13.1	15.2	14.3	13.8	39.3
6.3	5.6	5.8	6.3	4.6	3.8	4.4	4.0	5.0		3.5	4.1	3.5	5.3	4.1	11.3	13.4	16.5	15.1	14.7	40.7
6.9	6.0	5.4	6.3	5.0	5.8	5.2	4.8	4.2	3.5		3.7	2.7	4.5	3.7	11.7	13.1	15.9	14.9	14.5	39.1
7.1	5.8	5.6	6.9	5.6	5.2	5.0	4.6	5.2	4.0	3.6		2.9	5.1	3.1	11.7	14.0	15.4	15.0	14.5	40.3
6.5	5.8	5.8	6.7	5.4	5.8	5.2	4.8	4.6	3.5	2.7	2.9		2.9	2.5	11.5	13.1	14.5	13.6	13.6	39.5
6.9	6.3	6.7	7.7	6.3	6.3	5.8	5.4	5.9	5.2	4.4	5.0	2.9		3.1	11.7	13.4	15.4	14.5	14.5	40.1
7.1	5.2	5.2	6.1	4.8	4.8	4.6	4.2	4.8	4.0	3.6	3.1	2.5	3.1		11.7	14.1	15.4	14.5	14.1	40.4
11.5	10.7	11.3	12.3	11.3	11.3	10.7	10.0	10.2	10.9	11.3	11.3	11.1	11.3	11.3		11.8	17.1	16.6	16.8	39.7
12.7	12.1	13.1	13.6	12.3	12.7	12.1	11.3	12.3	12.3	12.1	12.9	12.1	12.3	12.9	11.3		14.0	13.1	13.6	37.2
14.8	14.2	16.1	17.1	15.5	15.7	15.5	14.8	14.0	15.0	14.4	14.0	13.2	14.0	14.0	15.7	12.9		1.1	3.3	39.8
13.8	13.4	15.7	16.3	14.8	15.0	15.2	14.4	13.2	13.8	13.6	13.6	12.5	13.2	13.2	15.3	12.1	1.2		2.1	38.3
13.6	13.6	15.4	15.9	14.4	14.6	14.4	14.0	12.8	13.4	13.2	13.2	12.5	13.2	12.9	15.5	12.5	3.3	2.1		38.9
28.7	29.3	30.7	31.0	30.7	29.7	30.1	30.1	29.3	30.3	29.3	30.1	29.5	29.9	30.1	29.4	28.4	29.9	28.9	29.3	

set *env* 1 and *env* 2 (*env* 1, 5′ TTT GAG CGG CCG CTC AAG CTA TAG TCT CCT CCC CCTG 3′; *env* 2, 5′ ACT TAG AAT TCG GGA GGT GTC GTA GCT GAC GGAGG 3′) containing the *Not*I and the *Eco*RI restriction sites, respectively. These primers amplify a 522-bp envelope fragment (nucleotides 6046 to 6567) (equivalent to European Molecular Biology Laboratory bases 6068 to 6589) of the Japanese prototype HTLV-I_{ATK} (52), spanning the carboxy terminus of the gp46 and almost the entire transmembrane protein gp21. PCR was carried out in a volume of 100 μl with 0.1× PCR buffer (10 mM Tris HCl [pH 8.3 at 25°C], 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% [wt/vol] gelatin); a 0.2 mM concentration of dATP, dCTP, dGTP, and dTTP; 200 ng of each primer; 2.5 U of *Thermus aquaticus* DNA polymerase (Perkin-Elmer Cetus); and 1 to 2 μg of each DNA. The reaction was carried out in a DNA thermal cycler (Perkin-Elmer Cetus) with an initial denaturation step for 5 min at 94°C and then 35 cycles for 1 min at 94°C, 1 min at 58°C, and 2 min at 72°C. The amplified fragments were cloned at the *Not*I and *Eco*RI sites of the Bluescript vector. The ^{32}P-labelled oligonucleotides PE 1.2 (nucleotides 6312 to 6336) and PR (nucleotides 6453 to 6433) were used as probes to select positive clones. Washing conditions were 5× SSC (1× SSC is 0.15 M sodium chloride plus 0.015 M sodium citrate) at 37°C for 10 min. The plasmid DNA of the positive recombinant clones was extracted, purified, and sequenced by the Sanger et al. method (50).

Phylogenetic analysis. HTLV-I and STLV-I *env* DNA sequences (522 bp, positions 6046 to 6567) were aligned with the GAP program of the Genetics Computer Group of the University of Wisconsin computer software package (8), which uses the algorithm of Needleman and Wunsch (38). Pairwise genetic distances between sequences were estimated by using the two-parameter model of Kimura (30) and also by direct estimation of percent aligned matches with gaps given a weight of a single residue regardless of their length (39). Results were comparable regardless of the distance measure employed. A phenetic analysis of distance matrices used the neighbor-joining algorithm (46). Both Kimura distance esti-

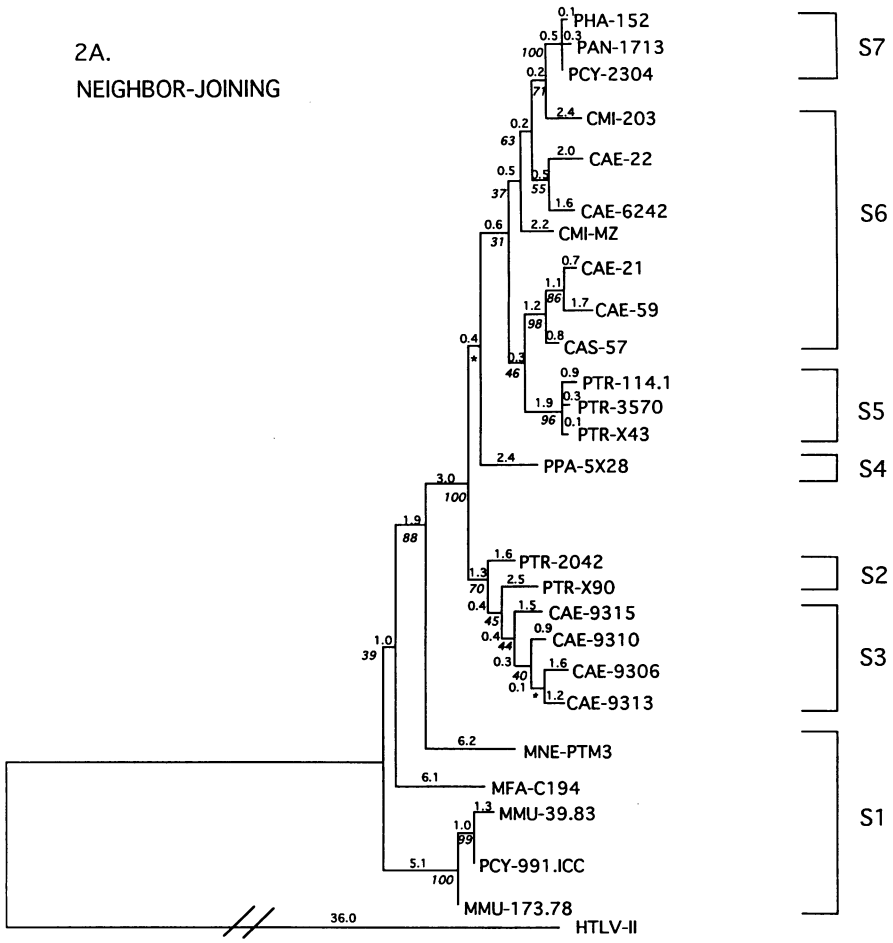


FIG. 2. Phylogenetic trees derived from analysis of 522 nucleotide bp from the envelope region of 25 STLV-I strains. (A) Neighbor-joining tree determined by using Kimura distance values given in Table 3. Branch lengths represent percent sequence divergence. Values in italics are numbers of trees in bootstrap analysis which support the adjacent node. Asterisks denote nodes not supported by bootstrap analysis. (B) Tree derived by maximum parsimony represents consensus tree from the bootstrap analysis and one of 20 trees of equivalent length determined by a heuristic search. Numbers on branches correspond to branch length and homoplasy, in that order. Numbers in italics are bootstrap values (100 iterations) in support of the node. (C) Maximum-likelihood tree (ln likelihood, −2912.47; 2,430 trees examined) generated with a transition/transversion ratio of 2. Limb lengths approximate the expected number of substitutions per site × 100. Branch lengths not significantly different from zero were collapsed into polytomies. In each method, different input orders did not alter topological associations.

mates and neighbor-joining trees were performed by using the PHYLIP-3.35 (phylogenetic inference package) computer program (10).

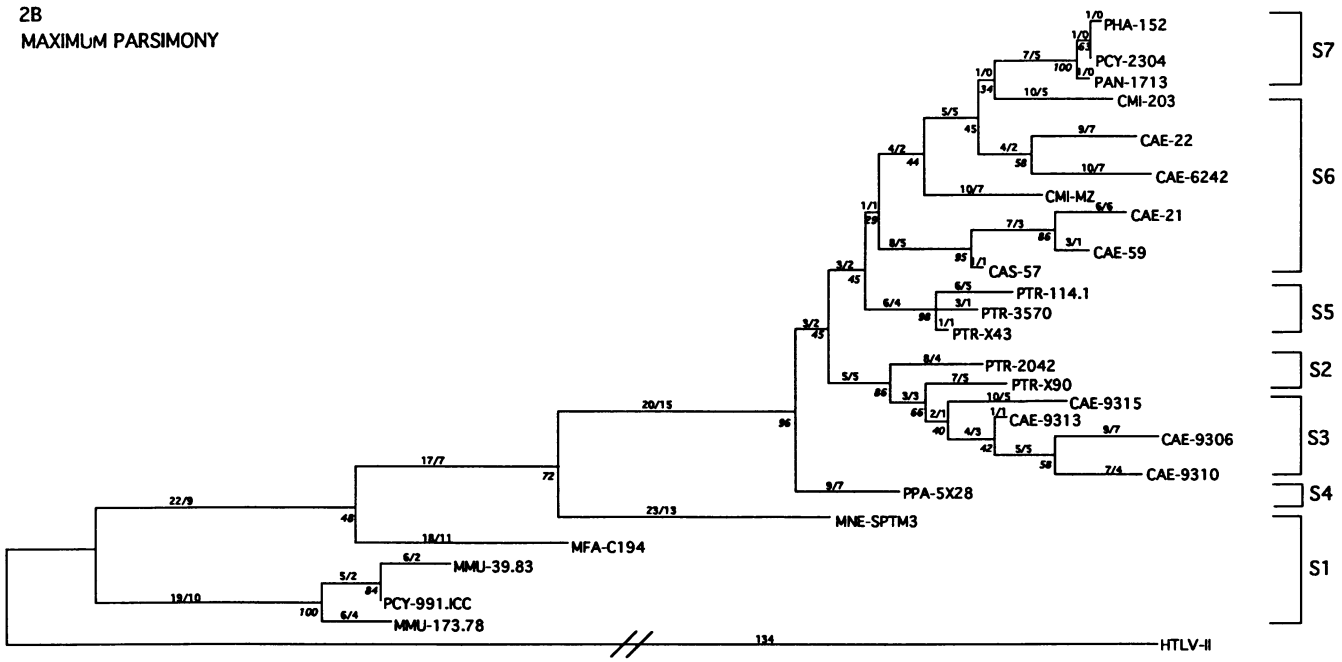
Sequences were transformed into character states for analysis by the maximum-parsimony method available in the PAUP (phylogenetic analysis using parsimony), version 3.1 (55, 56). Both neighbor-joining and PAUP were evaluated statistically for each tree, using 100 bootstrap iterations. This procedure provides an index of confidence for each topological node, as the fraction of bootstrap runs supporting the node. The maximum-likelihood analysis of representatives of each of the HTLV-I and STLV-I phylogenetic clusters was performed by using actual base pair frequencies of the data in conjunction with the DNAML program available in PHYLIP (10, 11). Evolutionary trees for each method were rooted with HTLV-II_{MO} (29) sequences. The strengths and limitations of the different phylogenetic methods are discussed elsewhere (10, 21, 45, 46, 55, 56).

Nucleotide sequence accession number. GenBank and EMBL accession numbers are U03122, U03124, and U03126 to U03160 for the 24 STLV-I sequences and the 13 HTLV-I sequences.

RESULTS

Phylogenetic reconstruction of HTLV-I sequences. We constructed phylogenetic trees based on aligned nucleotide sequences of 522 bp from the *env* gene of HTLV-I. The gene segment was amplified from genomic DNA of infected patients and spans the carboxy-terminal region of gp46 and nearly all of the gp21 coding gene. The pairwise genetic divergence, represented by both percent sequence divergence or by Kimura distance values are presented in Table 3. Different HTLV-I sequence comparisons range from 0 to 9% divergence and were 39 to 40% divergent from the HTLV-II outgroup sequence.

Evolutionary phenograms of HTLV-I sequences based on Kimura distances (Table 3) were constructed by the neighbor-



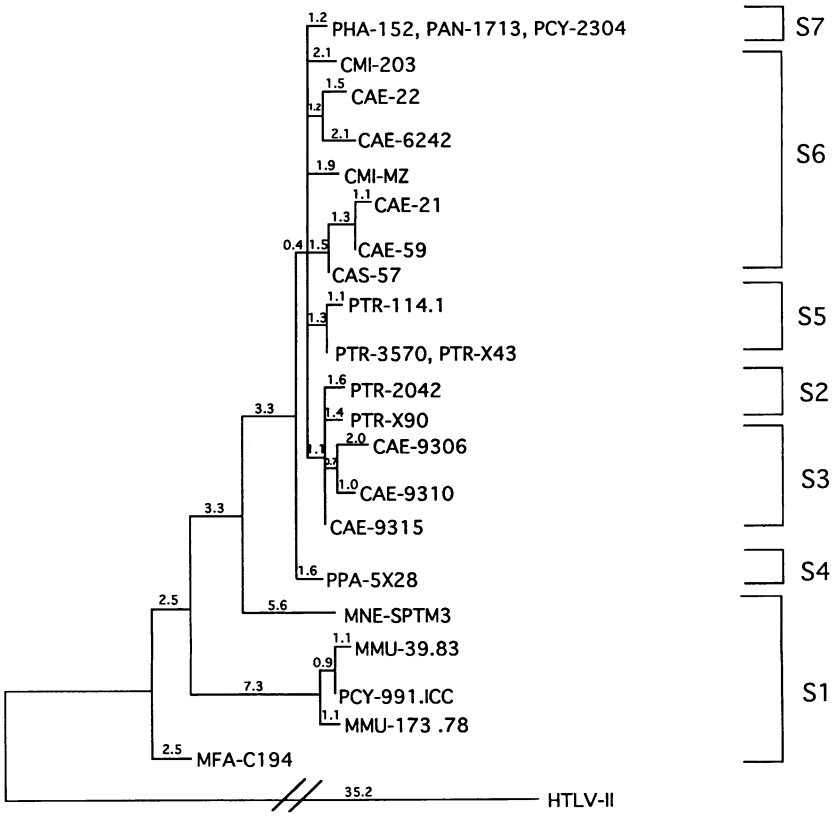


FIG. 2—*Continued.*

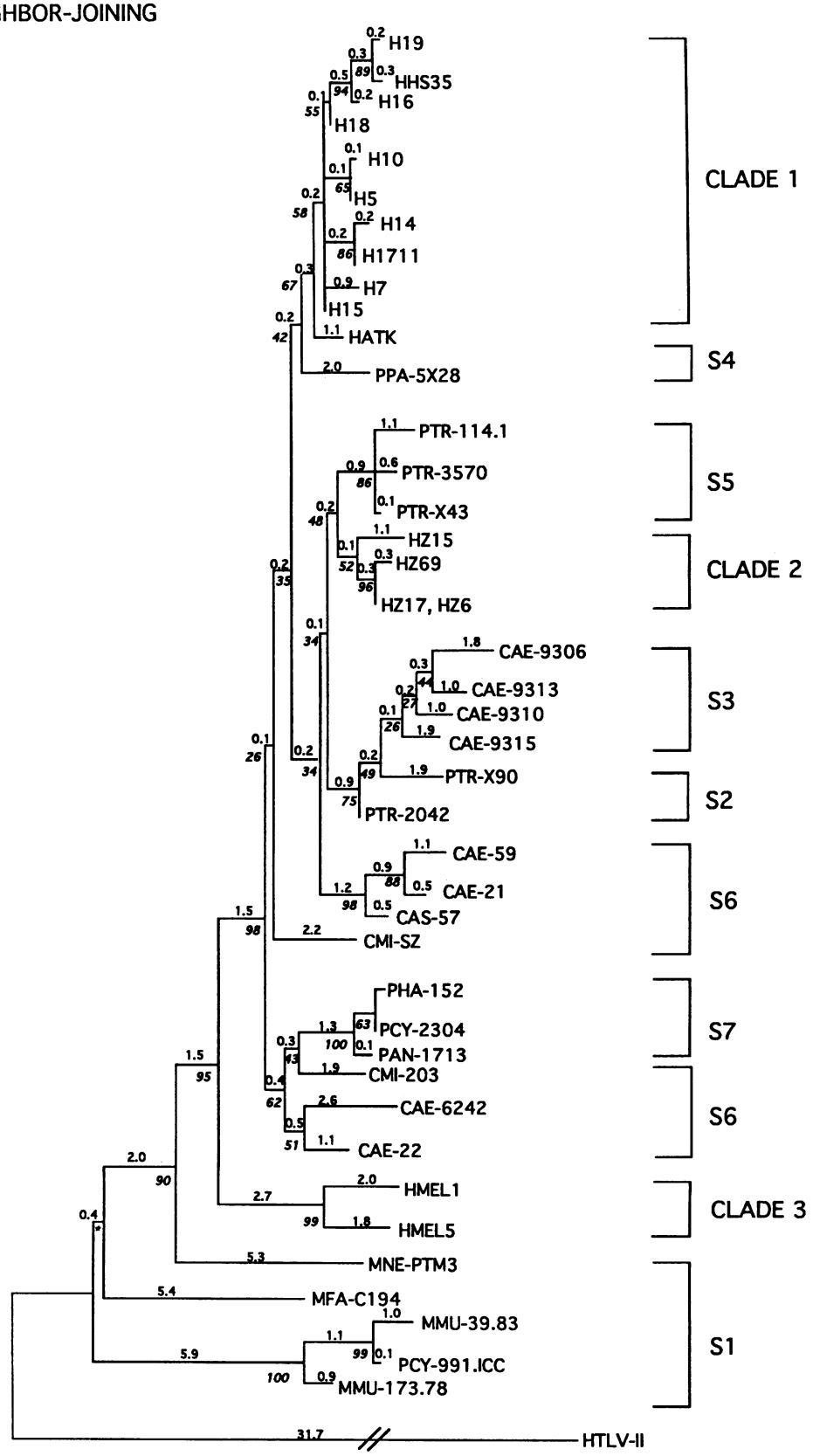


FIG. 3. Neighbor-joining tree of all human and simian isolates used in this study, using distance values from Table 3. Values on tree limbs represent percent sequence divergence. Numbers in italics are numbers of trees within 100 bootstrap replications which support the adjacent node.

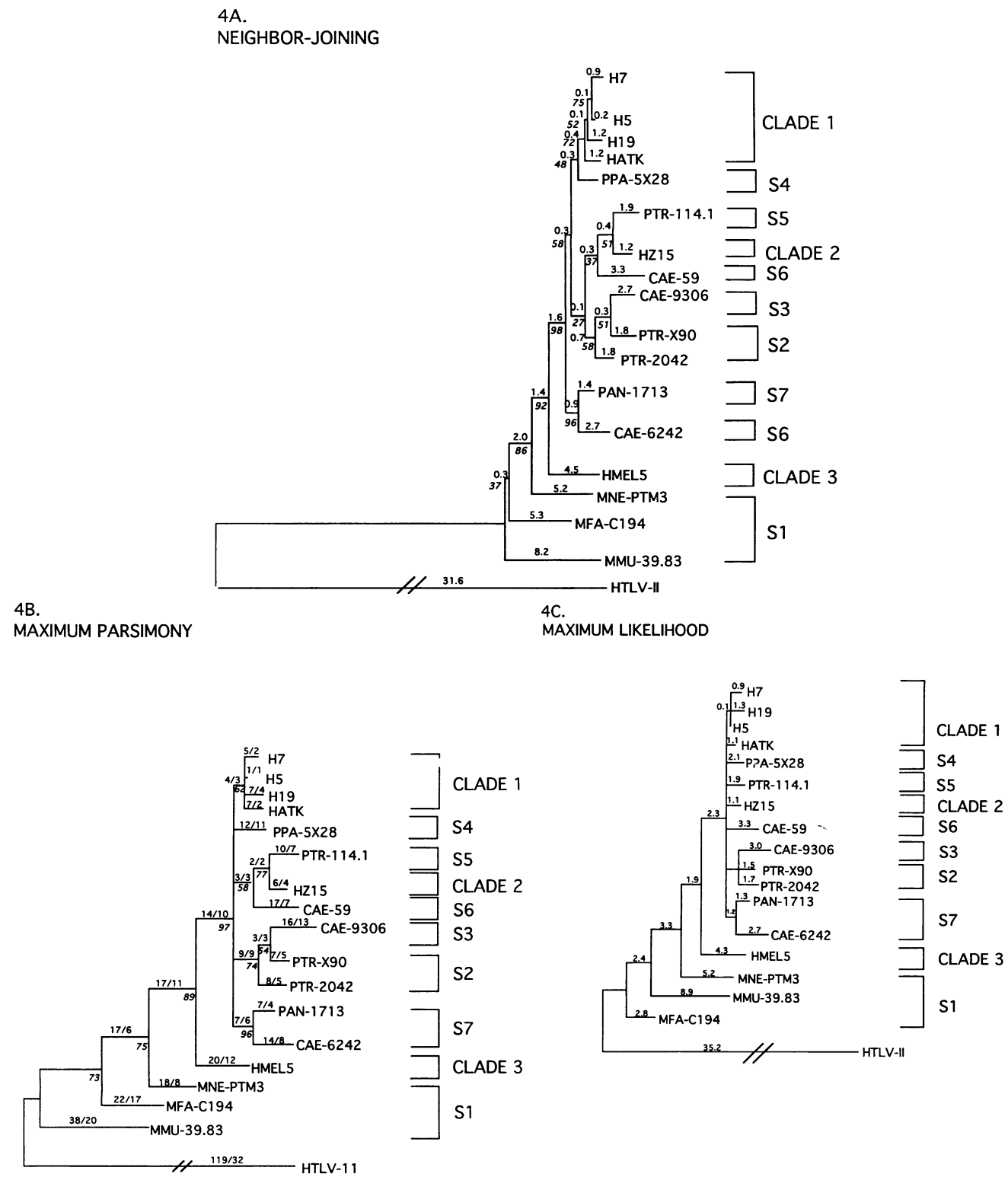


FIG. 4. Phylogenetic analysis of 18 isolates representing the 10 major clades of STLV-I and HTLV-I. (A) Neighbor-joining tree based on distance values given in Table 3. Limb lengths correspond to percent sequence divergence. Numbers in italics are bootstrap values (100 iterations) in support for the adjacent node. (B) Maximum-parsimony tree representing the majority rule consensus tree derived from bootstrap analysis and 1 of 29 trees of equivalent length (432 steps; consistency index, 0.678) determined by a heuristic search. (C) Maximum-likelihood tree (in likelihood, −2689.06; 4,154 trees examined) with a transition/transversion ratio of 2. Limb lengths approximate the expected number of substitutions per site × 100. Branch lengths not significantly different from zero were collapsed into polytomies.

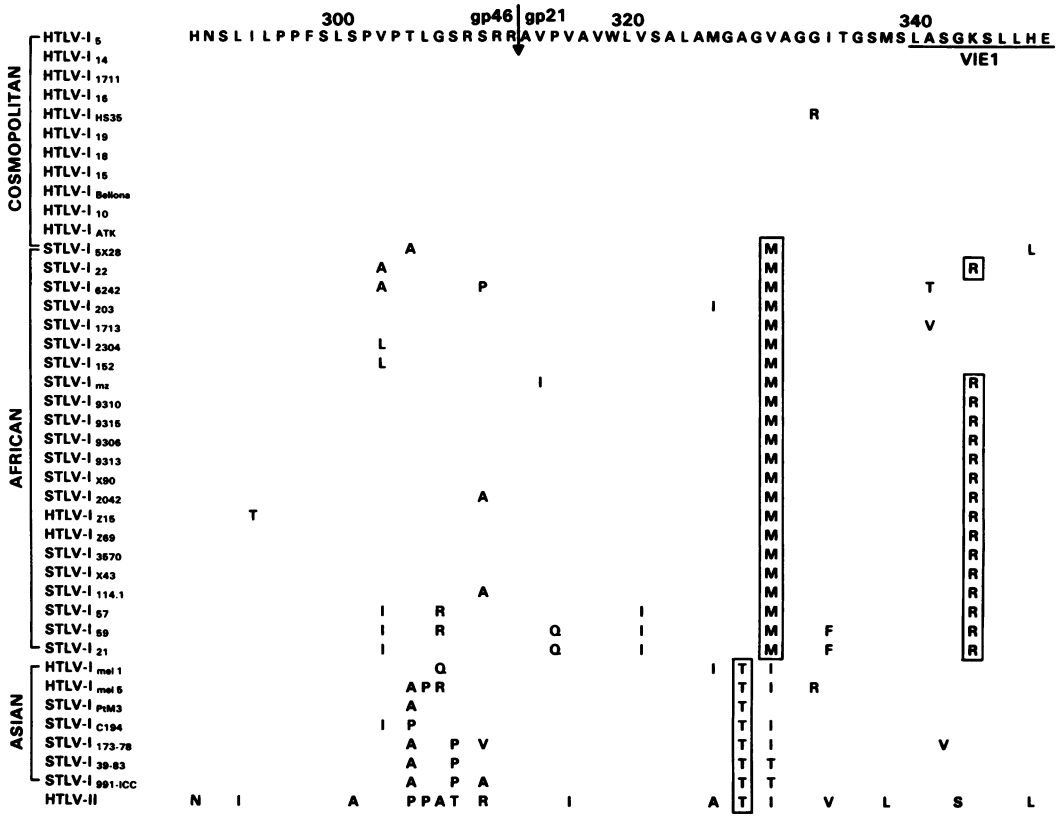


FIG. 5. Alignment of the putative amino acid envelope sequences of the cosmopolitan, Zairian, and Melanesian HTLV-I variants with the African and Asian STLV-I isolates and HTLV-II. Single-letter amino acid code was used. Asterisks indicate translational termination codon. Blanks indicate sequence identity with HTLV-I_5. Shared amino acids in the boxes are between the Zairian HTLV-I and the African STLV-I at positions 330 and 344 and among the Melanesian HTLV-I, the Asian STLV-I, and HTLV-II at position 328 underline the major conserved amino acid. A stretch of complete amino acid identity among all the isolates studied is indicated at positions 437 through 448. VIE1 and the underlined sequence epitope identified correspond to a specific linear epitope identified at the carboxy terminus of the gp21 (25).

joining method with bootstrap resampling methods (Fig. 1A). The nucleotide sequence data were also treated as character states and analyzed for production of minimum-length trees under the principle of maximum parsimony, using the PAUP program (Fig. 1B). Finally, the aligned HTLV-I sequences were analyzed for the phylogenetic relationship on the basis of the principle of maximum likelihood (Fig. 1C). We used three different phylogenetic methods to increase the reliability of the derived topologies, since tree-building algorithms rely on different assumptions (11, 21, 46, 56).

The neighbor-joining trees of HTLV-I sequences (Fig. 1A) resolved the strains into three distinct phylogenetic clusters. The first to diverge from a common ancestor were the two isolates from Melanesia (HTLV-I-Melanesia). These sequences were distantly related (7.3 to 9.1% divergence) from all other HTLV-I sequences. The second cluster was a group of four sequences from Zaire (HTLV-I-Zaire), and the third cluster was a group of 11 sequences which were geographically heterogeneous but closely related (HTLV-I-cosmopolitan). The three clusters were supported by the bootstrap iterations at levels of 75, 94, and 77%, respectively (Fig. 1A).

The existence of the three HTLV-I clusters was strongly corroborated by maximum parsimony (Fig. 1B, bootstrap = 60, 97, and 79%, respectively) and by maximum-likelihood analysis (Fig. 1C). Evolutionary trees constructed by each analytical method indicated that the HTLV-I-Melanesia clade was the earliest divergence preceding the split between the HTLV-I-Zaire and HTLV-I-cosmopolitan groups. Statistically significant resolution of branching order within the HTLV-I-cosmopolitan clade was not possible because of the low level of genetic diversity among sequences within each group.

Phylogenetic analysis of STLV sequences. Twenty-five STLV-I sequences from 12 nonhuman primate species (Table 2) were aligned separately, and their pairwise genetic divergence was estimated (Table 3). Each of the phylogenetic analyses revealed seven groups (or clades) within the STLV-I sequences examined. The clades were identified on the basis of consistent topological association in each of the three phylogenetic analyses (Fig. 2) plus the combining of clustered sequences from species of a single primate genus (e.g., clade S7 includes sequences from three *Papio* species) and separation of sequences from different simian genera (e.g., clade S2 and S3 from *Pan* and *Cercopithecus* species, respectively). The single exception is the grouping in S1 of five ancestral Asian monkey species STLV-I sequences (see below). A phenetic (neighbor-joining) analysis (Fig. 2A) indicated that the earliest radiation consisted of a heterogeneous mixture of five divergent sequences isolated from three species of *Macaca* from India and Asia and one yellow baboon (*Papio cynocephalus*) from Russia (clade S1). Three of these isolates, MMU-39.83, MMU-73.78, and PCY-991, form a monophyletic group that is genetically equidistant from the two other Asian macaque strains, MNE-

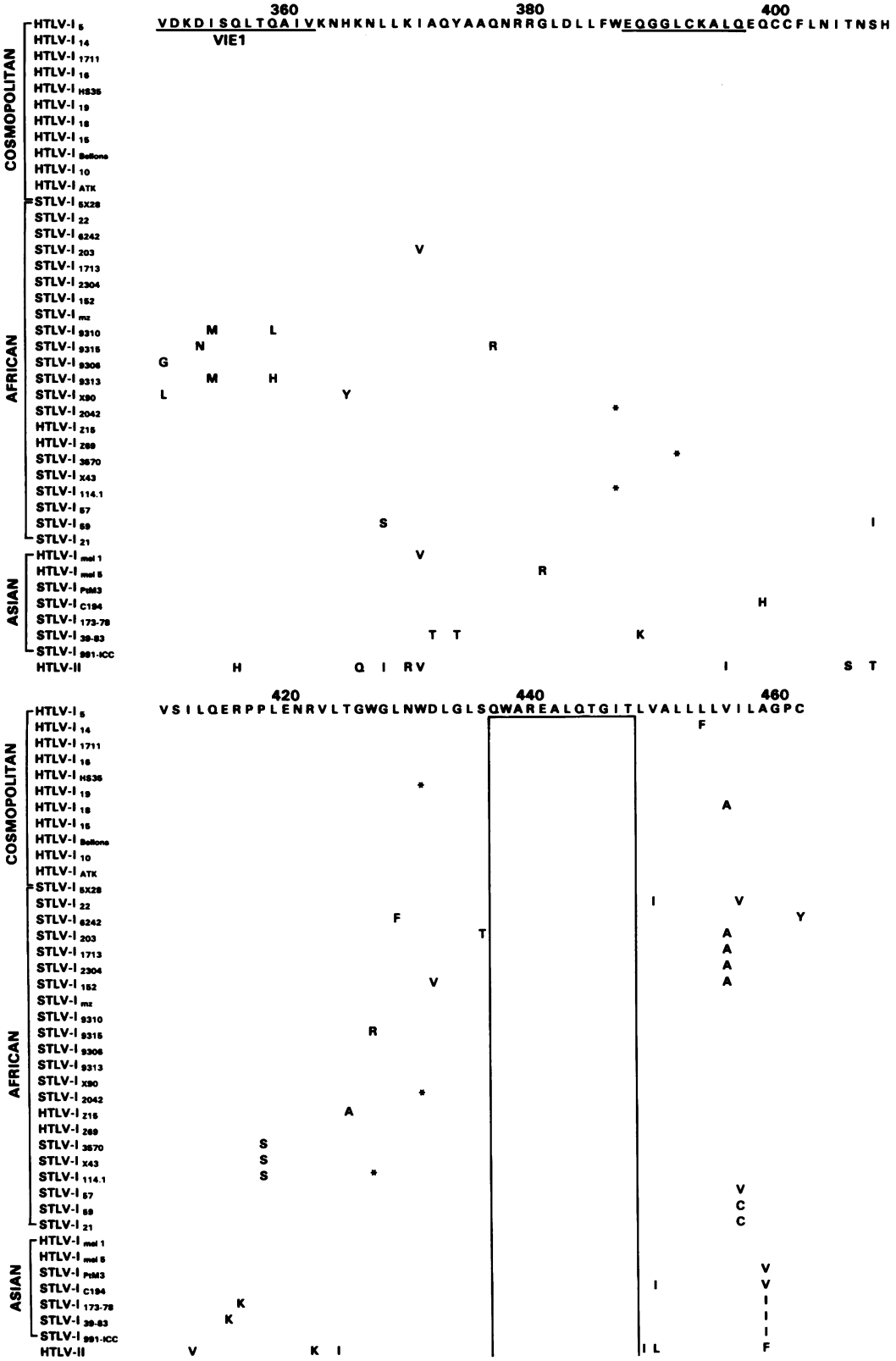


FIG. 5—*Continued.*

PTM3 and MFA-C194. The next major radiation consists of two groups: STLV strains from two chimpanzees (*Pan troglodytes*) at the Texas research facility, with one of indeterminate geographic origin and one from Sierra Leone (clade S2), and four sabeus monkeys (*Cercopithecus aethiops*) from Senegal (clade S3). The next group to diverge consists of a single individual representing the red baboon (*Papio papio*) from West Africa (clade S4). The placement of this sequence is not strongly supported (29%) by the bootstrap analysis, which instead puts this sequence distal to the radiation of another group of chimpanzees (one from Sierra Leone and two captive born but of unknown origin [clade S5]). A more recent radiation consisted of a mixture of STLV-I isolates from African species of *Cercopithecus* from Kenya, Ethiopia, Zaire, and Uganda (clade S6) and three species of baboon (*Papio*) from Kenya and Tanzania (clade S7 for STLV-East Africa). The seven clades are indicated in Table 2 and in Fig. 2.

The minimum-length tree for STLV-I nucleotide sequences produced with maximum parsimony (Fig. 2B) reaffirmed the existence of each of these seven clades. The only disagreement with the phenetic result involves the relative position of clade S4 (consisting of a single baboon sequence, PPA-5x28) which occurred as more primitive in both maximum-parsimony and maximum-likelihood trees (Fig. 2C). The maximum-likelihood tree also clustered the seven groups and reinforced the three-way divergence of the Asian macaque-baboon sequence within clade S1. In general, bootstrap support was moderate to good for each of the seven clades, but sequence strain differentiation was less well resolved within the identified clade groups.

The combined results of the three phylogenetic analyses of STLV-I strains lead to several important conclusions. First, there are several examples of STLV-I sequences from a single species that assort into genetically distinct clades. For example, STLV-I from *C. aethiops* occurs in two distinct phylogenetic clusters, S3 and S6; STLV-I from *P. troglodytes* occurs in two clades, S2 and S5; and STLV-I isolates from baboon (*Papio*) species occur in three distinct clades, S1, S4, and S7. Second, and conversely, STLV-I sequences within the seven different clades are composed of strains from related primate species. For example, clade 6 contains sequences from three distinct *Cercopithecus* species and clade 7 contains STLV-I from three species of *Papio*. These results suggest very recent interspecies transfer between species within these primate genera. Third, certain clades are themselves associated by monophyletic phylogenetic nodes, indicating a recency of common ancestry for the clades that is consistent with interspecies exchange between primates belonging to different genera. Thus, clade S2 (STLV-I from *P. troglodytes*) and clade S3 (STLV-I from *C. aethiops*) are strongly associated (high bootstrap significance, 70 and 80%, in all three analyses), indicating a recent cross-species exchange of STLV-I between chimpanzee and *Cercopithecus* species. Fourth, there are several examples of STLV-I sequences from one species being closer to STLV-I from a distant primate species than they are to a second clade from the same species. Thus, the two chimpanzee clades (S2 and S5) are more related to distinct STLV-I sequences from *Cercopithecus* species than they are to each other.

Taken together, the results of these analyses would be consistent with recent interspecies transfer of STLV-I sequences between distantly related (intergenera) primate species. The wide species distribution of distinct phylogenetic clusters within the genera *Pan*, *Papio*, and *Cercopithecus* suggests that more frequent, perhaps free, exchange has occurred below the genus level. Lastly, support for the interspecies transfer is evident by examination of the concordance of geographic origins of associated clades from distinct species. Thus, clades S6 and S7 are from sympatric African primates, as are species from associated clades S2 and S3. The deepest evolutionary divergence occurs between the three Asian groups within clade S1, and these come from both *Papio* and *Cercopithecus* species.

Combined phylogenetic analysis of HTLV-I and STLV-I sequences. A phenetic analysis of all HTLV-I and STLV-I sequences, using the Kimura distance matrix (Table 3), is presented in Fig. 3. The three HTLV-I and seven STLV-I clades were resolved by the topology, but the relative position of HTLV-I was polyphyletic with respect to the simian clades; that is, the HTLV-I clusters were mixed and interspersed within the STLV-I topology. Furthermore, the HTLV-I-cosmopolitan clade was most closely aligned (bootstrap, 34%) with the baboon clade S4 (PPA-5x28). The HTLV-I-Zaire group was aligned with the chimpanzee isolates of clade S5 (bootstrap, 58%), and the HTLV-I-Melanesia isolates were deeply nested within the nearly contemporaneous divergence of isolates from the three Asian groups with clade S1. These results affirm the evolution of three clades in the human species and suggest that at least three independent human-simian exchanges leading to the three human clades have occurred during the evolution of this virus.

A jackknife analysis of the neighbor-joining tree was performed to determine the stability of the placement of HTLV-I groups among those for STLV-I. Overall, topology did not change: HTLV-I-Zaire clearly clustered with clade S5, HTLV-I-Melanesia remained with the early Asian STLV-I divergence (S1 clade), and the STLV-I-west Africa sequence (clade S3) continued to precede the divergence of the HTLV-I-cosmopolitan sequences.

In order to analyze the combined data set with maximum parsimony and maximum likelihood, representative sequences were selected from each STLV-I and HTLV-I group. A subset of data was necessary because of limitations within the programs to distinguish among a large data set of closely related sequences. Results did not change with neighbor-joining with the 18 subset sequences, except for the movement of CAE-59 (*C. aethiops*) to the cluster of HTLV-I-Zaire and PTR-114.1 (Fig. 4A). With maximum parsimony, a single most parsimonious tree had the same topology as with neighbor joining. However, the results of a heuristic search with a bootstrap collapsed the internal nodes into a polytomy (Fig. 4B). Maximum-likelihood analysis revealed a consistent resolution and association of HTLV-I and STLV-I clades (Fig. 4C).

Amino acid sequence analysis of the Env proteins of HTLV-I and STLV-I. Nucleotide sequences of HTLV-I and STLV-I isolates were translated by computer and aligned (Fig. 5) to examine the pattern of amino acid sequence variation in the population. The results emphasize the extreme conservation of amino acids among these viral strains, as had been previously reported among HTLV-I-cosmopolitan isolates (17, 51). There is one region (amino acids 436 to 447) where there is 100% amino acid identity among all isolates. With a few exceptions (amino acids 328, 329, and 344 of the gp21 gene), the variation is spotty and dispersed throughout the sequence and the isolates. The nonrandom positions reinforce the divergence of and are diagnostic for HTLV-I-cosmopolitan, Asian STLV-I plus HTLV-I, and African STLV-I. For example, at position 328, all Asian sequences have a threonine compared with a glycine in all other isolates. Similarly, the three groups have distinct shared, derived residues (cladistic synapomorphies) at position 329.

DISCUSSION

The phylogenetic analysis of *env* nucleotide sequence divergence between HTLV-I and STLV-I strains in nature provided a robust method for interpreting the natural history of these viruses. In general, several important conclusions can be derived from analyses presented in Fig. 1 to 4. First, the sequence comparison led to the recognition of three HTLV-I and seven STLV-I phylogenetic clusters. The association of these 10 groups was apparent with several phylogenetic approaches and also provided a relative chronology of divergence, with Asian species (human and simian) being the oldest and the cosmopolitan HTLV-I among the more recent viral radiations. Second, neither HTLV-I nor STLV-I sequence divergence conformed to host species monophyly; rather, different clades within the same species had as closest relatives clades from more divergent species. For example, an HTLV-I clade from Zaire (four strains) was most closely related to an STLV-I clade (S5) consisting of three strains from chimpanzees. Human strains had two additional distantly related HTLV-I clades, and chimpanzees have at least one divergent STLV-I clade. These associations are interpreted as evidence for separate evolutionary divergence of the ancestral viruses for the clades (likely in different species, but possibly in the same species in different geographic locales), followed by transmission to modern species that now retain multiple phylogenetic clades. Third, the deeply rooted divergence of Asian isolates (human and simian) reflecting the greatest pairwise sequence divergence suggests an origin of these viruses in Asia and then migration to African primates and later spread of the cosmopolitan HTLV-I throughout the world. Fourth, there are several associations that suggest rare, but significant, transmission of viral ancestors between primate genera (*Homo-Pan*, *Papio-Homo*, *Papio-Cercopithecus*, and *Homo-Macaca*). We cannot exclude that some of these may be very recent in primate facilities. Finally, there seems to be more frequent transmission between species within certain genera (e.g., *Cercopithecus* spp. and *Papio* species). This should not be overly surprising, since many of these species divergences are rather recent, <200,000 years ago, which is of the order of divergence of human racial groups. As these groups may freely interbreed, virus transmission in sympatric locales (e.g., Africa) would be common.

The simplest scenario for the natural history of these viruses that is consistent with phylogenetic, epidemiologic, and geographic data would be some variation on the following. The ancestors of HTLV-I and STLV-I first entered primates (perhaps from an artiodactyl) somewhere in Asia and were transmitted to multiple species of Asian primates. The primate-adapted STLV-I infection migrated to Africa, where it infected several primate genera, notably *Pan*, *Cercopithecus*, *Papio*, and *Homo*. Meanwhile, somewhere in southeast Asia, HTLV-I emerged in Melanesia. In Africa, the STLV-I ancestors crossed genus barriers only a few times, but transmission between species in the same genus was more common. Recent facility in human migratory pattern, including the slave trade, led to dissemination of the cosmopolitan HTLV-I strain as worldwide.

This hypothesis must be considered as tentative, as several unanswered questions remain. These include the following: in which direction did each various virus transmission occur? Although it is likely that the original virus infected nonhuman simian ancestors that eventually infected humans, the directions of other transmissions are difficult to know for certain. In which species did the 10 clades evolve? In which geographic locale? What was the timing of the primary intergenus transfers? Answering this last question with molecular data is ambiguous because of uncertainty of the assumption of a constant molecular clock in retrovirus sequences. Calibration of each clock is difficult, since retroviral sequences, especially those which persist through speciation events, will saturate mutational substitutions and render clocklike assumptions invalid. Nonetheless, the provocative interplay of retroviral sequences offers some compelling insight into the natural history of these sequences which contributes to our further understanding of the coevolution of pathogens and their hosts (1, 37).

ACKNOWLEDGMENTS

We are grateful to all the support personnel in the primate centers that made the animal samples available; to R. Desrosiers, G. de Thé, D. Messinger, and H. M. Mc Clure for providing well-characterized animal samples; to L. Anderson for editorial assistance; and to A. F. Voevodin for helpful discussion.

We acknowledge the National Cancer Institute for allocation of computing time and staff support at the Frederick Biomedical Supercomputing Center of the Frederick Cancer Research and Development Center. I. J. Koralnik is a recipient of a grant from the Swiss National Fund for Scientific Research.

REFERENCES

1. **Anderson, R. M., and R. M. May.** 1991. Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford.
2. **Bastian, I. B., J. Gardner, D. Webb, R. R. Doherty, I. Gardner, and K. S. Sriprakash.** 1991. Isolation of an HTLV-I strain from Australian aboriginals. 4th International Conference on Human Retrovirology: HTLV. Montego Bay, Jamaica.
3. **Becker, W. B., M. L. B. Becker, T. Homma, H. D. Brede, and R. Kurth.** 1985. Serum antibodies to human T-cell leukaemia virus type I in different ethnic groups and in non-human primates in South Africa. S. Afr. Med. J. **67:**445–449.
4. **Blakeslee, J. R., Jr., H. M. McClure, D. C. Anderson, R. M. Bauer, L. Y. Huff, and R. G. Olsen.** 1987. Chronic fatal disease in gorillas seropositive for simian T-lymphotropic virus I antibodies. Cancer Lett. **37:**1–6.
5. **Botha, M. C., M. Jones, W. A. DeKlerk, and N. Yamamoto.** 1985. Spread and distribution of human T-cell leukaemia virus type I-reactive antibody among baboons and monkeys in the northern and eastern Transvaal. S. Afr. Med. J. **67:**665–668.
6. **Daniel, M. D., N. L. Letvin, P. K. Sehgal, D. K. Schmidt, D. P. Silva, K. R. Solomon, F. S. Hodi, Jr., D. J. Ringler, R. D. Hunt, N. W. King, and R. C. Desrosiers.** 1988. Prevalence of antibodies to 3 retroviruses in a captive colony of macaque monkeys. Int. J. Cancer **41:**601–608.
7. **De, B. K., M. D. Lairmore, K. Griffis, L. J. Williams, F. Villanger, T. C. Quinn, C. Brown, N. Nzilambi, M. Sugimoto, S. Araki, and T. M. Folks.** 1991. Comparative analysis of nucleotide sequences of the partial envelope gene (5′ domain) among human T lymphotropic virus type I (HTLV-I) isolates. Virology **182:**413–419.
8. **Devereaux, J.** 1991. Sequence analysis software package program manual, version 7.0. Genetics Computer Group, Madison, Wis.
9. **Dracopoli, N. C., T. R. Turner, J. G. Else, C. J. Jolly, R. Anthony, R. C. Gallo, and W. C. Saxinger.** 1986. STLV-I antibodies in feral populations of East African vervet monkeys (*Cercopithecus aethiops*). Int. J. Cancer **38:**523–529.
10. **Felsenstein, J.** 1990. PHYLIP (phylogeny inference package), version 3.3. University of Washington, Seattle.
11. **Felsenstein, J.** 1992. Phylogenies from restriction sites: a maximum-likelihood approach. Evolution **46:**159–173.
12. **Fukasawa, M., H. Tsujimoto, K. Ishikawa, T. Miura, B. Ivanoff, R. W. Cooper, E. Frost, E. Delaporte, J. A. A. Mingle, F. C. Grant, and M. Hayami.** 1987. Human T-cell leukemia virus type 1 isolates from Gabon and Ghana: comparative analysis of proviral genomes. Virology **161:**315–320.
13. **Gallo, R. C.** 1986. The first human retrovirus. Sci. Am. **255:**88–98.
14. **Gallo, R. C., A. Sliski, and F. Wong-Staal.** 1983. Origin of human T-cell leukaemia-lymphoma virus. Lancet **ii:**962–963.

15. **Gessain, A., F. Barin, J. Vernant, O. Gout, L. Maurs, A. Calender, and G. deThe.** 1985. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. Lancet **ii:**407–409.
16. **Gessain, A., E. Boeri, R. Yanagihara, R. C. Gallo, and G. Franchini.** 1993. Complete nucleotide sequence of a highly divergent human T-cell leukemia (lymphotropic) virus type I (HTLV-I) variant from Melanesia: genetics and phylogenetic relationship to HTLV-I strains from other geographical regions. J. Virol. **67:** 1015–1023.
17. **Gessain, A., R. C. 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. **66:**2288–2295.
18. **Gessain, A., R. Yanagihara, G. Franchini, R. M. Garruto, C. L. Jenkins, A. B. Ajdukiewicz, R. C. Gallo, and D. C. Gajdusek.** 1991. Highly divergent molecular variants of human T-lymphotropic virus type I from isolated populations in Papua New Guinea and the Solomon Islands. Proc. Natl. Acad. Sci. USA **88:**7694–7698.
19. **Guo, H. G., F. Wong-Staal, and R. C. Gallo.** 1984. Novel viral sequences related to human T-cell leukemia virus in T cells of a seropositive baboon. Science **223:**1195–1197.
20. **Harrison, R. E., K. K. Murthy, G. I. Melendez, P. A. Tuck, and J. W. Eichberg.** Submitted for publication.
21. **Hasegawa, M., and M. Fujiwara.** 1993. Relative efficiencies of the maximum likelihood, maximum parsimony, and neighbor-joining methods for estimating protein phylogeny. Mol. Phylogenet. Evol. **2:**1–5.
22. **Hayami, M., A. Komuro, K. Nozawa, T. Shotake, K. Ishikawa, K. Yamamoto, T. Ishida, S. Honjo, and Y. Hinuma.** 1984. Prevalence of antibody to adult T-cell leukemia virus-associated antigens (ATLA) in Japanese monkeys and other non-human primates. Int. J. Cancer **33:**179–183.
23. **Hinuma, Y., K. Nagata, M. Hanaoka, M. Nakai, T. Matsumoto, K. Kinoshita, S. Shirakawa, and I. Miyoshi.** 1981. Adult T-cell leukemia: antigen in an ATL cell line and detection of antibodies to the antigen in human sera. Proc. Natl. Acad. Sci. USA **78:**6476–6480.
24. **Homma, T., P. J. Kanki, R. D. Hunt, N. W. King, M. J. O'Connell, N. L. Letvin, M. D. Daniel, R. C. Desrosiers, C. S. Yang, and M. Essex.** 1984. Lymphoma in macaques: association with virus of human T lymphotropic family. Science **225:**716–718.
25. **Horal, P., W. W. Hall, P. Svennerholm, J. Lycke, S. Jeansson, L. Rymo, M. H. Kaplan, and A. Vahlne.** 1991. Identification of type-specific linear epitopes in the glycoproteins gp46 and gp21 of human T-cell leukemia viruses type I and type II using synthetic peptides. Proc. Natl. Acad. Sci. USA **88:**5754–5758.
26. **Hunsmann, G., J. Schneider, J. Schmitt, and N. Yamamoto.** 1983. Detection of serum antibodies to adult T-cell leukemia virus in non-human primates and in people from Africa. Int. J. Cancer **32:**329–332.
27. **Inoue, J. I., T. Watanabe, M. Sato, A. Oda, K. Toyoshima, M. Yoshida, and M. Seiki.** 1986. Nucleotide sequence of the protease-coding region in an infectious DNA of simian retrovirus (STLV) of the HTLV-I family. Virology **150:**187–195.
28. **Ishikawa, K., M. Fukasawa, H. Tsujimoto, J. G. Else, M. Isahakia, N. K. Ubhi, T. Ishida, O. Takenaka, Y. Kawamoto, T. Shotake, H. Ohsawa, B. Ivanoff, R. W. Cooper, E. Frost, F. C. Grant, Y. Spriatna, Y. Sutarman, K. Abe, K. Yamamoto, and M. Hayami.** 1987. Serological survey and virus isolation of simian T-cell leukemia/T-lymphotropic virus type I (STLV-I) in non-human primates in their native countries. Int. J. Cancer **40:**233–239.
29. **Kalyanaraman, V. S., M. G. Sarngadharan, M. Robert-Guroff, I. Miyoshi, D. Blayney, D. 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.
30. **Kimura, M.** 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:**111–120.
31. **Komurian, F., F. Pelloquin, and G. deThe.** 1991. In vivo genomic variability of human T-cell leukemia virus type 1 depends more upon geography than upon pathologies. J. Virol. **65:**3770–3778.
32. **Komuro, A., T. Watanabe, I. Miyoshi, M. Hayami, H. Tsujimoto, M. Seiki, and M. Yoshida.** 1984. Detection and characterization of simian retroviruses homologous to human T-cell leukemia virus type 1. Virology **138:**373–378.
33. **Lee, R. V., A. W. Prowten, S. K. Satchidanand, and B. I. S. Srivastava.** 1985. Non-Hodgkin's lymphoma and HTLV-1 antibodies in a gorilla. N. Engl. J. Med. **312:**118–119.
34. **Miyoshi, I., S. Yoshimoto, M. Fujishita, H. Taguchi, I. Kubonishi, K. Niiya, and M. Minezawa.** 1982. Natural adult T-cell leukemia virus infection in Japanese monkeys. Lancet **ii:**658.
35. **Monath, T. P.** 1990. Yellow fever, p. 797–802. *In* B. N. Fields, D. M. Knipe, R. M. Chanock, M. S. Hirsch, J. L. Melnick, T. P. Monath, and B. Roizman (ed.), Virology. Raven Press, New York.
36. **Mone, J., E. Whitehead, M. Leland, G. Hubbard, and J. S. Allan.** Simian T-cell leukemia virus type 1 infection in captive baboons. AIDS Res. Hum. Retroviruses, in press.
37. **Morse, S. S. (ed.).** 1992. Emerging viruses. Oxford University Press, New York.
38. **Needleman, S. B., and C. D. Wunsch.** 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48:**443–453.
39. **Olmsted, R. A., R. Langley, M. E. Roelke, R. M. Goeken, D. Adger-Johnson, J. P. Goff, J. P. Albert, C. Packer, M. K. Laurenson, T. M. Caro, L. Scheepers, D. E. Wildt, M. Bush, J. S. Martenson, and S. J. O'Brien.** 1992. Worldwide prevalence of lentivirus infection in wild feline species: epidemiologic and phylogenetic aspects. J. Virol. **66:**6008–6018.
40. **Osame, M., K. Usuku, S. Izumo, N. Ijichi, H. Amitini, and A. Igata.** 1986. HTLV-I associated myelopathy, a new clinical entity. Lancet **i:**1031–1032.
41. **Paine, E., J. Garcia, T. C. Philpott, G. Shaw, and L. Ratner.** 1991. Limited sequence variation in human T-lymphotropic virus type 1 isolates from North American and African patients. Virology **182:**111–123.
42. **Parrish, C. R., C. F. Aquadro, M. L. Strassheim, J. F. Evermann, J.-Y. Sgro, and H. O. Mohhammed.** 1991. Rapid antigenic-type replacement and DNA sequence evolution of canine parvovirus. J. Virol. **65:**6544–6552.
43. **Poiesz, B. J., 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. USA **77:**7415–7419.
44. **Rodgers-Johnson, P., D. C. Gajdusek, O. S. Morgan, V. Zaninovic, P. S. Sarin, and D. S. Graham.** 1985. HTLV-I and HTLV-II antibodies and tropical spastic paraparesis. Lancet **ii:**1247–1248.
45. **Saitou, N., and T. Imanishi.** 1989. Relative efficiencies of the Fitch-Margolish, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. Mol. Biol. Evol. **6:**514–525.
46. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
47. **Sakakibara, I., Y. Sugimoto, A. Sasagawa, S. Honjo, H. Tsujimoto, H. Nakamura, and M. Hayami.** 1986. Spontaneous malignant lymphoma in an African green monkey naturally infected with simian T-lymphotropic virus (STLV). J. Med. Primatol. **15:**311–318.
48. **Saksena, N. K., V. Herve, J. P. Durand, B. LeGuenno, O. M. Diop, J. P. Digoutte, C. Mathiot, M. C. Muller, J. L. Love, S. Dube, M. P. Sherman, P. M. Benz, S. Erensoy, A. Galat-Luong, G. Galat, B. Paul, D. K. Dube, F. Barre-Sinoussi, and B. J. Poiesz.** Seroepidemiologic, molecular and phylogenetic analyses of simian T-cell leukemia viruses (STLV-I) from various naturally infected monkey species from central and western Africa. Virology, in press.
49. **Saksena, N. K., V. Herve, M. P. Sherman, J. P. Durand, C. Mathiot, M. Muller, J. L. Love, F. B. Sinoussi, D. K. Dube, and B. J. Poiesz.** 1993. Sequence and phylogenetic analysis of a new STLV-I from a naturally infected Tantalus monkey from Central Africa. Virology **192:**312–320.
50. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.

51. **Schulz, T. F., M. L. Calabro, J. G. Hoad, C. V. F. Carrington, E. Matutes, D. Catovsky, and R. A. Weiss.** 1991. HTLV-I envelope sequences from Brazil, the Caribbean, and Romania: clustering of sequences according to geographic origin and variability in an antibody epitope. Virology **184:**483–491.
52. **Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida.** 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA **80:**3618–3622.
53. **Sherman, M. P., N. K. Saksena, D. K. Dube, R. Yanagihara, and B. J. Poiesz.** 1992. Evolutionary insights on the origin of human T-cell lymphoma/leukemia virus type I (HTLV-I) derived from sequence analysis of a new HTLV-I variant from Papua New Guinea. J. Virol. **66:**2556–2563.
54. **Siegl, G.** 1984. Biology and pathogenicity of autonomous parvoviruses, p. 297–347. *In* K. I. Beras (ed.), The parvoviruses. Plenum Press, New York.
55. **Swofford, D. L.** 1984. Phylogenetic analysis using parsimony (PAUP). Illinois Natural History Survey, Champaign.
56. **Swofford, D. L., and G. J. Olsen.** 1990. Phylogeny reconstruction, p. 411–515. *In* Molecular systematics. Sinauer, Sunderland.
57. **Tsujimoto, H., Y. Noda, K. Ishikawa, H. Nakamura, M. Fukasawa, I. Sakakibara, A. Sasagawa, S. Honjo, and M. Hayami.** 1986. Development of adult T-cell leukemia-like disease in an African green monkey associated with clonal integration of simian T-cell leukemia virus type-1. Cancer Res. **47:**269–274.
58. **Voevodin, A. F., B. A. Lapin, L. A. Yakovleva, T. I. Ponomaryeva, T. E. Oganyan, and E. N. Razmadze.** 1985. Antibodies reacting with human T-lymphotropic retrovirus (HTLV-I) or related antigens in lymphomatous and healthy hamadryas baboons. Int. J. Cancer **36:**579–584.
59. **Watanabe, T., M. Seiki, Y. Hirayama, and M. Yoshida.** 1986. Human T-cell leukemia virus type 1 is a member of the African subtype of simian viruses (STLV). Virology **148:**385–388.
60. **Watanabe, T., M. Seiki, H. Tsujimoto, I. Miyoshi, M. Hayami, and M. Yoshida.** 1985. Sequence homology of the simian retrovirus genome with human T-cell leukemia virus type 1. Virology **144:** 59–65.
61. **Yanagihara, R., V. R. Nerurkar, and A. B. Ajdukiewicz.** 1991. Comparison between strains of human T lymphotropic virus type I isolated from inhabitants of the Solomon Islands and Papua New Guinea. J. Infect. Dis. **164:**443–449.
62. **Yanagihara, R., V. R. Nerurkar, R. M. Garruto, M. A. Miller, M. E. Leon-Monzon, C. L. Jenkins, R. C. Sanders, P. P. Liberski, M. P. Alpers, and D. C. Gajdusek.** 1991. Characterization of a variant of human T-lymphotropic virus type I isolated from a healthy member of a remote, recently contacted group in Papua New Guinea. Proc. Natl. Acad. Sci. USA **88:**1446–1450.