Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 1124–1127, February 1994
Evolution

Phylogenetic subtypes of human T-lymphotropic virus type I and their relations to the anthropological background

TOMOYUKI MIURA[a], TAKUYA FUKUNAGA[a], TATSUHIKO IGARASHI[a,b], MASAHIRO YAMASHITA[a], EIJI IDO[a], SHIN-ICHI FUNAHASHI[a], TAKAFUMI ISHIDA[c], KEIKO WASHIO[c], SHINTAROU UEDA[c], KEN-ICHIRO HASHIMOTO[d,e], MITSUAKI YOSHIDA[d], MITSUHIRO OSAME[e], BHIM SEN SINGHAL[f], VLADIMIR ZANINOVIC[g], LUIS CARTIER[h], SHUNRO SONODA[e], KAZUO TAJIMA[i], YASUO INA[j], TAKASHI GOJOBORI[j], AND MASANORI HAYAMI[a,k]

[a]Institute for Virus Research, Kyoto University, Kyoto 606, Japan; [b]Research Institute for Microbial Disease, Osaka University, Osaka 565, Japan; [c]Faculty of Science, University of Tokyo, Tokyo 113, Japan; [d]Institute of Medical Science, University of Tokyo, Tokyo 108, Japan; [e]Faculty of Medicine, Kagoshima University, Kagoshima 890, Japan; [f]Bombay Hospital, Institute of Medical Sciences, Bombay, India; [g]Faculty of Medicine, University of Valle, Cali, Colombia; [h]Faculty of Medicine, University of Chile, Santiago, Chile; [i]Aichi Cancer Center Research Institute, Nagoya 464, Japan; and [j]National Institute of Genetics, Mishima 411, Japan

Communicated by Motoo Kimura, October 22, 1993

ABSTRACT **Isolates of human T-lymphotropic virus type I (HTLV-I) were phylogenetically analyzed from native inhabitants in India and South America (Colombia and Chile) and from Ainu (regarded as pure Japanese descendants from the preagricultural "Jomon" period). Their genomes were partially sequenced together with isolates from Gabon in central Africa and from Ghana in West Africa. The phylogenetic tree was constructed from the sequence data obtained and those of previously reported HTLV-I isolates and simian T-lymphotropic virus type I (STLV-I) isolates. The heterogeneity of HTLV-I was recently recognized, and one major type, generally called the "cosmopolitan" type, contained Japanese, Caribbean, and West African isolates. The phylogenetic tree constructed in the present study has shown that this cosmopolitan type can be further grouped into three lineages (subtypes A, B, and C). Subtype A consists of some Caribbean, two South American, and some Japanese isolates, including that from the Ainu, in addition to an Indian isolate, and subtype B consists of other Japanese isolates in addition to another Indian isolate, suggesting that there might be at least two ancestral lineages of the Japanese HTLV-I. Subtype A implies a close connection of the Caribbean and South American natives with the Japanese and thereby a possible migration of the lineage to the American continent via Beringia in the Paleolithic era. Subtype C consists of the West African and other Caribbean isolates, indicating that not all but part of the Caribbean strains directly originated from West Africa probably during the period of slave trade. The tree also has shown that the HTLV-I isolate from Gabon in central Africa forms a cluster with STLV-I from a chimpanzee, suggesting a possible interspecies transmission between man and the chimpanzee in the past. No specific clustering was observed in the tree in relation to manifestations of the disease such as adult T-cell leukemia and HTLV-I-related neurological disorders. Thus, the topology of the phylogenetic tree reflects the movement of people carrying the virus in the past.**

Human T-lymphotropic virus type I (HTLV-I) (1, 2), the retrovirus causing adult T-cell leukemia (ATL) (3) and HTLV-I-associated myelopathy or tropical spastic paraparesis (HAM/TSP) (4), was considered to be mainly endemic in Japanese in East Asia and Blacks in Africa (5) and the Caribbean basin (6). Recently, foci of HTLV-I carriers were found in other various regions in the world. From genetic analysis, three major types of HTLV-I are known: the first is the most divergent type, recently isolated from remote Melanesians in Papua New Guinea and the Solomon Islands in the South Pacific Ocean and from Australian aborigines (7–9); the second is a distinct type isolated from a Zairian (10); the third is the so-called "cosmopolitan" type, which includes the majority of isolates from various areas (1–6). HTLV-I is known to be transmitted only by cell-associated infection and mainly by vertical infection from mother to child; consequently, its horizontal spread throughout a population is unlikely. Thus, the distribution of the virus is considered to be related to the anthropological background and past human movement. So we examined the phylogenetic relationship of isolates of HTLV-I and simian T-lymphotropic virus type I (STLV-I), which is closely related to HTLV-I (11), from various areas of the world and report on the phylogenetic characterization of the isolates from the Ainu in Japan and from native inhabitants in India; Colombia and Chile in South America; Gabon in central Africa; and Ghana in West Africa. The Ainu in particular are regarded as relatively pure descendants of the native population inhabiting mainly northern Japan from the preagricultural "Jomon" period more than 2300 years ago.

We selected the long terminal repeat (LTR) region for phylogenetic analysis of the viruses for the following reasons: (*i*) the HTLV-I/STLV-I genome is highly conserved, yet its LTR region is one of the most variable regions; (*ii*) positive selection by environmental pressure such as the immune system affects the LTR less than the *env*, which is another variable region; and (*iii*) many sequence data on this region have been reported previously and are available.

Table 1 shows the isolates of HTLV-I/STLV-I used in the present study. Using the PCR method (22) or the molecular cloning of provirus genomes, we newly sequenced seven strains that had been isolated from a Ghanaian in West Africa, a Gabonaise in central Africa (20), a Chilean (21) and a Colombian in South America, two Indians (KK and SG) in India, and an Ainu in North Japan. The Colombian strain (SIB170) was isolated from an indigenous 51-year-old female inhabitant in the Andes area in 1991. Both Indian strains were isolated in 1991 from HAM/TSP patients in Bombay. KK was a 52-year-old Hindu female, and SG was a 32-year-old male born in Kerala in South India. The Ainu female in Hokkaido island in northern Japan, from whom the virus was isolated in 1983, was 84 years old. The sequences analyzed

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HTLV-I and HTLV-II, human T-lymphotropic virus types I and II; STLV-I, simian T-lymphotropic virus type I; LTR, long terminal repeat; ATL, adult T-cell leukemia; HAM/TSP, HTLV-I-associated myelopathy or tropical spastic paraparesis; N-J, neighbor joining; UPG, unweighted pair group.

[k]To whom reprint requests should be addressed.

Table 1. Isolates of HTLV-I/STLV-I used in the present study

Isolate	Location	Pathology	Ref.
HTLV-I			
ATM	Japan	ATL	12
ATK	Japan	ATL	13
CR1	Caribbean	ATL	14
HS35	Caribbean	ATL	15
1010	Caribbean	ATL	15
H5	Japan	HAM/TSP	16
TSP1	Caribbean	HAM/TSP	17
HCT	Japan	HAM/TSP	18
XAV	Caribbean	HAM/TSP	19
BOU	Caribbean	HAM/TSP	19
GRO	Caribbean	B lymphoma	19
SIE	Ivory Coast	ATL	19
AKR	Ivory Coast	HAM/TSP	19
JAP1	Japan	ATL	19
JAP2	Japan	ATL	19
JAP3	Japan	HAM/TSP	19
JAP4	Japan	HAM/TSP	19
CH	Caribbean	ND	10
EL	Zaire	ND	10
PNG1	Papua New Guinea	ND	7
MEL5	Solomon Islands	ND	8
GB233	Gabon	ND	20
GH78	Ghana	ND	20
CH26	Chile	ND	21
SIB170	Colombia	ND	UD
AINU	Japan	ND	UD
KK	India	HAM/TSP	41
SG	India	HAM/TSP	41
STLV-I*			
AGM	Kenya	ND	1
CPZ	Sierra Leone	ND	11
PTM	Asia	ND	11

ND, no disease or disease other than that caused by HTLV-I; UD, unpublished data; *STLV-I isolates: AGM, African green monkey; CPZ, chimpanzee; PTM, pig-tailed macaque.

were about 500 base pairs long, which is the maximum length of the common region of all the references used (7–21, 23), corresponding to positions 144–650 in ATK (an ATL provirus clone) (13).

From the LTR sequence data, the total number of nucleotide substitutions was estimated for each pair of isolates by the six-parameter method (24). Using these numbers, phylogenetic trees were constructed by two different methods, the neighbor joining (N–J) (25) and the unweighted pair group (UPG) methods (26). Both of these methods gave similar trees. Because two methods based on different principles provided similar topologies, the result is considered reliable (27). Furthermore, bootstrap estimation of phylogenetic variability using 2000 replications was performed on the tree (28).

Fig. 1 shows the phylogenetic tree constructed by the N–J method. A long time after the divergence of human T-lymphotropic virus type II (HTLV-II) (23), the branches of Melanesian HTLV-I and STLV-I isolated from an Asian pig-tailed macaque (11) diverged in order. After STLV-I from an African green monkey (11) diverged, central African isolates from Zaire and Gabon form a cluster with STLV-I from a chimpanzee. This suggests a likelihood of interspecies transmission between man and the chimpanzee in the relatively near past. All the other HTLV-I isolates are the cosmopolitan type of HTLV-I. However, the tree clearly shows the existence of three lineages (subtypes A, B, and C) of HTLV-I in the cosmopolitan type. In bootstrap estimation, subtypes A and C were quite stable (more than 80%) while subtype B was not so stable (about 40%). Subtype A consists of Caribbean, South American (Chilean and Colombian) and

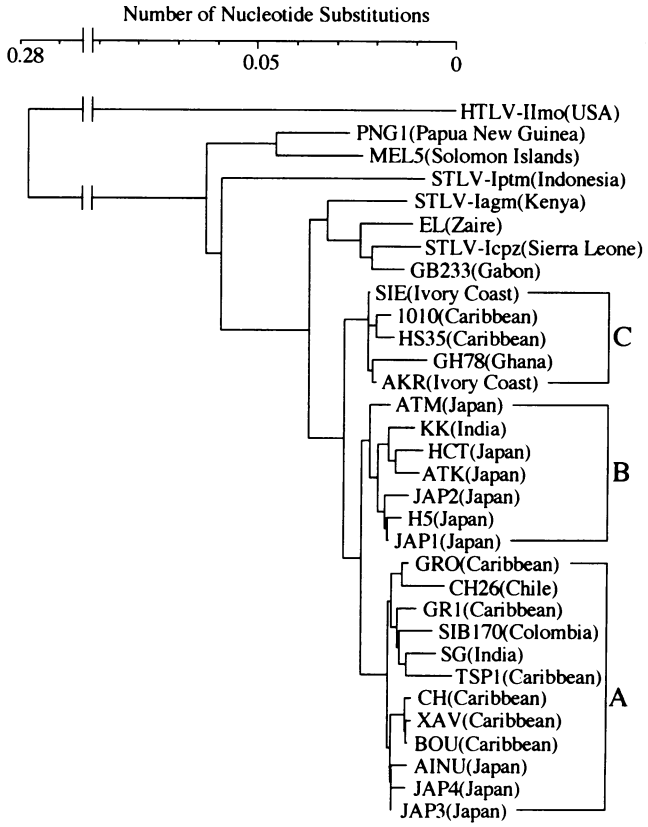

FIG. 1. Phylogenetic tree showing the evolutionary relationship of HTLV-I/STLV-I in the world, including recently sequenced isolates from native Indian, Colombian, and Chilean individuals and from the Ainu in Japan. The tree was constructed by the N–J method (25) after alignment of the nucleotide sequence of the LTR region [144–650 in ATK (13)] using an ODEN program package for a computer at the National Institute of Genetics in Japan. A scale on the tree is the number of nucleotide substitutions per site estimated by the six-parameter method (24), and the horizontal branch length indicates the genetic distance.

Japanese isolates, including that from the Ainu together with one Indian isolate, whereas subtype B consists of other Japanese isolates and another Indian isolate. In other words, there are two (or more) lineages of Japanese HTLV-I because subtype B was not so stable in bootstrap estimation. Both Japanese lineages (subtypes A and B) are related to the Indian isolates, but only one of them—that is, stable subtype A—is closely related to Caribbean and South American isolates. This is not consistent with the general idea that all Caribbean HTLV-I was directly transmitted from West Africa during the slave trade period. However, its transmission to the Caribbean from West Africa is supported by the third lineage, stable subtype C, which consists of West African and Caribbean isolates. Thus, our analysis indicates the existence of mixed populations with at least two lineages in Japan (subtypes A and B) and two in the Caribbean basin (subtypes A and C) and one stable common lineage (subtype A) of Caribbean, South American, and Japanese HTLV-I including the Ainu isolate. The samples analyzed include isolates from ATL and HAM/TSP patients and healthy carriers, but this tree showed no specific clustering linked to the pathogenicity.

Previously, we proposed that interspecies transmission of HTLV-I/STLV-I between humans and nonhuman primates was unlikely (29) and that it evolved in concert with the host species after the ancestral virus infected the ancestor of primates, resulting in the present species-specific viruses. However, rather recent divergence of STLV-I from a chim-

panzee and HTLV-I isolated in central Africa shown by the tree is against this idea. In addition, the phylogenetic tree of this virus does not coincide with that of the host primate species, as pointed out by recent reports (30, 31). The viruses clearly started to diverge later than primates.

Fig. 2 shows the geographical distribution of HTLV-I in the world and a possible movement of the cosmopolitan type in the past based on the topology of the phylogenetical tree and the anthropological background. We think the present distribution of the cosmopolitan type can be explained by human movements in the past. Namely, the West African lineage was brought to the Caribbean basin by the slave trade several hundred years ago, forming the subtype C. This possibility is also supported by recent analysis of the *env* region of HTLV-I by Gessain *et al.* (32). The other lineage possibly moved to Japan as more than two Paleo-Mongoloid lineages in the Paleolithic period, resulting in the present existence of at least two lineages of HTLV-I in Japan (subtypes A and B).

In Japan, HTLV-I is mainly endemic in the southwest and in remote populations of frontier areas such as "Ryukyuans" on Okinawa and Ainu people on Hokkaido island. Ishida and Hinuma proposed that "Jomon" people, ancient inhabitants of Japan who carried HTLV-I, were shoved away by the HTLV-I-free Yayoi people, who invaded Japan from the Asian continent more than 2300 years ago (33). The first immigration to the Japanese archipelago is considered to have occurred more than 10,000 years ago (34), and the mixed HTLV-I population of subtypes A and B in Japan revealed by the tree may indicate plural immigrations of HTLV-I-carrying people before or during the "Jomon" period.

From the paleoanthropological viewpoint, it is suggested that the first immigrants to the North American continent arrived about 12,000 years ago from North Asia, passing through Beringia and then gradually spreading from the North American continent to the South American continent (35). Phylogenetic analysis of mitochondrial DNA (36) and HLA type analysis (37) also suggest the relationship between Japanese and Paleo-Indians in South America. Only one of the two lineages in Japan, subtype A, may have crossed Beringia; the other lineage, subtype B, probably reached Japan after Beringia was disconnected.

Caribbean HTLV-I has generally been considered to have been brought from West Africa during the period of slave trade. This idea is also supported by the existence of subtype C in our analysis. Our study, however, indicates the existence of another lineage, subtype A, related to Asian HTLV-I, in the Caribbean basin. This suggests a possibility of mixed blood between Asian and African populations in the Caribbean.

A highly divergent type of HTLV-I in remote Melanesian in Papua New Guinea and the Solomon Islands of Oceania and Australian aborigines was recently reported (7–9), and it was proposed that HTLV-I originated in the Indo-Malay region rather than Africa. The topology of our tree is slightly different from one by them, though the same LTR region was analyzed. This is due to the difference in the methods used to estimate nucleotide substitutions. They used simple homology percent for the analysis, which does not take overlapping substitutions into consideration, resulting in a relatively higher calculation error. We used the six-parameter method that is considered to be more reasonable. The reliability of our analysis was supported by the similar results obtained by the UPG and N–J methods, and the results were further evaluated by bootstrap estimation.

Saksena *et al.* proposed that settlers from the Indonesian archipelago spread HTLV-I to Africa thousands of years ago and that European adventurers and slave traders disseminated the infection to the New World and Japan. However, this proposition is unlikely for the following reasons. (*i*) Our present data and serological data described previously (33) do not support their proposition about the origin of Japanese HTLV-I. (*ii*) Their proposition does not explain why most African nonhuman primates naturally carry HTLV-I-related virus (38–40). (*iii*) Their idea of the Indo-Malay origin of the virus is not supported by results on the Indian isolates analyzed in this study. There is also a possibility that the Melanesian lineage was brought to the Indo-Malay region a very long time ago, such as in the period of *Homo erectus* (hundreds of thousands of years before), from Africa, which is considered to be the birthplace of human beings. We think that it is premature to draw a definite conclusion on the origin of HTLV-I at this moment. The connection between Melanesian and African HTLV-I should be further investigated by analyzing isolates from the region of the Pan-Indian-Ocean, such as Seychelles islanders and minor populations of the Indian continent and the Malay Peninsula, and STLV-I isolates from various species of Asian and African nonhuman primates. Now, new light is being shed on the question of the origin and evolution of HTLV-I. These studies may also

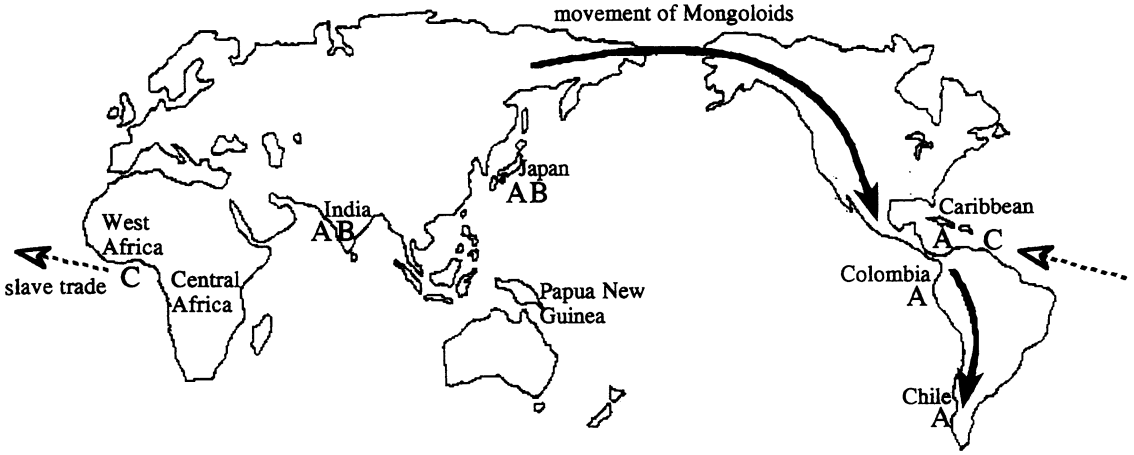

FIG. 2. Possible movement of the cosmopolitan type of HTLV-I in the world, deduced from the topology of the phylogenetic tree and anthropological background. The distribution of the cosmopolitan type of HTLV-I can be explained by the movement of human beings. The West African lineage was brought to the Caribbean basin by the slave trade several hundred years ago, forming subtype C. Other lineages possibly moved to Japan as two (or more) lineages with the movement of mongoloid people in the paleolithic period, resulting in the existence of two (or more) lineages of HTLV-I in Japan, subtypes A and B. Only subtype A crossed Beringia and reached the Caribbean basin and South America. The other Japanese lineage, subtype B, possibly reached Japan after Beringia was disconnected.

provide new information on the evolutional history of our ancestors.

We thank Dr. K. Katayama, Kyoto University, for fruitful discussion of anthropological aspects. This work was supported by Grants-in-Aid for AIDS Research and Cancer Research from the Ministry of Health and Welfare and Ministry of Education, Science and Culture of Japan.

1. Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D. & Gallo, R. C. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7415–7419.
2. Hinuma, Y., Nagata, K., Hanaoka, M., Nakai, M., Matsumoto, T., Kinoshita, K., Shirakawa, S. & Miyoshi, I. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6476–6480.
3. Uchiyama, T., Yodoi, J., Sagawa, K., Takatsuki, K. & Uchino, H. (1977) *Blood* **50**, 481–492.
4. Osame, M., Usuku, K., Izumo, S., Ijichi, N., Amitani, H., Igata, A., Matsumoto, M. & Tara, M. (1986) *Lancet* **i**, 1031–1032.
5. Biggar, R. J., Saxinger, C., Gardiner, C., Collins, W. E., Levine, P. H., Clark, J. W., Nkruman, F. K. & Blattner, W. A. (1984) *Int. J. Cancer* **34**, 215–219.
6. Blattner, W. A., Kalyanaraman, V. S., Robert-Guroff, M., Lister, T. A., Galton, D. A. G., Sarin, P. S., Crawford, M. H., Catovsky, D., Greaves, M. & Gallo, R. C. (1982) *Int. J. Cancer* **30**, 257–264.
7. Saksena, N. K., Sherman, M. P., Yanagihara, R., Dube, D. K. & Poiesz, B. J. (1992) *Virology* **189**, 1–9.
8. Gessain, A., Boeri, E., Yanagihara, R., Gollo, R. C. & Franchini, G. (1993) *J. Virol.* **67**, 1015–1023.
9. Bastian, I., Gardner, J., Webb, D. & Gardner, I. (1993) *J. Virol.* **67**, 843–851.
10. Ratner, L., Philpott, T. & Trowbridge, D. B. (1991) *AIDS Res. Hum. Retroviruses* **7**, 923–941.
11. Watanabe, T., Seiki, M., Hirayama, Y. & Yoshida, M. (1986) *Virology* **148**, 385–388.
12. Seiki, M., Hattori, S. & Yoshida, M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6899–6902.
13. Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3618–3622.
14. Josephs, S. F., Wong-Staal, F., Manzari, V., Gallo, R. C., Sodroski, J. G., Trus, M. D., Perkins, D., Patarca, R. & Haseltine, W. A. (1984) *Virology* **139**, 340–345.
15. Malik, K. T. A., Even, J. & Karpas, A. (1988) *J. Gen. Virol.* **69**, 1695–1710.
16. Tsujimoto, A., Teruuchi, T., Imamura, J., Shimotohno, K., Miyoshi, I. & Miwa, M. (1988) *Mol. Biol. Med.* **5**, 29–42.
17. Evangelista, A., Maroushek, S., Minnigan, H., Larson, A., Retzel, E., Haase, A., Gonzallezdunia, D., McFarlin, D., Mingioli, E., Jacobson, S., Osame, M. & Sonoda, S. (1990) *Microb. Pathogen.* **8**, 259–278.
18. Shirabe, S., Nakamura, T., Tsujihata, M., Nagataki, S., Seiki, M. & Yoshida, M. (1990) *Arch. Neurol.* **47**, 1258–1260.
19. Komurian, F., Pelloquin, F. & De The, G. (1991) *J. Virol.* **65**, 3770–3778.
20. Fukasawa, M., Tsujimoto, H., Ishikawa, K., Miura, T., Ivanoff, B., Cooper, R. W., Frost, E., Delaporte, E., Mingle, J. A. A., Grant, F. C. & Hayami, M. (1987) *Virology* **161**, 315–320.
21. Cartier, L., Araya, F., Castillo, J. L., Zaninovic, V., Hayami, M., Miura, T., Imai, J., Sonoda, S., Shiraki, H., Miyamoto, K. & Tajima, K. (1993) *Jpn. J. Cancer Res.* **84**, 1–3.
22. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491.
23. Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D. W., Chen, I. S. Y., Miwa, M. & Sugimura, T. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3101–3105.
24. Gojobori, T., Ishii, K. & Nei, M. (1982) *J. Mol. Evol.* **18**, 414–423.
25. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
26. Nei, M. (1975) *Molecular Population Genetics and Evolution* (Columbia Univ. Press, Irvington-on-Hudson, NY).
27. Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, Irvington-on-Hudson, NY).
28. Felsenstein, J. (1985) *Evolution* **39**, 783–791.
29. Komuro, A., Watanabe, T., Miyoshi, I., Hayami, M., Tsujimoto, H., Seiki, M. & Yoshida, M. (1984) *Virology* **138**, 373–378.
30. Ina, Y. & Gojobori, T. (1990) *J. Mol. Evol.* **31**, 493–499.
31. Saksena, N. K., Hervé, V., Sherman, M. P., Durand, J. P., Mathiot, C., Müller, M., Love, J. L., LeGuenno, B., Sinoussi, F. B., Dube, D. K. & Poiesz, B. J. (1993) *Virology* **192**, 312–320.
32. Gessain, A., Gallo, R. C. & Franchini, G. (1992) *J. Virol.* **66**, 2288–2295.
33. Ishida, T. & Hinuma, Y. (1986) *Nature (London)* **322**, 504.
34. Suzuki, H. (1982) *J. Anthrop. Soc. Nippon* **90**, 11–26.
35. Bryan, A. L. (1983) in *Early Man in the New World*, ed. Shulter, R. (Saga Publications, Beverly Hills, CA), pp. 137–146.
36. Horai, S., Kondo, R., Murayama, K., Hayashi, S., Koike, H. & Nakai, N. (1991) *Philos. Trans. R. Soc. London B* **333**, 409–417.
37. Sonoda, S., Arce-Gomez, B., Satz, M. L., Gorodezky, C., Juarez, V. *et al.* (1992) in *HLA 1991*, eds. Tsuji, K., Aizawa, M. & Sasazuki, T. (Oxford Univ. Press, Oxford), pp. 685–688.
38. Hayami, M., Komuro, A., Nozawa, K., Shotake, T., Ishikawa, K., Yamamoto, K., Ishida, T., Honjo, S. & Hinuma, Y. (1984) *Int. J. Cancer* **33**, 179–183.
39. Tsujimoto, H., Komuro, A., Iijima, K., Miyamoto, J., Ishikawa, K. & Hayami, M. (1985) *Int. J. Cancer* **35**, 377–384.
40. Ishikawa, K., Fukasawa, M., Tsujimoto, H., Else, J. G., Isahakia, M., Ubhi, N. K., Ishida, T., Takenaka, O., Kawamoto, Y., Shotake, T., Ohsawa, H., Ivanoff, B., Cooper, R. W., Frost, E., Grant, F. C., Spriatna, Y., Sutarman, Abe, K., Yamamoto, K. & Hayami, M. (1987) *Int. J. Cancer* **40**, 233–239.
41. Hashimoto, K., Lalkaka, J., Fujisawa, J., Singhal, B. S., Machigashira, K., Kubota, R., Suehara, M., Osame, M. & Yoshida, M. (1993) *AIDS Res. Hum. Retroviruses* **9**, 495–498.