

The Low Evolutionary Rate of Human T-Cell Lymphotropic Virus Type-1 Confirmed by Analysis of Vertical Transmission Chains

S. Van Dooren,* O. G. Pybus,† M. Salemi,* H.-F. Liu,‡ P. Goubau,§ C. Remondégui,|| A. Talarmin,¶ E. Gotuzzo,# L. C. J. Alcantara,** B. Galvão-Castro,** and A.-M. Vandamme*

*Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven, Belgium; †Department of Zoology, University of Oxford, Oxford, U.K.; ‡Department of Medical Research, Mackay Memorial Hospital, Taipei, Taiwan; §Université Catholique de Louvain, Unité de Virologie, Bruxelles, Belgium; ||Hospital San Roque, San Salvador de Jujuy, Argentina; ¶Laboratoire de Rétrovirologie, Institut Pasteur de la Guyane, Cayenne, French Guiana; #Instituto de Medicina Tropical 'Alexander Von Humboldt', Universidad Peruana Cayetano Heredia, Lima, Peru; **Gonçalo Moniz Research Center, Oswaldo Cruz Foundation, Bahia School of Medicine and Public Health, Salvador, Bahia, Brazil

The evolutionary rate of the human T-cell lymphotropic virus type-1 (HTLV-1) is considered to be very low, in strong contrast to the related human retrovirus HIV. However, current estimates of the HTLV-1 rate rely on the anthropological calibration of phylogenies using assumed dates of human migration events. To obtain an independent rate estimate, we analyzed two variable regions of the HTLV-1 genome (LTR and env) from eight infected families. Remarkable genetic stability was observed, as only two mutations in LTR (756 bp) and three mutations in env (522 bp) occurred within the 16 vertical transmission chains, including one ambiguous position in each region. The evolutionary rate in HTLV-1 was then calculated using a maximum-likelihood approach that used the highest and lowest possible times of HTLV-1 shared ancestry, given the known transmission histories. The rates for the LTR and env regions were 9.58×10^{-8} – 1.25×10^{-5} and 7.84×10^{-7} – 2.33×10^{-5} nucleotide substitutions per site per year, respectively. A more precise estimate was obtained for the combined LTR-env data set, which was 7.06×10^{-7} – 1.38×10^{-5} substitutions per site per year. We also note an interesting correlation between the occurrence of mutations in HTLV-1 and the age of the individual infected.

Introduction

The two types of human T-cell lymphotropic virus (HTLV), type 1 (HTLV-1) and type 2 (HTLV-2), belong to the family Retroviridae, together with the human immunodeficiency viruses (HIV-1 and HIV-2). Although HTLV and HIV have a comparable morphology, life cycle, and genetic structure, their replication strategies seem to be quite different, resulting in different evolutionary behaviors. In particular, the HTLV genome is remarkably stable compared with that of HIV. Additionally, HTLV is not as transmissible as HIV, even though both viruses are blood borne and use similar transmission routes: (1) vertical, from mother-to-infant (especially via breastfeeding, for HTLV) (Wiktor et al. 1997; Bittencourt 1998; Ando et al. 2003), (2) horizontal, via sexual contact (Murphy et al. 1989; Kaplan et al. 1996; Plancoulaine et al. 1998), and (3) parenteral, via injection equipment. (Lee et al. 1989; Hjelle et al. 1990; Sullivan et al. 1991; Manns et al. 1992; Take et al. 1993). HTLV is mainly infectious through cell-bound virus and not through freely circulating virus. Initially, a short period of active viral replication is needed to establish infection of the lymphocytes. After transformation and immortalization of the lymphocytes, HTLV replication is largely maintained through clonal expansion of the infected cells (Wattel et al. 1995, 1996), rather than by the replication and propagation of new virions by reverse transcription of integrated proviral

DNA. Thus, the sequence divergence observed among HTLV strains results from mutations incorporated during active viral replication and from somatic mutations acquired by the provirus during cellular mitosis.

The genetic stability of HTLV has been indicated by several estimates of the evolutionary rate of the virus. The investigation of intrafamilial HTLV infections has demonstrated the presence of almost identical HTLV-1 sequences in several family members sampled over several generations (Nerurkar et al. 1993; Liu et al. 1994). In a more precise manner, the evolutionary rate of HTLV-1, measured as the number of nucleotide substitutions per site per year, has been estimated by phylogenetic analysis using two different methods. The conventional approach, which involves dividing the difference in branch lengths in a phylogenetic tree by the difference in isolation time (Li, Tanimura, and Sharp 1988), although straightforward, is not ideally suited for HTLV strains, as the observed sequence divergence is usually too small for reliable estimates to be obtained. This method has only been applied to HTLV-2 strains from intravenous drug users (Salemi et al. 1999). The second approach assumes a known time for a particular node in the HTLV phylogeny. All current calculations of the HTLV-1 evolutionary rate are based on an assumed time point in the migration history of the human host population (Yanagihara et al. 1995; Salemi et al. 1998; Van Dooren et al. 1998; Salemi, Desmyter, and Vandamme 2000; Van Dooren, Salemi, and Vandamme 2001). This time point is obtained from anthropological studies, so the estimated rates are dependent on the accuracy of the anthropological date and rely heavily on the assumption that the phylogenetic node in question coincides with the anthropological event.

Herein, we report an alternative approach that uses HTLV-1 familial transmission data that is independent of anthropological dates. Familial HTLV-1 is predominantly

Key words: HTLV-1, vertical transmission, evolutionary rate, molecular clock.

E-mail: Sonia.VanDooren@uz.kuleuven.ac.be.

Mol. Biol. Evol. 21(3):603–611. 2004

DOI: 10.1093/molbev/msh053

Advance Access publication January 22, 2004

Molecular Biology and Evolution vol. 21 no. 3

© Society for Molecular Biology and Evolution 2004; all rights reserved.

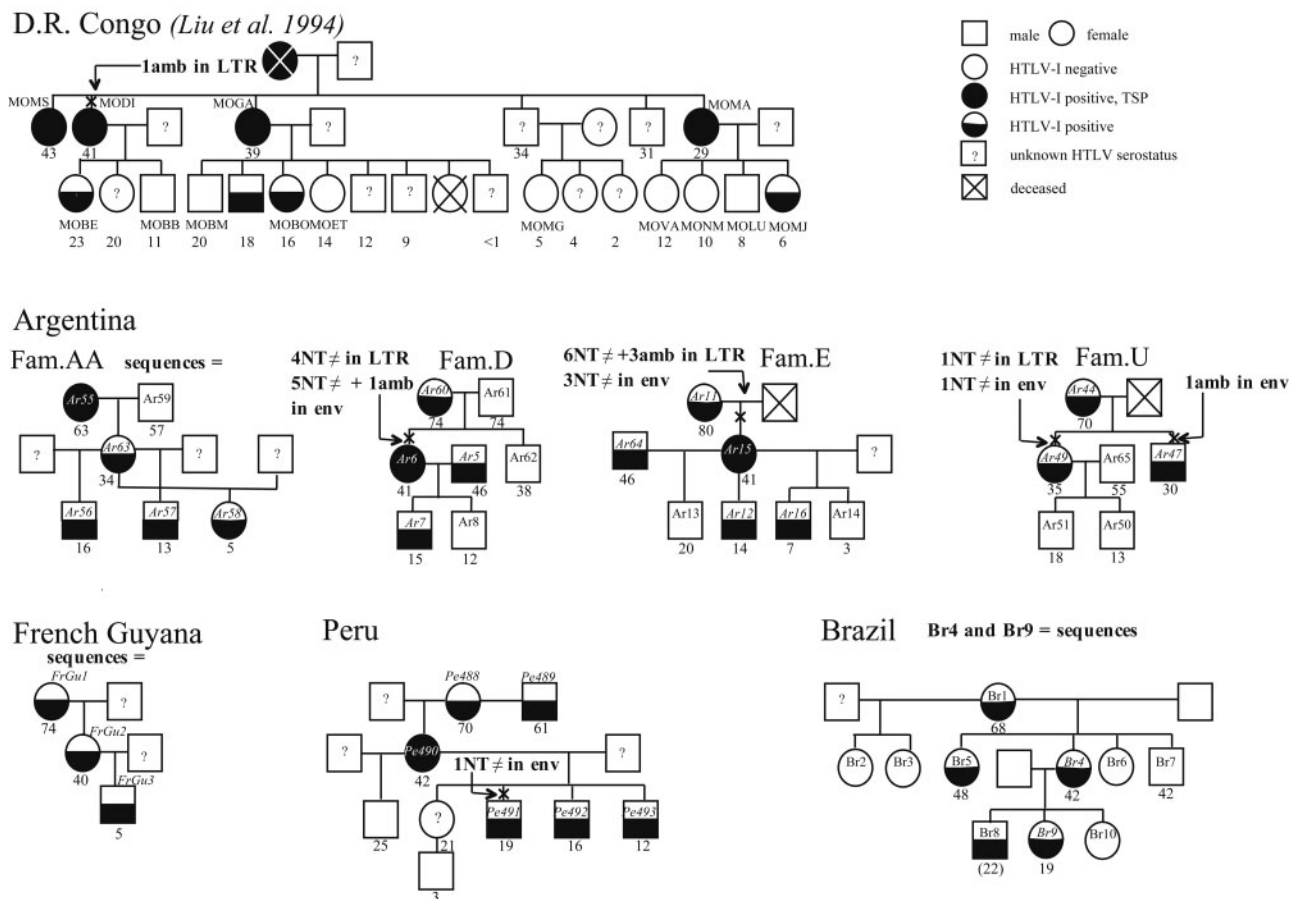


FIG. 1.—The pedigrees of the eight HTLV-1 infected families. The age of the infected patient is noted under the sample name. LTR and env sequences were obtained from samples with names in italic. The observed mutations are indicated by the symbol \times on the branches of the pedigree. The genomic locations of the mutations are noted. “NT” denotes an unambiguous mutation, “amb” denotes an ambiguous position.

transmitted vertically from mother to offspring or horizontally through sexual contact. However, only HTLV-1 sequences from vertically infected family members can contribute to the estimation of evolutionary rates, as data on the timing of horizontal transmissions are almost impossible to collect, forcing us to exclude these events from our calculations.

To obtain an estimate of the molecular evolutionary rate of HTLV-1 that is independent of anthropological data, we have sequenced the complete LTR region and a 522 bp fragment of the gp21 *env* region from HTLV-1 strains in eight families. First, the phylogenetic relationships of the strains were studied. Subsequently, we estimated the evolutionary rate of the sequenced genome regions from the total number of mutations accumulated during the combined period of vertical infection that was represented by the pedigrees of the infected families.

Materials and Methods

Description of the Pedigrees of HTLV-1 Infected Families

The pedigrees of the eight families are presented in figure 1. The Congolese family “MO,” previously published by Liu et al. (1994), has seven HTLV-1b-infected members spanning two generations, in which all

first-generation members (four sisters) have tropical spastic paraparesis (TSP). Four Argentinean families (all with three generations) of mixed, but predominantly Amerindian, origin were investigated. Each Argentinean family had three to five HTLV-1 infected members. Three of the Argentinean families harbored TSP cases in the first or second generations. A three-generation Noir-Marron family from French Guyana, containing some members who were infected with strongyloides, was also studied. Two other three-generation families were included: one from Peru of mixed origin (mestizo) and one from Brazil of black origin. Both families have five to six HTLV-1-infected family members, with a TSP case in the second generation of the Peruvian family. Blood samples from all members of the same family were collected in the same year.

DNA Isolation, PCR, and Sequencing

For all non-Argentinean families, DNA was extracted from lymphocytes isolated from blood that was collected on EDTA tubes or Vacutainer Cell Preparation tubes (Becton Dickinson). For the Argentinean families, DNA was extracted from dried whole-blood spots collected on filter paper. DNA was extracted using the classical proteinase K digestion and phenol-chloroform extraction,

followed by ethanol precipitation or using the QiaAmp Mini Blood kit (Qiagen). The LTR and env regions were PCR amplified using previously described primers and PCR conditions (Van Dooren et al. 1998). The PCR products were agarose gel purified using the Qiaquick Gel extraction kit (Qiagen). Population sequencing was done by direct sequencing of the purified PCR products using the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Version 3.0 mixture (Applied Biosystems), and sequences were generated on an ABI Prism 310 according to the manufacturers' instructions (Applied Biosystems). Sequence assembly was done using the ABI Sequence Navigator software (Applied Biosystems).

The accession numbers of the new Latin American sequences are AY324777 to AY324788 (LTR) and AY324789 to AY324800 (env). For the African family members, the accession numbers are Z31659 (MOM LTR) and X88884 (MOMS env).

Phylogeny

Only the phylogeny of the LTR region was studied, as closely related strains are best investigated using a highly variable genome region. A Blast search (www.ncbi.nlm.gov/blast) was conducted on the Latin American isolates as an aid to HTLV-1 subtype identification. A detailed phylogenetic analysis was then performed. ClustalX (Jeanmougin et al. 1998) was used to align the nine nonidentical familial sequences with 59 HTLV-1 reference sequences from GenBank that belonged to the same subtype. Reference strains from the same geographic area were preferentially chosen, and three strains from other subtypes were included as outgroups. Minor editing of the alignment was performed manually in MacClade version 3.04 (Maddison and Maddison 1992). Phylogenetic trees were estimated with PAUP* version 4.0b10 (Swofford 1998). The Tamura-Nei substitution model with gamma-distributed rate heterogeneity among sites was chosen as best model for PTLV-1 strains (Salemi, Desmyter, and Vandamme 2000) and was thus used to construct neighbor-joining (NJ) and maximum-likelihood (ML) trees. Using empirical base frequencies, the NJ tree was constructed by optimizing the substitution rate matrix and gamma shape parameter three times; this was followed by a bootstrap analysis (1,000 replicates). The ML tree was estimated using the substitution model and parameters obtained above. A heuristic ML search was performed using the subtree-pruning-regrafting branch-swapping algorithm and an NJ starting tree. Statistical support for the ML tree branches was calculated in PAUP* using a likelihood ratio test that compared the likelihood of the estimated branch length with that of a zero branch length.

HTLV-1 Evolutionary Rate Estimation

The sequences from each family were aligned and the position and number of observed mutations in the LTR and env regions were scored (see fig. 1).

The HTLV evolutionary rate was calculated using a homogenous Poisson process model. Taken together, the pedigrees represent a total amount of time (t years), during

which n mutations were observed in the combined LTR-env region. Therefore, the likelihood of the evolutionary rate (λ) in this region is

$$L(\lambda | t, n) = \frac{e^{-\lambda t}}{n!} (\lambda t)^n. \quad (1)$$

The pedigree data can be used to calculate the highest (y) and lowest (x) possible values of t . For each pedigree, the lowest possible value of t is obtained by summing the ages of the different offspring at the time of sampling, and then adding the age of the oldest child to account for evolution that occurred within the mother. This procedure assumes a maximum amount of shared ancestry. Summing this value across all pedigrees gives a lowest possible value of $x = 419$ years. For the highest possible t , the age of the mother at the time of sampling was summed for each offspring and also for the mother. This procedure assumes no shared ancestry within the pedigree of each family. When summed across all pedigrees, this yields a maximum value of $y = 1,101$ years. Therefore, the pedigrees represent a total of 419 to 1,101 years of HTLV-1 evolution. When this range is incorporated uniformly into the Poisson model we obtain the marginal likelihood

$$L(\lambda | n) = \int_{t=x}^y \frac{e^{-\lambda t}}{n!} \frac{(\lambda t)^n}{(y-x)} dt \quad (2)$$

which has the solution

$$L(\lambda | n) = \frac{1}{\lambda(y-x)} \times \sum_{i=0}^n \left\{ \frac{1}{i!} (e^{-\lambda x} (\lambda x)^i - e^{-\lambda y} (\lambda y)^i) \right\}. \quad (3)$$

Equation 3 was used to obtain maximum-likelihood estimates of the evolutionary rate λ . Because two of the proposed positions are ambiguous, separate estimates of λ were obtained for $n = 3$, $n = 4$, and $n = 5$. Approximate 95% confidence intervals were obtained using the likelihood ratio statistic. The value λ represents the number of mutations per year in the investigated LTR and/or env region. Thus, the standard rate of molecular evolution (mutations per nucleotide site per year) is obtained by dividing λ by the length of the investigated region (1,278 nucleotides for LTR-env).

Results

PCR and Sequencing

PCR amplification of HTLV DNA was successful for all blood samples (indicated in italic in fig. 1), except for a few members of the Brazilian family (Br1, Br5, and Br8). Different combinations of primers and extra inner PCR cycles were tried out, and alternative primer sets for LTR (Alcantara et al. 2003) and env (Yang et al. 1997) were used, resulting in negative or very faint PCR bands. These three samples were therefore not included in the study.

Phylogeny

Figure 2 shows the results of the phylogenetic analysis. Both NJ and ML trees showed similar topologies

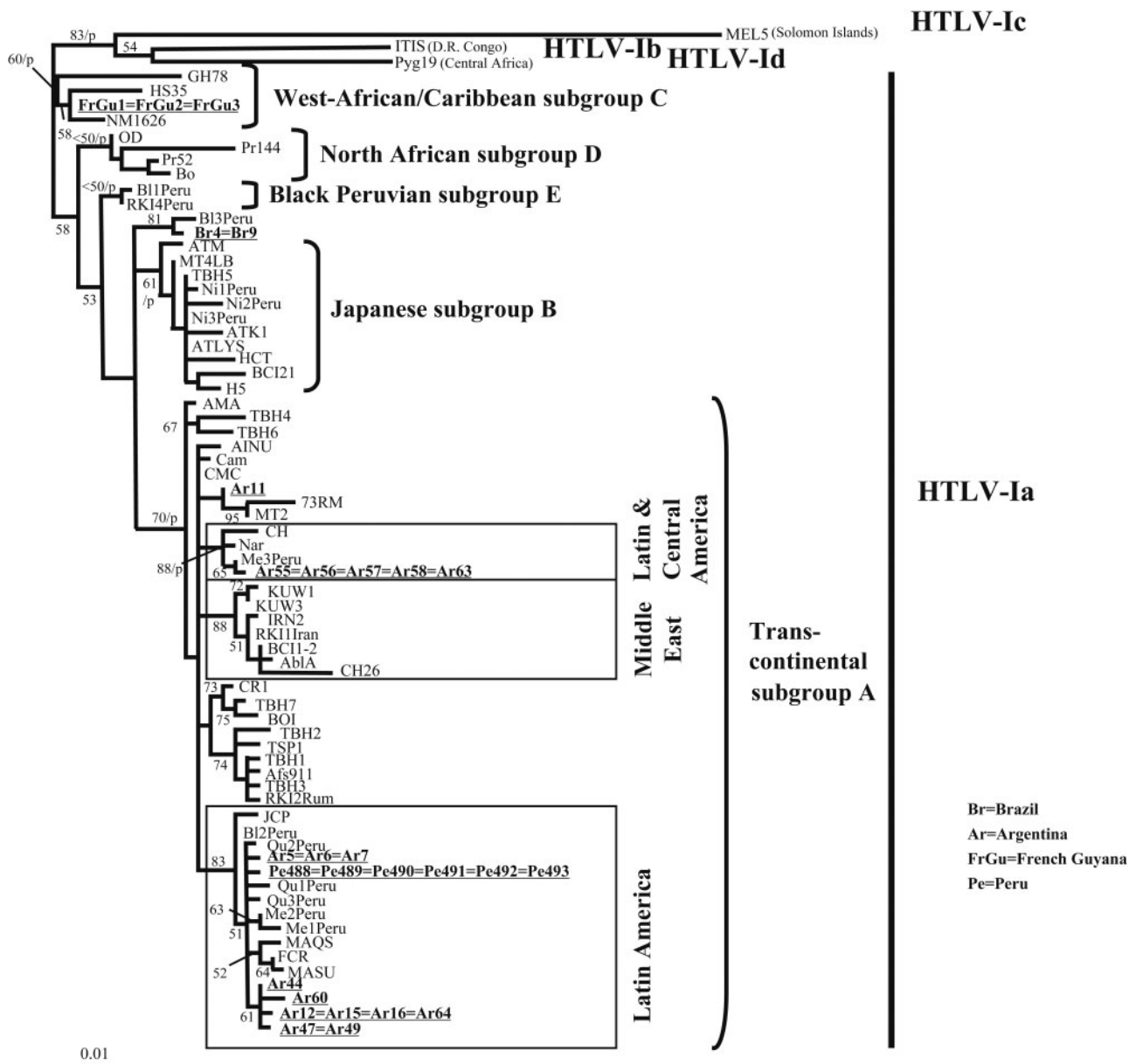


Fig. 2.—ML tree of a 722-bp LTR fragment containing 68 HTLV-1a strains, including the strains reported here, rooted with three HTLV-1 strains from other subtypes (HTLV-1b, HTLV-1c, and HTLV-1d). The phylogenetic tree was inferred under the Tamura-Nei evolutionary model with a gamma model of among-site substitution rate heterogeneity. The NJ bootstrap and maximum-likelihood statistic P -values (denoted by 'p' if $P < 0.01$).

and demonstrated that HTLV-1b, HTLV-1c, and HTLV-1d were appropriate outgroups (NJ bootstrap, 83%; ML, $P < 0.01$). Within the HTLV-1a part of the tree, different subgroups could be identified (as previously noted by Yamashita et al. [1996] and Van Dooren et al. [1998]). These subgroups were strongly supported in the ML analysis ($P < 0.01$) but only moderately in the NJ bootstrap analysis. The nine nonidentical familial LTR sequences belonged to the cosmopolitan subtype HTLV-1a. All the Peruvian and Argentinean familial HTLV-1 strains cluster within the transcontinental subgroup A of subtype HTLV-1a. More specifically, they group within the previously described Latin American clade (NJ, 83%; ML, $P < 0.01$). The strains of family AA (Ar55, Ar56, Ar57, Ar58, and Ar63) were an exception, as they seemed to

belong to a different Latin American clade of subgroup A (NJ, 88%; ML, $P < 0.01$). Nonidentical HTLV-1 strains from the same family cluster closely together, with the exception of Ar11 from family E, which is more distantly related to other strains from the same family (Ar12, Ar15, Ar16, and Ar64). The family from French Guyana cluster within the West African/Caribbean subgroup C of subtype HTLV-1a (NJ, 60%; ML, $P < 0.01$), together with a strain (NM1626) from the same geographic area. The strains sampled from the Brazilian family are very closely related to B13.Peru and were all obtained from individuals of black origin. The NJ analysis clustered the Brazilian and Peruvian strains and positioned them as the most divergent lineage within the Japanese subgroup B, although this was not supported by the bootstrap analysis. However, the ML

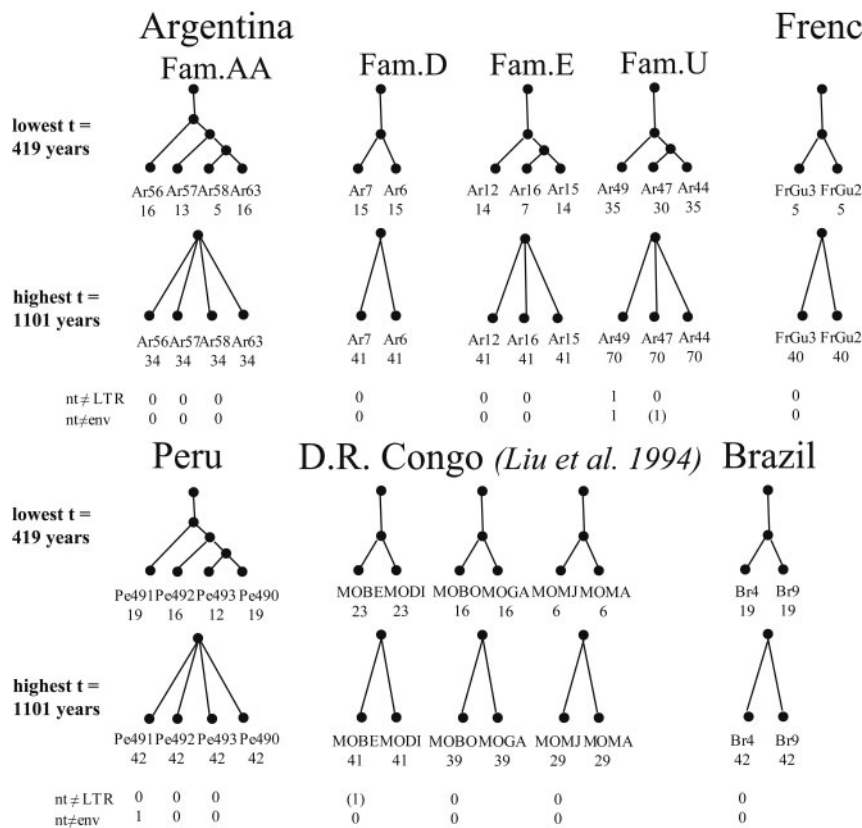


FIG. 3.—The HTLV-1 vertical transmission chains, demonstrating the calculation of the lowest ($t = 419$ years) and highest ($t = 1,101$ years) possible times during which the mutations within the vertically infected family members occurred. The lowest time is the result of the sum of the ages of the different offspring at the time of sampling, plus the age of the oldest child for the mother to correct for shared ancestry. For the highest time, the age of the mother at the time of sampling was counted for each offspring and also for the mother (thus assuming no shared ancestry). The number of mutations observed in the LTR and env regions are shown below the transmission chains. The observed ambiguities are noted in parentheses.

analysis supported a branch that clustered Br4 and Br9 with B13.Peru, separate from any other known HTLV-1a subgroup.

HTLV-1 Evolutionary Rate Estimate

The sequencing results revealed 11 mutations in the LTR and 12 mutations in the env region, across all investigated strains. The LTR mutations were found in three different regions: U3, R, and U5. The majority of the env mutations were synonymous (eight out of 12). As noted earlier, only sequences obtained from vertically infected family members were used in the rate estimations. Further, HTLV-1 strains were omitted from the calculations when the possibility of horizontal transmission could not be ruled out. In cases where the grandmother, mother, and some of the children are seropositive, such that vertical transmission over three generations can be assumed, the transmission from grandmother to mother was not included if the HTLV-1 serostatus of the father was positive or unknown, because the origin of the mother's infection cannot be determined.

Among the eight investigated families, 16 vertical transmission chains were available for study (see fig. 3), containing two mutations in the LTR region and three mutations in the env region. Two of the proposed positions were ambiguous (one in the LTR sequence from MODI that was not fixed in the next generation and one in the

env sequence from Ar47), making it impossible to draw conclusions regarding mutation fixation. Mutations in the LTR sequences occurred only in the R and U5 regions, and mutations in the env region occurred only at third codon positions (two synonymous mutations in family U and one nonsynonymous mutation in the Peruvian family).

It is noteworthy that all mutations, including ambiguous ones, occurred in the oldest individuals. The HTLV-1 strains of the family members between 5 and 19 years of age had no mutations, whereas sequences from individuals older than 18 years (19, 23, 30, and 35 years of age, respectively) showed at least one mutation in the LTR and/or env. To assess the statistical uncertainty arising from the presence of ambiguous mutations, separate evolutionary rate estimates were calculated for (1) no ambiguous changes, (2) one ambiguous change, and (3) two ambiguous changes (see table 1).

The HTLV-1 evolutionary rate was first estimated separately for the LTR and env regions (table 1). The average evolutionary rate for the LTR (756 sites) was estimated to be 1.77×10^{-6} subs./site/year for one fixed mutation and 3.56×10^{-6} subs./site/year, including the ambiguous position. The env data (522 sites) resulted in an average rate of 5.15×10^{-6} subs./site/year for two fixed mutations and 7.30×10^{-6} subs./site/year, including the ambiguity. However, the confidence intervals were large

Table 1
Estimated HTLV-1 Evolutionary Rates (λ) Obtained Using Maximum Likelihood Under a Homogenous Poisson Process Model

Region (Number of Sites)	N	Estimate of λ	Lower 95% Confidence Limit	Upper 95% Confidence Limit
LTR (756)	1	1.77×10^{-6}	9.58×10^{-8}	8.73×10^{-6}
	2	3.56×10^{-6}	5.41×10^{-7}	1.25×10^{-5}
Env (522)	2	5.15×10^{-6}	7.84×10^{-7}	1.82×10^{-5}
	3	7.30×10^{-6}	1.73×10^{-6}	2.33×10^{-5}
LTR-env (1,278)	3	2.98×10^{-6}	7.06×10^{-7}	9.53×10^{-6}
	4	4.00×10^{-6}	1.16×10^{-6}	1.17×10^{-5}
	5	5.04×10^{-6}	1.64×10^{-6}	1.38×10^{-5}
LTR-env 3rd cdp (930)	3	4.10×10^{-6}	9.70×10^{-7}	1.31×10^{-5}
	4	5.49×10^{-6}	1.59×10^{-6}	1.61×10^{-5}
	5	6.93×10^{-6}	2.25×10^{-6}	1.89×10^{-5}

NOTE.—N denotes the number of mutations accumulated in each region during the time represented by the family pedigrees, which ranges from 1,101 years to 419 years. The upper and lower 95% confidence limits were obtained using the likelihood ratio statistic.

because of the small number of mutations observed. For a more accurate estimate with narrower confidence intervals, we combined the complete LTR and env regions. This resulted in an average evolutionary rate estimate ranging from 2.98 to 5.04×10^{-6} subs./site/year, depending on the number of mutations (three to five) included in the analysis. To make a valid comparison between our new estimates and those previously obtained using phylogenetic and anthropological analyses, we also combined the LTR data with the third codon position of env. This resulted in an average rate ranging from 4×10^{-6} subs./site/year (based on the three fixed mutations) to 5.49 to 6.93×10^{-6} subs./site/year (three fixed mutations + ambiguous positions). Figure 4 shows the likelihood curves obtained for the LTR + env3rd codon position data set.

Discussion

Only a small number of mutations were observed in the LTR and env sequences that were sampled from the vertical transmission chains, confirming the genetic stability of HTLV. The observation that mutations only occur in the oldest family members implies that sequence divergence increases with age, suggesting continued viral replication and/or oligoclonal expansion after childhood. The appearance of mutations after adolescence is unlikely to be caused by sexual transmission of a new variant. This would result in a higher observed sequence divergence, as seen in the horizontally infected persons belonging to the Argentinean families D and E (four to nine mutations + ambiguities in the LTR and three to six mutations + ambiguities in env).

Calculation of evolutionary rates usually requires the assumption of a molecular clock (Zuckerandl and Pauling 1962). However, the validity of the clock assumption for viruses is still a matter of controversy (Holmes 2003). Although the hypothesis of a strict molecular clock is rejected for many viral data sets, most likely because

evolutionary rates vary among lineages, the resulting effect on estimated dates may be small if the rate variation is not large (Jenkins et al. 2002). Previous studies of PTLV have shown that the clock assumption may be applied but only under certain conditions. Either lineages (or complete clades) deviating from the clock must be removed from the analysis (Van Dooren et al. 1998), or sites affecting the molecular clock must be eliminated (Salemi et al. 2000; Van Dooren et al. 2001), resulting in the loss of molecular information. Alternatively, statistical methods that explicitly incorporate rate variation could be used (e.g., Thorne et al. 1998). HTLV-1 sequences from vertical transmission chains might be expected to evolve in a more clocklike manner, as the heterogeneity in the evolutionary rate is probably smaller when closely related sequences are considered.

The estimation of the HTLV-1 evolutionary rate using a molecular clock requires an accurate calibration date for at least one node in the phylogeny. Previous analyses (Yanagihara et al. 1995; Salemi et al. 1998; Salemi, Desmyter, and Vandamme 2000; Van Dooren et al. 1998; Van Dooren, Salemi, and Vandamme 2001) have used time points based on anthropological events in the history of the viral host. In the analysis presented here, the calibration dates represent a well-known time frame, corresponding to the total number of years during which mutations in the vertical transmission chains occurred. More specifically, this time frame was expressed as the highest and lowest possible times for the accumulation of mutations among the different pedigrees. The reasoning behind the highest possible time is that, theoretically, one cannot know whether a mutation observed in the HTLV strain of the offspring was already present in the HTLV population of the mother before she gave birth (e.g., as a minor variant among the major HTLV clones). The lowest possible time assumes that the mutation really arose within the offspring. Thus for each mother-offspring pair, the highest or lowest possible times represent the minimum and maximum amount of shared common ancestry.

The inclusion of ambiguous positions in the calculations led to several different rate estimates. The estimates presented in table 1 revealed that the LTR and env regions evolve at approximately equivalent rates, with a slightly higher rate for the envelope gene compared with the LTR noncoding region. The combined LTR-env region resulted in estimates with smaller confidence intervals, thanks to an increased total amount of mutations.

It is important to note that the rate estimates provided in table 1 reflect the evolution of HTLV-1 within vertically infected family members. As vertical HTLV-1 transmission is one of the main transmission routes in endemically infected areas, our estimate should reflect the true HTLV-1 evolutionary rate in such populations. All previously published estimates lie within the confidence limits of our new estimates. Our estimates for the LTR correspond to those of a previous study that investigated the cosmopolitan HTLV-1a subtype and its introduction into Latin America. (Van Dooren et al. 1998). The hypothesis concerning the dissemination of HTLV-1a in the New World is still a matter of controversy. Some analyses suggest the virus was first introduced on the

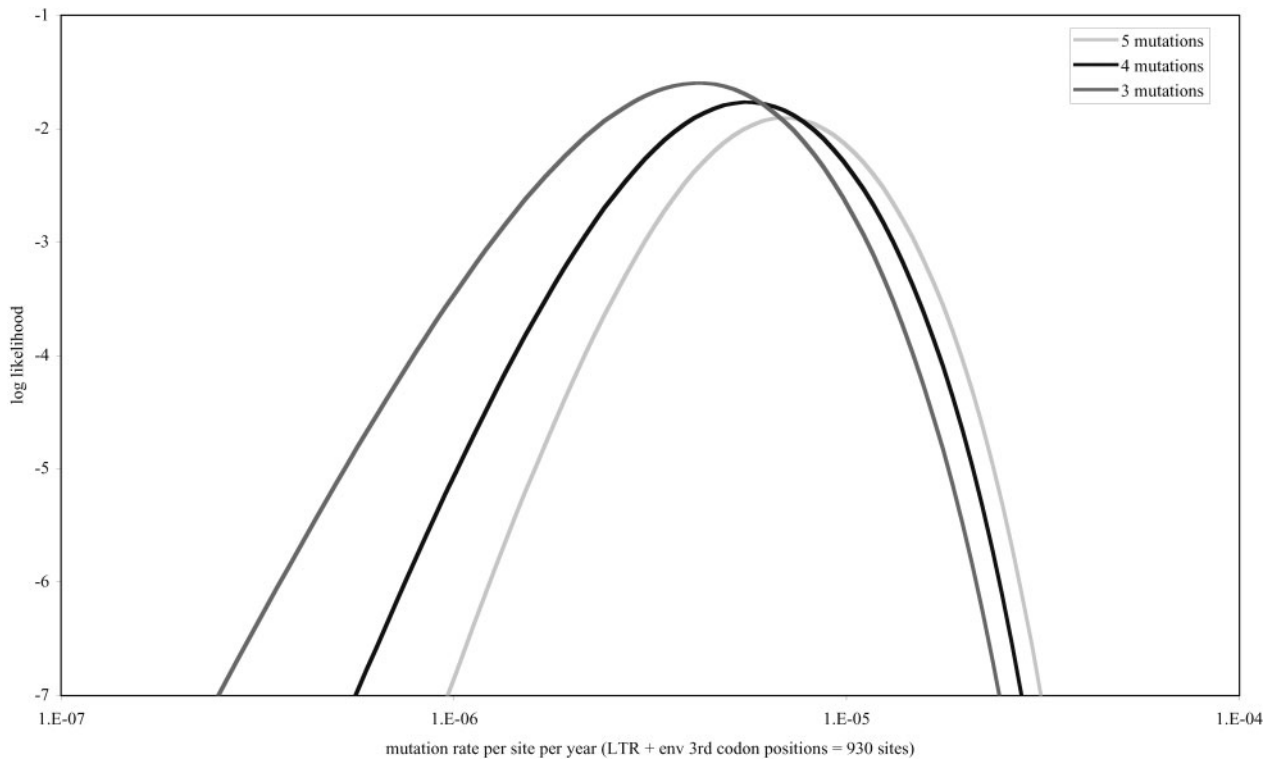


FIG. 4.—The log likelihood curves for the HTLV-1 evolutionary rate estimate from the LTR-env third codon position data set (= 930 sites). Separate estimates of λ were obtained without ($n = 3$) and with ($n = 4$ and $n = 5$) the two proposed ambiguous positions.

American continent by an ancient migration of mongoloids across the Bering Strait 40,000 to 10,000 years ago (Miura et al. 1994, 1997; Yamashita et al. 1998; Ohkura et al. 1999; Li et al. 1999, Ramirez et al. 2002). Other molecular epidemiological studies indicate that HTLV was first introduced along with the post-Columbian African migration, which started approximately 400 years ago (Gessain et al. 1992, Gessain, Gallo, and Franchini 2000; Vandamme et al. 1994, 2000). Assuming either an ancient introduction or a post-Columbian introduction, the HTLV-1 LTR rate was estimated to be around 1.25 to 5×10^{-7} or 1.25×10^{-5} subs./site/year, respectively. These estimates correspond to the upper and lower 95% confidence limits of the estimates for the LTR provided here. However, this comparison is complicated by the worldwide distribution of HTLV-1a strains included in the previous study; these strains are not necessarily restricted to the endemically infected populations. If horizontal transmission is greater among these nonendemic HTLV-1a isolates, then a rate that only takes into account vertical transmission will be an underestimate. In that case, the recent introduction theory would be more plausible. We have previously published an estimated rate for the combined LTR+ env third codon position (Van Dooren, Salemi, and Vandamme 2001), using a data set containing African HTLV-1 subtypes and some STLV-1 strains that cluster with the human strains. In that study, the earliest human migrations to Melanesia and Australia 60,000 to 40,000 years ago were used as an anthropological calibration point, as these dates could be correlated with the isolated presence of HTLV-1 subtype c in Australia and Melanesia. This study reported a rate of $1.56 \pm 0.43 \times 10^{-6}$ subs./site/year, similar to the range

calculated here for four mutations. The analysis of Van Dooren, Salemi, and Vandamme (2001) contained simian strains, and the presence of this cross-species transmission event could have led to a difference between the estimated rates. Other published estimates of the evolutionary rate in PTLV, based on larger concatenated gene regions, resulted in rate estimates of 2.5 to 6.8×10^{-7} subs./site/year (Yanagihara et al. 1995) and $\leq 1.67 \pm 0.17 \times 10^{-6}$ subs./site/year at the third codon position, using the Melanesian migration calibration (Salemi, Desmyter, and Vandamme 2000). These data cannot be compared directly with the vertical transmission chain estimates presented here, because different coding regions and/or different PTLV types were investigated.

Using 16 HTLV-1 vertical transmission chains from eight HTLV-1 infected families, we have confirmed the genetic stability of the virus, even in genomic regions such as LTR and env that are expected to be relatively variable. Our estimates are based on a new approach that is independent of anthropological calibrations. In the future, more precise rate estimates could be calculated if samples and medical histories from patients were gathered over more than three generations, illustrating the importance of long-term sample collection, curation, and storage.

Acknowledgments

We thank Eddie Holmes and Philippe Lemey for the helpful discussions and for the editorial advice. This work was supported by the Fonds voor Wetenschappelijk

Onderzoek (Grant 0288.01). O.G.P. was funded by The Wellcome Trust.

Literature Cited

- Ando, Y., Y. Matsumoto, S. Nakano, K. Saito, K. Kakimoto, T. Tanigawa, Y. Ekuni, M. Kawa, and T. Toyama. 2003. Long-term follow up study of vertical HTLV-I infection in children breast-fed by seropositive mothers. *J. Infect.* **46**:177–179.
- Alcantara, L. C., S. Van Dooren, M. S. Gonçalves et al. (12 co-authors). 2003. Globin haplotypes of human T-cell lymphotropic virus type I (HTLV-I) infected individuals in Salvador, Bahia, Brazil suggest a postcolumbian African origin of this virus. *J. Acquir. Immune Defic. Syndr.* **33**:536–542.
- Bittencourt, A. L. 1998. Vertical transmission of HTLV-I/II: a review. *Rev. Inst. Med. Trop. Sao Paulo* **40**:245–251.
- 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.
- Gessain, A., J. Pecon-Slattery, L. Meertens, and R. Mahieux. 2000. Origins of HTLV-I in South America. *Nat. Med.* **6**:232.
- Hjelle, B., R. Mills, G. Mertz, and S. Swenson. 1990. Transmission of HTLV-II via blood transfusion. *Vox. Sang.* **59**:119–122.
- Holmes, E. C. 2003. Molecular clocks and the puzzle of RNA virus origins. *J. Virol.* **77**:3893–3897.
- Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson. 1998. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**:403–405.
- Jenkins, G. M., A. Rambaut, O. G. Pybus, and E. C. Holmes. 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J. Mol. Evol.* **54**:156–165.
- Kaplan, J. E., R. F. Khabbaz, E. L. Murphy et al. (15 co-authors). 1996. Male-to-female transmission of human T-cell lymphotropic virus types I and II: association with viral load. The Retrovirus Epidemiology Donor Study Group. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **1**:193–201.
- Lee, H., P. Swanson, V. S. Shorty, J. A. Zack, J. D. Rosenblatt, and I. S. Chen. 1989. High rate of HTLV-II infection in seropositive i.v. drug abusers in New Orleans. *Science.* **28**:471–475.
- Li, H.C., T. Fujiyoshi, H. Lou, S. Yashiki, S. Sonoda, L. Cartier, L. Nunez, I. Munoz, S. Horai, and K. Tajima. 1999. The presence of ancient human T-cell lymphotropic virus type I provirus DNA in an Andean mummy. *Nat. Med.* **5**:1428–1432.
- Li, W. H., M. Tanimura, and P. M. Sharp. 1988. Rates and dates of divergence between AIDS virus nucleotide sequences. *Mol. Biol. Evol.* **5**:313–330.
- Liu H.F., A. M. Vandamme, K. Kazadi, H. Carton, J. Desmyter, and P. Goubau. 1994. Familial transmission and minimal sequence variability of human T-lymphotropic virus type I (HTLV-I) in Zaire. *AIDS Res. Hum. Retrovir.* **10**:1135–1142.
- Maddison W., and R. Maddison. 1992. MacClade: analysis of phylogeny and character evolution. Version 3. Sinauer Associates, Sunderland, Mass.
- Manns, A., R. J. Wilks, E. L. Murphy, G. Haynes, J. P. Figueroa, M. Barnett, B. Hanchard, and W. A. Blattner. 1992. A prospective study of transmission by transfusion of HTLV-I and risk factors associated with seroconversion. *Int. J. Cancer* **30**:886–891.
- Miura, T., T. Fukunaga, T. Igarashi et al. (20 co-authors) 1994. Phylogenetic subtypes of human T-lymphotropic virus type I and their relations to the anthropological background. *Proc. Natl. Acad. Sci. USA* 1994 **91**:1124–1127.
- Miura, T., M. Yamashita, V. Zaninovic et al. (11 co-authors). 1997. Molecular phylogeny of human T-cell leukemia virus type I and II of Amerindians in Colombia and Chile. *J. Mol. Evol.* **44**:S76–82.
- Murphy, E. L., J. P. Figueroa, W. N. Gibbs, A. Brathwaite, M. Holding-Cobham, D. Waters, B. Cranston, B. Hanchard, and W. A. Blattner. 1989. Sexual transmission of human T-lymphotropic virus type I (HTLV-I). *Ann. Intern. Med.* **1**:555–560.
- Nerurkar, V. R., K. J. Song, N. Saitou, R. R. Melland, and R. Yanagihara. 1993. Interfamilial and intrafamilial genomic diversity and molecular phylogeny of human T-cell lymphotropic virus type I from Papua New Guinea and the Solomon Islands. *Virology* **196**:506–513.
- Ohkura, S., M. Yamashita, L. Cartier, D. G. Tanabe, M. Hayami, S. Sonoda, and K. Tajima. 1999. Identification and phylogenetic characterization of a human T-cell leukaemia virus type I isolate from a native inhabitant (Rapa Nui) of Easter Island. *J. Gen. Virol.* **80**:1995–2001.
- Plancouline, S., R. P. Buigues, E. L. Murphy et al. (12 co-authors). 1998. Demographic and familial characteristics of HTLV-I infection among an isolated, highly endemic population of African origin in French Guiana. *Int J Cancer* **4**:331–336.
- Ramirez, E., L. Cartier, C. Villota, and J. Fernandez. 2002. Genetic characterization and phylogeny of human T-cell lymphotropic virus type I from Chile. *Virus Res.* **20**:135–149.
- Salemi, M., J. Desmyter, and A. M. Vandamme. 2000. Tempo and mode of human and simian T-lymphotropic virus (HTLV/STLV) evolution revealed by analyses of full-genome sequences. *Mol. Biol. Evol.* **17**:374–386.
- Salemi, M., A. M. Vandamme, J. Desmyter, C. Casoli, and U. Bertazzoni. 1999. The origin and evolution of human T-cell lymphotropic virus type II (HTLV-II) and the relationship with its replication strategy. *Gene* **24**:11–21.
- Salemi, M., A. M. Vandamme, C. Gradozzi, K. Van Laethem, E. Cattaneo, G. Taylor, C. Casoli, P. Goubau, J. Desmyter, and U. Bertazzoni. 1998. Evolutionary rate and genetic heterogeneity of human T-cell lymphotropic virus type II (HTLV-II) using isolates from European injecting drug users. *J. Mol. Evol.* **46**:602–611.
- Sullivan, M. T., A. E. Williams, C. T. Fang, T. Grandinetti, B. J. Poiesz, and G. D. Ehrlich. 1991. Transmission of human T-lymphotropic virus types I and II by blood transfusion: a retrospective study of recipients of blood components (1983 through 1988). The American Red Cross HTLV-I/II Collaborative Study Group. *Arch. Intern. Med.* **151**:2043–2048.
- Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b5. Sinauer Associates, Sunderland, Mass.
- Take, H., M. Umemoto, K. Kusuhara, and K. Kuraya. 1993. Transmission routes of HTLV-I: an analysis of 66 families. *Jpn. J. Cancer Res.* **84**:1265–1267.
- Thorne, J. L., H. Kishino, and I. S. Painter I. S. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**:1647–1657.
- Vandamme, A. M., W. W. Hall, M. Lewis, P. Goubau, and M. Salemi. 2000. Origins of HTLV-I in South America. *Nat. Med.* **6**:232–233.
- Vandamme, A. M., H. F. Liu, P. Goubau, and J. Desmyter. 1994. Primate T-lymphotropic virus type I LTR sequence variation and its phylogenetic analysis: compatibility with an African origin of PTLV-I. *Virology.* **202**:212–223.
- Van Dooren, S. E. Gotuzzo, M. Salemi et al. (11 co-authors) 1998. Evidence for a post-Columbian introduction of human

- T-cell lymphotropic virus type I in Latin America. *J. Gen. Virol.* **79**:2695–2708.
- Van Dooren, S., M. Salemi, and A. M. Vandamme. 2001. Dating the origin of the African human T-cell lymphotropic virus type-I (HTLV-I) subtypes. *Mol. Biol. Evol.* **18**:661–671.
- Wattel, E., M. Cavrois, A. Gessain, and S. Wain-Hobson. 1996. Clonal expansion of infected cells: a way of life for HTLV-I. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **13**:S92–S99.
- Wattel, E., J. P. Vartanian, C. Pannetier, and S. Wain-Hobson. 1995. Clonal expansion of human T-cell leukemia virus type I-infected cells in asymptomatic and symptomatic carriers without malignancy. *J. Virol.* **69**:2863–2868.
- Wiktor, S. Z., E. J. Pate, P. S. Rosenberg, M. Barnett, P. Palmer, D. Medeiros, E. M. Maloney, and W. A. Blattner. 1997. Mother-to-child transmission of human T-cell lymphotropic virus type I associated with prolonged breast-feeding. *J. Hum. Virol.* **1**:37–44.
- Yamashita, M., E. Ido, T. Miura, and M. Hayami. 1996. Molecular epidemiology of HTLV-I in the world. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **1**(Suppl):S124–S131.
- Yamashita, M., G. Picchio, R. Veronesi, S. Ohkura, P. Bare, and M. Hayami. 1998. HTLV-Is in Argentina are phylogenetically similar to those of other South American countries, but different from HTLV-Is in Africa. *J. Med. Virol.* **55**:152–160.
- Yanagihara, R., N. Saitou, V. R. Nerurkar, K. J. Song, I. Bastian, G. Franchini, and D. C. Gajdusek. 1995. Molecular phylogeny and dissemination of human T-cell lymphotropic virus type I viewed within the context of primate evolution and human migration. *Cell Mol. Biol.* **41**:S145–161.
- Yang, Y. C., T. Y. Hsu, M. Y. Liu, M. T. Lin, J. Y. Chen, and C. S. Yang. 1997. Molecular subtyping of human T-lymphotropic virus type I (HTLV-I) by a nested polymerase chain reaction-restriction fragment length polymorphism analysis of the envelope gene: two distinct lineages of HTLV-I in Taiwan. *J. Med. Virol.* **51**:25–31.
- Zuckerkindl, E., and L. Pauling. 1962. Molecular disease, evolution, and genetic heterogeneity. Pp 189–225 *in* M. Kasha and B. Pullman, eds. *Horizons in Biochemistry* Academic Press, New York.

Dan Graur, Associate Editor

Accepted November 4, 2003