J Mol Evol (1990) 31:493–499

Journal of Molecular Evolution

Molecular Evolution of Human T-Cell Leukemia Virus

Yasuo Ina and Takashi Gojobori

National Institute of Genetics, Mishima 411, Japan

Summary. Phylogenetic trees for the human T-cell leukemia virus type I (HTLV-I) and its related viruses were constructed by use of nucleotide sequences of the long terminal repeat (LTR) and the *tax* gene. The trees showed that the viruses diverged from a common ancestral virus and that they are classified into two groups whose hosts are either primates or bovines. However, the topology of the trees for the viruses differed from that for the hosts. This suggests that HTLV-I and HTLV-I-related viruses evolved independently of host-species divergence and that interspecies transmission between human and monkeys occurred in the past. The nucleotide diversity of the *tax* genes of HTLV-I was estimated to be 0.025. This value is more than 10 times larger than that of human globin genes, but it is about 20 times smaller than that of hemagglutinin genes of influenza A viruses. Thus, the genetic variability of the HTLV-I genes seems to be higher than that of nuclear genes but much lower than the genes of typical RNA viruses. Furthermore, we examined functional constraints on the overlapping region of the *rex* and *tax* genes. The results obtained imply that for the overlapping region, the *tax* gene has much stronger constraints against amino acid changes than the *rex* gene.

Key words: Human T-cell leukemia virus — *rex* gene — *tax* gene — Phylogenetic tree — Nucleotide diversity

Introduction

Adult T-cell leukemia (ATL) is a virus-associated disease with T-lymphocyte malignancies in adults. ATL is geographically distributed in the limited areas of the world: southwestern Japan (Hinuma et al. 1981, 1982), the Caribbean basin (Blatter et al. 1982), and equatorial Africa (Hunsmann et al. 1983). Human T-cell leukemia virus type I (HTLV-I) was discovered as the etiological agent of ATL by Poiesz et al. (1980) and Hinuma et al. (1981). Studies of familial occurrence of ATL and the prevalence of antibodies of HTLV-I in family members implied that HTLV-I was transmitted by vertical infection, particularly by some kinds of intimate contact. It is now known that the main routes of natural transmission of HTLV-I are mother's milk and sexual contact. However, the reason for the endemic distribution of ATL remains unknown.

There has been much speculation regarding the origin of HTLV-I. HTLV-I-specific antibodies were found in Japanese macaques (Miyoshi et al. 1982; Hayami et al. 1984). Moreover, simian T-cell leukemia viruses (STLVs), which are related to HTLV-I, were detected in various nonhuman primates such as chimpanzee, pig-tailed monkey, and gorilla (Miyoshi et al. 1982, 1983; Hayami et al. 1983, 1984; Yamamoto et al. 1983, 1984; Guo et al. 1984; Komuro et al. 1984; Saxinger et al. 1984). Thus, it is possible that HTLV-I originated from interspecies transmission of STLV between monkeys and humans. Besides these STLVs, bovine leukemia virus (BLV) was discovered from cows as the HTLV-I-related virus. Furthermore, human T-cell leukemia virus type II (HTLV-II) was isolated from a patient with hairy T-cell leukemia.

In this paper, we constructed phylogenetic trees for the HTLV family in order to elucidate the evolutionary origin of HTLV-I. The trees suggest that the HTLV family evolved independently of host-species divergence and that interspecies transmission between human and monkeys occurred in the past. We also estimated the nucleotide diversity of

Offprint requests to: T. Gojobori

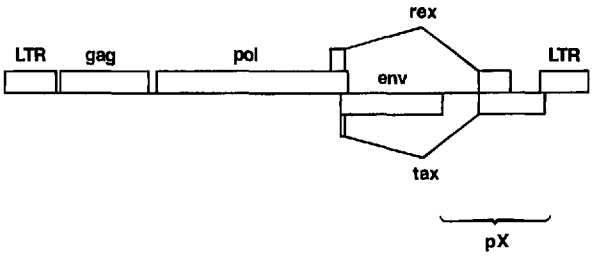

Fig. 1. Genomic structure of the HTLV family. The *tax* and *rex* genes are located in the *pX* region. They are translated by the different reading frame.

the *tax* gene of HTLV-I for the purpose of evaluating the degree of genetic variability for this virus. We show that the genetic variability of HTLV-I is much lower than that of influenza A viruses. Furthermore, we examined the functional constraints in the overlapping region of the *tax* and *rex* genes. The result obtained implies that for the overlapping region the *tax* gene has much stronger constraints against amino acid changes than the *rex* gene.

Nucleotide Sequences Used

All of the HTLV-I and related viruses belong to the family Retroviridae. Figure 1 shows that their provirus, a DNA form embedded in the host genome, has the structure of long terminal repeat (LTR)–*gag*–*pol*–*env*–*pX*–LTR in this order from the 5′ end of the genome. This structure is unique to the HTLV family. The entire genome is about 9000 bp long. The LTRs are located at both 5′ and 3′ ends of the provirus. They regulate the expression of the viral genes and work as promoter, enhancer, and terminator of transcription. The *gag* gene encodes the viral core proteins. The *pol* gene is the largest one of the viral genes. It encodes a reverse transcriptase and an endonuclease. These enzymes play an important role in reverse transcription and integration of the virus into the host genome. The *env* gene encodes the surface glycoprotein and the transmembrane protein. The former protein is required when the virus infects the host cell. The *pX* region is unique to the HTLV family. It contains three kinds of partially overlapping genes; *rex, tax* and $p21^{x\text{-}III}$. The *rex* gene regulates viral transcription at the post-transcriptional level. The *tax* gene product is associated with viral replication and transformation of the host cell. The function of $p21^{x\text{-}III}$ is not well known at present.

In the present study, we used LTR and the *tax* gene in order to examine the evolutionary relationships among HTLV-I and HTLV-I-related viruses. This is because LTR and the *tax* gene have been sequenced in the largest numbers of isolates of the viruses (Table 1).

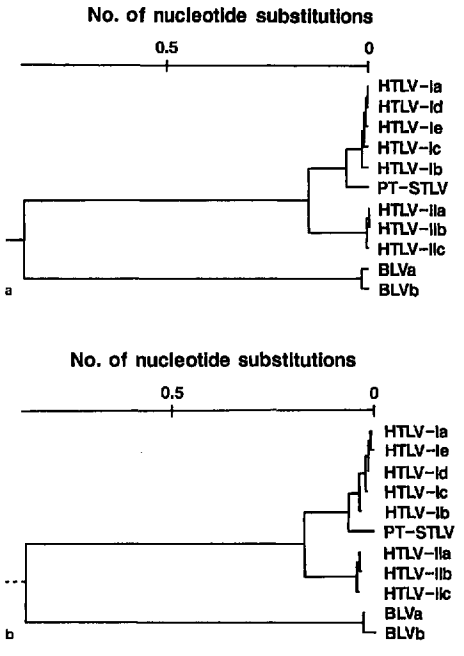

Fig. 2. Phylogenetic trees for the HTLV family constructed by the UPG method (a; Nei 1975, 1987) and the N-J method (b; Saitou and Nei 1987), using the numbers of nucleotide substitutions at all positions of codons for the *tax* gene.

Results and Discussion

Evolutionary Relationships Among Members of the HTLV Family

The nucleotide sequences for the *tax* genes of HTLV-I and the HTLV-I-related viruses were aligned with each other to maximize the sequence homology. The number of nucleotide substitutions at all positions of codons was estimated for each pair of viral isolates by the six-parameter method (Gojobori et al. 1982). Using these estimates, a phylogenetic tree was constructed by the unweighted pair-group (UPG) method (Sneath and Sokal 1973; Nei 1975, 1987).

The phylogenetic tree constructed shows that an ancestral virus of the HTLV family diverged into two groups, depending on the hosts (Fig. 2a). One group consists of primate viruses [HTLV-Ia, HTLV-Ib, HTLV-Ic, HTLV-Id, HTLV-Ie, PT-STLV (pig-tailed monkey STLV), HTLV-IIa, HTLV-IIb, and HTLV-IIc] and the other group is bovine viruses (BLVa and BLVb). After the ancestral virus of the HTLV family diverged into the two groups, the HTLV-II isolates (HTLV-IIa, HTLV-IIb, and HTLV-IIc) first diverged from an ancestor of the HTLV-I group and PT-STLV. Then, separation between the HTLV-I group and PT-STLV took place, and finally divergence within the HTLV-I group (HTLV-Ia, HTLV-Ib, HTLV-Ic, HTLV-Id, and HTLV-Ie) occurred. This branching order of the phylogenetic tree does not follow the evolution of

Table 1. Seventeen isolates of the HTLV family used in the present study

Isolate[a]	Host	Clone	Location[b]	Reference for nucleotide sequence[c]
HTLV-Ia	Human	λ ATK-1	Japan	[1]
HTLV-Ib		λ MC-1	Zaire	[2]
HTLV-Ic		HS-35	Caribbean	[3]
HTLV-Id		pATK08	Japan	[4]
HTLV-Ie		pHTLV707	Japan	[5]
HTLV-If		pATM-3	Japan	[6]
HTLV-Ig		pCR1	USA	[7]
HTLV-IIa	Human	λ H6.0	USA	[8]
HTLV-IIb		pH6-B3.5	USA	[9]
HTLV-IIc		MO15A	USA	[10]
HTLV-IId		pH6-R0 .8	USA	[11]
HTLV-IIe		MO15A	USA	[12]
AG-STLV	African green monkey		Africa	[13]
CH-STLV	Chimpanzee		Africa	[13]
PT-STLV	Pig-tailed monkey	λ PT-2	Southeast Asia	[14]
BLVa	Bovine	λ BLV-1	Japan	[15]
BLVb			Belgium	[16]

[a] Isolates within a given group of viruses are named alphabetically

[b] It represents the location where a strain was isolated

[c] Data source: [1] Seiki et al. (1983), [2] Ratner et al. (1985), [3] Malik et al. (1988), [4] Seiki et al. (1985), [5] Takeuchi et al. (1985), [6] Seiki et al. (1982), [7] Josephs et al. (1984), [8] Shimotohno et al. (1985), [9] Shimotohno et al. (1984b), [10] Haseltine et al. (1984), [11] Shimotohno et al. (1984a), [12] Sodroski et al. (1984), [13] Watanabe et al. (1986), [14] Watanabe et al. (1985), [15] Sagata et al. (1985), [16] Rice et al. (1984)

the host species, implying that interspecies transmission between human and monkeys occurred in the past. However, this implication is based on the assumption that nucleotide substitutions occurred at a constant rate.

To examine whether the tree topology remains the same when the rate of nucleotide substitution has changed during evolution, we also constructed a phylogenetic tree by the neighbor-joining (N-J) method (Fig. 2b; Saitou and Nei 1987). This method does not require the assumption of rate constancy. The root of this tree was taken as the midpoint of the longest path, which was between HTLV-Ie and BLVb. This tree has the same topology as the tree constructed by the UPG method, except for the positions of HTLV-Id and HTLV-Ie. Moreover, all branching lengths remain almost unchanged. It suggests that the HTLV family evolved with a relatively constant rate and that the HTLV-II isolates diverged from the common ancestor of PT-STLV and the HTLV-I isolates. Furthermore, we constructed the phylogenetic tree by use of the numbers of synonymous (silent) substitutions (Nei and Gojobori 1986), showing that it has the same topology (data not shown). This topology is also supported by the tree constructed by the maximum likelihood method (Fukami and Tateno 1990).

Although the phylogenetic trees constructed above showed the possibility of interspecies transmission between human and monkeys, the tree contains only one nonhuman primate virus, PT-STLV. To confirm this possibility, we also constructed the phylogenetic tree for LTRs by the N-J method, because the nucleotide sequences of LTRs for other nonhuman primate viruses are now available. The tree shown in Fig. 3 indicated that the human virus (HTLV-I) and the monkey viruses [CH (chimpanzee)-STLV, AG (African green monkey)-STLV, and PT-STLV] diverged from another human virus (HTLV-II). Moreover, the evolutionary relationships between CH-STLV, AG-STLV, PT-STLV, and HTLV-I are completely different from those between respective host species in the primate evo-

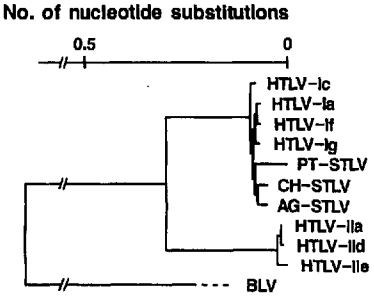

Fig. 3. Phylogenetic tree for the HTLV family constructed by the N-J method, using the numbers of nucleotide substitutions for LTR. Because the numbers of nucleotide substitutions were unreliably large for some comparisons between BLV and other viruses, we first constructed a phylogenetic tree excluding BLV, and then the root of this tree was determined with use of the BLV sequence.

Table 2. Nucleotide diversity for HTLV-I compared with those for human[a] and influenza virus

Organism	Gene	Sample size	Number of sites compared (bp)	Nucleotide diversity
HTLV-I	*tax*	5	548[b]	0.025
Human	β-globin[c]	50	35,000	0.002
	Growth hormone[c]	52	50,000	0.002
	Insulin	2	1431	0.003
Influenza A virus	Hemagglutinin	12	320	0.510

[a] The estimates of nucleotide diversities for man and influenza A virus are from Nei (1987)

[b] The overlapping region was excluded

[c] The estimate was obtained by the restriction enzyme technique

lution. It also supports the view that the HTLV family evolved independently of the host-species divergence. However, it does not rule out the possibility that the divergence between HTLV-I and HTLV-II was earlier than that between human and monkeys.

If the divergence of HTLV-I and STLV was much earlier than that of human and monkeys, the tree topology for the viruses may not be identical to that for the hosts. If this is the case, the rate of nucleotide substitution for HTLV-I should be slower than or nearly equal to that for the host genome. However, it seems less likely that HTLV-I has evolved at such a rate because the virus replicates itself by reverse transcriptase, which is known to be more error-prone than the polymerase of the host.

The above discussion should be taken with caution because we assumed that all nonhuman primate viruses have infected their hosts in the wild. In particular, no HTLV-I carriers have been found so far in Southeast Asia, where pig-tailed monkeys (host of PT-STLV) inhabit. It is thus possible that the interspecies transmission of the viruses between human and monkeys should have occurred a long time ago and then HTLV-I may have disappeared from the human population in Southeast Asia. Of course, we cannot exclude the possibility that the pig-tailed monkey was infected by human viruses by some other mechanism.

Nucleotide Diversity for HTLV-I

It was reported that the rates of nucleotide substitutions for RNA viruses are about 10^6 times higher than those for DNA genomes because of a high mutation rate and a high replication frequency (Holland et al. 1982; Gojobori and Yokoyama 1985; Hayashida et al. 1985; Saitou and Nei 1986). However, no one has reported the rate of nucleotide substitution for the HTLV family. This was due to the difficulty of knowing the divergence time between the viruses they compared. In fact, HTLV-I can be embedded in the DNA genome of the host as provirus for a long time, although it replicates itself as an RNA virus.

In order to evaluate the degree of genetic variability for the HTLV family, we computed the nucleotide diversity (Nei and Li 1979) for the *tax* gene of HTLV-I. As shown in Table 2, the nucleotide diversity for HTLV-I is about 10 times higher than that for the host genome. The high nucleotide diversity for HTLV-I may be attributed to a high mutation rate in reverse transcription. However, the nucleotide diversity for HTLV-I is much lower than that for influenza A virus (Nei 1987). Because HTLV-I can be embedded in the host genome as provirus, as mentioned earlier, the replication frequency of the HTLV-I genome may not be as high as that of the influenza viral genome.

It is reported that the rate of nucleotide substitution for human immunodeficiency virus (HIV) is as high as that for influenza A virus (Yokoyama and Gojobori 1987; Li et al. 1988). Although HTLV-I and HIV both are human retroviruses, it is known that the latent time for HTLV-I within the DNA is much longer than that for HIV. These features of HTLV-I may account for the observation that the nucleotide diversity of HTLV-I is much lower than that of influenza A virus.

Previous studies showed that interspecies transmission of HIV between human and monkeys occurred about a few hundred years ago (Yokoyama and Gojobori 1987; Sharp and Li 1988). Judging from the results on nucleotide diversity for HTLV-I, we speculate that interspecies transmission of T-cell leukemia virus between human and monkeys may have occurred much earlier than that of the AIDS virus.

Functional Constraints on the Overlapping Region of the rex and tax Genes

The *rex* and *tax* genes are located in the *pX* region. These genes are translated into their respective amino acid sequences by different but overlapping reading frames (Seiki et al. 1985; Nagashima et al. 1986). Thus, these overlapping genes are exposed to two different kinds of functional constraints at the protein level. One is against amino acid changes of its own product and another is against that of the counterpart's product. It is generally known that proteins with a biologically important function have stronger constraints against amino acid changes than proteins with a less important function (Kimura 1983).

To examine which of the *rex* and *tax* genes is more constrained in the overlapping region, we es-

Table 3. Numbers of synonymous and nonsynonymous substitutions per site for the overlapping regions of the *rex* and *tax* genes

Comparison[a]	*tax*		*rex*	
	SYN[b]	NON[c]	SYN	NON
HTLV-Ia vs HTLV-Ib	0.048 (±0.020)	0.003 (±0.003)	0.000 (±0.000)	0.019 (±0.007)
HTLV-Ia vs HTLV-Ic	0.016 (±0.011)	0.000 (±0.000)	0.000 (±0.000)	0.005 (±0.004)
HTLV-Ib vs HTLV-Ic	0.065 (±0.023)	0.003 (±0.003)	0.000 (±0.000)	0.025 (±0.008)
HTLV-Ib vs HTLV-Id	0.048 (±0.020)	0.003 (±0.003)	0.000 (±0.000)	0.019 (±0.007)
HTLV-Ib vs HTLV-Ie	0.053 (±0.031)	0.000 (±0.000)	0.000 (±0.000)	0.018 (±0.010)
HTLV-Ic vs HTLV-Id	0.016 (±0.011)	0.000 (±0.000)	0.000 (±0.000)	0.005 (±0.004)
HTLV-Ic vs HTLV-Ie	0.017 (±0.017)	0.000 (±0.000)	0.000 (±0.000)	0.006 (±0.006)
Average	0.038 (±0.020)	0.001 (±0.002)	0.000 (±0.000)	0.014 (±0.007)

[a] No substitutions were observed between HTLV-Ia and HTLV-Id, HTLV-Ia and HTLV-Ie, and HTLV-Id and HTLV-Ie
[b] Number and standard deviation of synonymous substitutions per site
[c] Number and standard deviation of nonsynonymous substitutions per site

timated the numbers of synonymous and nonsynonymous (amino acid-altering) substitutions for each gene in the overlapping region by Nei and Gojobori's method (Nei and Gojobori 1986). For the *tax* gene, the number of synonymous substitutions is larger than that of nonsynonymous substitutions for any comparisons (Table 3). For the *rex* gene, on the other hand, no synonymous substitutions were observed, and all the nucleotide substitutions occurred at nonsynonymous sites for any comparisons. These results suggest that in their overlapping region, the *tax* gene has much stronger constraints than the *rex* gene.

To evaluate the difference in degrees of the functional constraints between the overlapping and nonoverlapping regions for the *tax* gene, we computed the ratio of the number of nonsynonymous substitutions of the overlapping region to that of the nonoverlapping region. The ratio obtained was less than 1/10 for most comparisons (data not shown). This value is more than two times smaller than other kinds of overlapping genes that have been examined previously (Miyata and Yasunaga 1978). Thus, the observed ratio may not be caused only by the fact that the *tax* gene overlaps with the *rex* gene. The overlapping region of the *tax* gene may encode a more important functional domain of its product than the nonoverlapping region.

Ninety percent of the *rex* gene overlaps with the *tax* gene. As mentioned earlier, for the overlapping region of the *rex* gene, nucleotide substitutions occurred only at nonsynonymous sites. However, the *rex* gene product is thought to be essential for expression and replication of the HTLV-I genes (Inoue et al. 1987). The nonoverlapping region of the *rex* gene shows a high homology even between BLV and HTLV-I, HTLV-II, or PT-STLV, whereas the other region shows only a low homology between them (data not shown). These features lead us to speculate that the nonoverlapping region of the *rex* gene product may be its functional domain, although it is encoded by only 18 codons.

The details of the evolution of the HTLV family should be studied more extensively. Biological functions of the regulatory genes such as *tax* and *rex* should also be elucidated more precisely. To carry out this kind of study, it will be important to accumulate molecular data for the HTLV family, including nucleotide sequences and epidemiological data for various strains of HTLV and STLV.

Acknowledgments. We thank Dr. E.N. Moriyama for her helpful comments. This research was supported by a Grant-in-Aid from the Ministry of Education, Science, and Culture of Japan.

References

Blatter WA, Kalyanaraman VS, Robert-Guaroff M, Lister A, Galton DAG, Sarin PS, Crawford MH, Catovsky D, Greaves M, Gallo RC (1982) The human type-C retrovirus HTLV in blacks from the Caribbean region and relationship to adult T-cell leukemia/lymphoma. Int J Cancer 30:257–264

Fukami K, Tateno Y (1990) Maximum likelihood tree estimation against different patterns of base substitutions. J Mol Evol (in press)

Gojobori T, Yokoyama S (1985) Rates of evolution of the retroviral oncogene of Moloney murine sarcoma virus and of its cellular homologues. Proc Natl Acad Sci USA 82:4198–4210

Gojobori T, Ishii K, Nei M (1982) Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. J Mol Evol 18:414–423

Guo HG, Wong-Staal F, Gallo RC (1984) Novel viral sequence related to human T cell leukemia virus in T cells of a seropositive baboon. Science 223:1195–1197

Haseltine WA, Sodroski J, Patarca R, Briggs D, Perkins D, Wong-Staal F (1984) Structure of 3′ terminal region of type II human T lymphotropic virus: evidence for new coding region. Science 225:419–421

Hayami M, Ishikawa K, Komuro A, Shida T, Hinuma Y (1983) ATLV antibody in cynomolgus monkeys in the wild. Lancet 2:620

Hayami M, Komuro A, Nozawa K, Shotake T, Ishikawa K, Yamamoto K, Ishida T, Honjo S, Hinuma Y (1984) Prevalence of anti-adult T cell leukemia virus (ATLV) antibody in Japanese monkeys and other non-human primates. Int J Cancer 33:179–183

Hayashida H, Toh H, Kikuno R, Miyata T (1985) Evolution of influenza virus genes. Mol Biol Evol 2:289–303

Hinuma Y, Nagata K, Hanaoka M, Misoka M, Nakai M, Matsumoto T, Kinoshita K, Shirakawa S, Miyoshi I (1981) Adult T-cell leukemia: antigen in a ATL cell line and detection of antibodies to the antigen in human sera. Proc Natl Acad Sci USA 78:6476–6480

Hinuma Y, Komoda H, Chosa T, Kondo T, Kohakura M, Takenaka T, Kikuchi M, Ichimaru M, Yunoki K, Sato I, Matsuo R, Takiuchi Y, Uchino H, Hanaoka M (1982) Antibodies to adult T-cell leukemia-virus-associated antigen (ATLA) in sera from patients with ATL and controls in Japan: a nationwide seroepidemiological study. Int J Cancer 29:631–635

Holland J, Spindler F, Horodyski F, Grabau E, Nichol S, Vandepol S (1982) Rapid evolution of RNA genomes. Science 215:1577–1585

Hunsmann G, Schneider J, Schmitt J, Yamamoto N (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

Inoue JI, Yoshida M, Seiki M (1987) Transcriptional ($p40^x$) and post-transcriptional ($p27^{x\text{-}III}$) regulators are required for the expression and replication of human T-cell leukemia virus type I genes. Proc Natl Acad Sci USA 84:3653–3657

Josephs SF, Wong-Staal F, Manzari V, Gallo RC, Sodoroski JG, Trus MD, Perkins D, Patarca R, Haseltine WA (1984) Long terminal repeat structure of an American isolate of type I human T-cell leukemia virus. Virology 139:340–345

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Komuro A, Watanabe T, Miyoshi I, Hayami M, Tsujimoto H, Seiki M, Yoshida M (1984) Detection and characterization of simian retroviruses homologous to human T-cell leukemia virus type I. Virology 138:373–378

Li WH, Tanimura M, Sharp PM (1988) Rates and dates of divergence between AIDS virus nucleotide sequences. Mol Biol Evol 5:313–330

Malik KTA, Even J, Karpas A (1988) Molecular cloning and complete nucleotide sequence of an adult T cell leukemia virus/human T cell leukemia virus type I (ATLV/HTLV-I) isolate of Caribbean origin: relationship to other members of the ATLV/HTLV-I subgroup. J Gen Virol 69:1695–1710

Miyata T, Yasunaga T (1978) Evolution of overlapping genes. Nature 272:532–535

Miyoshi I, Yoshitomo S, Fujishita M, Taguchi H, Kubonishi I, Niiya K, Minezawa M (1982) Natural adult T-cell leukemia virus infection in Japanese monkeys. Lancet 11:658

Miyoshi I, Fujishita M, Taguchi H, Matsubayashi K, Miwa N, Tanioka Y (1983) Natural infection in non-human primates with adult T-cell leukemia virus or a closely related agent. Int J Cancer 32:333–336

Nagashima K, Yoshida M, Seiki M (1986) A single species of *pX* mRNA of human T-cell leukemia virus type I encodes *trans*-activator $p40^x$ and two other phosphoproteins. J Virol 60:394–399

Nei M (1975) Molecular population genetics and evolution. North-Holland, Amsterdam

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC (1980) Detection and isolation of type C retrovirus particle from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc Natl Acad Sci USA 77: 7415–7419

Ratner L, Josephs SF, Starcich B, Hahn B, Shaw GM, Gallo RC, Wong-Staal F (1985) Nucleotide sequence analysis of a variant human T-cell leukemia virus (HTLV-Ib) provirus with a deletion in pX-I. J Virol 54:781–790

Rice NR, Stephens RM, Couez D, Deschamps J, Kettmann R, Burny A, Gilden RV (1984) The nucleotide sequence of the *env* gene and post-*env* region of bovine leukemia virus. Virology 138:82–93

Sagata N, Yasunaga T, Tsuzuku-Kawamura J, Ohishi K, Ogawa Y, Ikawa Y (1985) Complete nucleotide sequence of the genome of bovine leukemia virus: its evolutionary relationship to other retroviruses. Proc Natl Acad Sci USA 82:677–681

Saitou N, Nei M (1986) Polymorphism and evolution of influenza A virus genes. Mol Biol Evol 3:57–74

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Saxinger WC, Lange-Wantzin G, Thompson K, Lapin B, Yokovleva L, Li Y, Guo H, Robert-Gutoff MJ, Blattner WA, Ito Y, Gallo RC (1984) In: Gallo RC, Essex M, Gross L (eds) Human T-cell leukemia/lymphoma virus, Cold Spring Harbor Laboratory, Cold Spring Harbor NY, pp 323–330

Seiki M, Hattori S, Yoshida M (1982) Human adult T-cell leukemia virus: molecular cloning of the provirus DNA and the unique terminal structure. Proc Natl Acad Sci USA 79: 6899–6902

Seiki M, Hattori S, Hirayama Y, Yoshida M (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

Seiki M, Hikikoshi A, Taniguchi T, Yoshida M (1985) Expression of the *pX* gene of HTLV-I: general splicing mechanism in the HTLV family. Science 228:1532–1534

Sharp PM, Li WH (1988) Understanding the origins of AIDS viruses. Nature 336:315

Shimotohno K, Golde DW, Miwa M, Sugimura T, Chen ISY (1984a) Nucleotide sequence analysis of the long terminal repeat of human T-cell leukemia virus type II. Proc Natl Acad Sci USA 81:1079–1083

Shimotohno K, Wachsman W, Takahashi Y, Golde DW, Miwa M, Sugimura T, Chen ISY (1984b) Nucleotide sequence of the 3′ region of an infectious human T-cell leukemia virus type II genome. Proc Natl Acad Sci USA 81:6657–6661

Shimotohno K, Takahashi Y, Shimizu N, Gojobori T, Golde DW, Chen ISY, Miwa M, Sugimura T (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 USA 82:3101–3105

Sneath PHA, Sokal RR (1973) Numerical taxonomy. Freeman, San Francisco

Sodroski J, Trus M, Perkins D, Patarca R, Wong-Staal F, Gelmann E, Gallo R, Haseltine WA (1984) Repetitive structure in the long-terminal-repeat element of a type II human T-cell leukemia virus. Proc Natl Acad Sci USA 81:4617–4621

Takeuchi K, Kobayashi N, Nam SH, Yamamoto N, Hatanaka M (1985) Molecular cloning of cDNA encoding gp68 of adult T-cell leukemia-associated antigen: evidence for expression of the *pX* IV region of human T-cell leukemia virus. J Gen Virol 66:1825–1829

Watanabe T, Seiki M, Tsujimoto H, Miyoshi I, Hayami M, Yoshida M (1985) Sequence homology of the simian retrovirus genome with human T-cell leukemia virus type I. Virology 144:59–65

Watanabe T, Seiki M, Hirayama Y, Yoshida M (1986) Human T-cell leukemia virus type I is a member of the African subtype of simian virus (STLV). Virology 148:385–388

Yamamoto N, Hinuma Y, Hausen HZ, Schneider J, Hunmann G (1983) African green monkeys are infected with adult T-cell leukemia virus or closely related agent. Lancet 1:240–241

Yamamoto N, Kobayashi N, Takeuchi K, Koyanagi Y, Hatanaka M, Hinuma Y, Chosa T, Schneider J, Hunsmann G (1984) Int J Cancer 34:77

Yokoyama S, Gojobori T (1987) Molecular evolution and phylogeny of the human AIDS viruses LAV, HTLV-III, and ARV. J Mol Evol 24:330–336

Received February 15, 1990/Revised June 20, 1990