The simian origins of the pathogenic human T-cell lymphotropic virus type I
Anne-Mieke Vandamme, Marco Salemi and Jan Desmyter

At least four, and possibly six, molecular subtypes of human T-cell lymphotropic virus type I (HTLV-I) exist; one is confined to Melanesia/Australia, one is ubiquitous, and the others are found only in Africa. Molecular epidemiology suggests that all subtypes arose from separate interspecies transmissions from simians to humans.

Origins of PTLV-I
HTLV is a complex retrovirus consisting of a long terminal repeat (LTR) promoter, structural (gag and env) and enzymatic (pro and pol) coding regions, and regulatory (tax and rex) and accessory coding regions. The accessory proteins differ between PTLV types I, II and L (Refs 8,22,23), and even between subtypes45,52. HTLV has several remarkable features that make it an exceptional tool and subject for phylogenetic analysis. First, HTLV is maintained in an endemic population, mainly by mother-to-child transmission, owing to limited horizontal transmission and the low mortality rate. In this way, it acts as a genetic marker for some populations44. Second, compared with HIV, HTLV has a very stable genome. The LTR of HTLV-II has a very high mutation rate. In drug users has been estimated to have an incidence rate of 20-30 per 100,000 person-years. Third, the suffix of HTLV-II, the accessory protein, is not found in other pathogens, such as Plasmodium falciparum, or to autologous reactions56. Most of the western blot indeterminates are PCR negative and are thus assumed to be unaffected, although some can be identified as being infected with HTLV-I or HTLV-II (Refs 18,20).

They indeterminate serological reactivity could result from infection with divergent strains or from a poor immune response57,58. Epidemiological studies have revealed that the endemic foci of HTLV-I are scattered across Asia, Africa, America and Oceania59. The areas with the highest prevalence are Japan and equatorial Africa, and the lowest prevalence is found in Europe.
cellular polymerases therefore limits the genomic variability found in a single patient. This raises the possibility that the evolutionary rate is slower for viruses that are transmitted vertically, with a longer period of clonal expansion between transmission events, than for viruses that are, for example, transmitted among drug users. The fact that HTLV acts as a genetic marker and has a stable genome means that phylogenetic analysis of HTLV strains from endemic populations can be used to trace the origin of the virus in the distant past. Thus, the spread of the virus can be correlated to human migrations of ethnic groups and to contacts between populations in a time-frame that can span up to 100,000 years.

Analysis of the gp21 env region
Although the env region is not the most divergent region, it gives the best available picture of the evolution of PTLV-I, as genetic information from the largest number of different PTLV-I strains is available for this region. The most striking characteristics of the env PTLV-I phylogenetic tree are that clades of HTLV-I strains are interspersed between STLV-I clades and that the root of the tree lies within the macaque clade (Fig. 1). To date, four HTLV-I clades or subtypes have been well characterized, and two potential new subtypes have recently been found. The HTLV-Ia, or cosmopolitan, clade contains strains from all over the world. HTLV-Ib is found only in central Africa, and HTLV-Ic is found only in Melanesia/Australia among aboriginals. HTLV-IId is also a central African clade found among Cameroonians.

HTLV-Ile D.R. Congo
HTLV-If Gabon
HTLV-Ia
Cosmopolitan

African/cosmopolitan
strains

HTLV-Ib
Central Africa

HTLV-Ia,
Cosmopolitan

HTLV-Ic
Melanesia

HTLV-Ia,
Cosmopolitan

India

Sukhumi Primate Centre

Indonesia

Host species
- Homo sapiens
- Cercopithecus allosi
- Cercopithecus anthops
- Cercopithecus mitis
- Macaca fascicularis
- Macaca mulatta
- Macaca nemestrina
- Macaca tonkeana
- Mandrillus sphinx
- Pan troglodytes
- Papio cynocephalus
- Papio hamadryas
- Papio anubis
- Papio papio

Fig. 1. Phylogenetic analysis of 75 human and simian T-cell lymphotropic virus type I (HTLV-I and STLV-I, respectively) strains, using a 522-nt consensus fragment of the env region, corresponding to almost the entire gp21 coding sequence (nt 6046-6567 of HTLV-I strain ATK1). The maximum-likelihood method was used (Phylip software package version 3.572 (Ref. 54)) with a transition/transversion bias of 5:1. This tree is rooted by using the HTLV-II strain Mo and the STLV-L strain PH969 as outgroup strains, and the root node is indicated. The tree is drawn in a star-like format, to illustrate clearly the real proportions of the branches and clades. Statistical evaluation of branch lengths results in some clades being better supported than others (** = p < 0.01). The branch lengths are proportional to the evolutionary distance (scale bar) between the taxa. The human and simian strains have symbols marked according to the host species, and the human clades are circled. The strains used in this analysis are described in Refs 19,39.
among the macaque strains, clearly implies a simian origin for PTLV-I and indicates that interspecies transmissions have occurred between simians and humans10–13.

The entire phylogenetic tree can be separated into an African part according to the geographical origin of the virus strains, with the exception of the origin of the cosmopolitan HTLV-Ib subtype (as indicated in Fig. 1). The Asian/Australian part of the tree is much more complicated and has many more HTLV-I and STLV-I clusters, all apparently having arisen during an explosive spread of PTLV-I in Africa. All human and simian clades have a common origin, with no significant clustering of any two clades. Within each clade, several strains from different species (simians and humans) can be found. Simian strains of similar geographic origin but from different species cluster together: for example, strains from West African common chimpanzees (Pan troglodytes) and from vervet monkeys (Cercopithecus aethiops) combine in one cluster; similarly, strains from East African baboon species (Papio spp.) combine with East African Cercopithecus spp. strains, suggesting several interspecies transmissions among simians. Four of the six human subtypes cluster with simian strains; the central African HTLV-Ib subtype clusters with the common chimpanzee strains from unknown origin or from West Africa; the Cameroon pygmy HTLV-IId subtype clusters with mandrils (Mandrillus sphinx) strains from Gabon; the Efe pygmy HTLV-Ile subtype clusters with Congolese (Zairean) and East African baboon and Cercopithecus spp. strains from various species; and the HTLV-Ie subtype from Gabon clusters with a mandril (M. sphinx) strain from Gabon.

The African part of the tree is much more complicated and has many more HTLV-I and STLV-I clusters, all apparently having arisen during an explosive spread of PTLV-I in Africa. All human and simian clades have a common origin, with no significant clustering of any two clades. Within each clade, several strains from different species (simians and humans) can be found. Simian strains of similar geographic origin but from different species cluster together: for example, strains from West African common chimpanzees (Pan troglodytes) and from vervet monkeys (Cercopithecus aethiops) combine in one cluster; similarly, strains from East African baboon species (Papio spp.) combine with East African Cercopithecus spp. strains, suggesting several interspecies transmissions among simians. Four of the six human subtypes cluster with simian strains; the central African HTLV-Ib subtype clusters with the common chimpanzee strains from unknown origin or from West Africa; the Cameroon pygmy HTLV-IId subtype clusters with mandrils (Mandrillus sphinx) strains from Gabon; the Efe pygmy HTLV-Ile subtype clusters with Congolese (Zairean) and East African baboon and Cercopithecus spp. strains from various species; and the HTLV-Ie subtype from Gabon clusters with a mandril (M. sphinx) strain from Gabon.

Thus, it is possible that in the past the habitat of the infected simians overlapped with the habitat of the affected humans. In the case of the central African HTLV-Ib and HTLV-IId subtypes, STLV-I strains cluster close to human strains, but the point where the clade branches is equally likely to be among the STLV-I strains as among the HTLV-I strains. This suggests that transmission from humans to simians is just as likely as transmission from simians to humans, which is supported by network analyses carried out before the slave trade. Transmission from monkeys, such as baboons or vervet monkeys, to both humans and chimpanzees is also possible. It is known that humans and chimpanzees hunt both these species. If the hunt ends in a bloody fight, this could result in STLV-I transmission from the prey to the hunter. However, the fact that there are more simian than human clusters, together with the fact that the origin of PTLV-I is among simian strains, suggests that the main direction of interspecies transmission is from simians to humans and not the reverse. Thus, although we cannot exclude simian strains arising from transmission of human strains, all human clades must have had a different simian origin and are therefore appropriately called subtypes. In view of these results, it is obvious that the cosmopolitan HTLV-Ia subtype clustering among African STLV-I clades has its origin among African simians. The human strains must have been carried out of Africa by their human host. This human migration might have been either pre-Columbian, during the ancient human migrations out of Africa, or post-Columbian, mainly via the slave trade.

Analysis of the LTR region

As long as sequences are so similar that homologous nucleotides can be aligned unambiguously, regions with the highest genetic diversity can provide most phylogenetic information. Therefore, the LTR region, one of the most divergent regions in PTLV-I, is very informative when tracing the molecular epidemiology of HTLV-I. There is no doubt about the Asian origin of HTLV-Ic or the African origin of the other five HTLV-I subtypes. It is also clear that human migrations are responsible for the worldwide spread of the cosmopolitan subtype HTLV-Ia. Human migrations out of Africa started 100 000–200 000 years ago10, and another major wave of human migrations out of Africa was initiated around the 16th century by the slave trade. During the past century, increased worldwide mobility and sexual contact have given the virus another opportunity to spread in new populations.

A detailed HTLV-Ia LTR tree is shown in Fig. 2. The analysis was done with all available HTLV-Ia strains, using a representative of each of the other well-established HTLV-I subtypes as outgroup strains. As genetic divergence increases with time, the horizontal branch length from left to right roughly represents an arrow in time; hence, the branch points on the left are the earliest. One of the oldest clades contains strains from west Africa and the Caribbean.
Bering Strait 10,000–40,000 years ago, one would speculate of HTLV-Ia via the human migration across the Bering Strait and disseminated among Amerindians in an attempt to solve this paradox.

**Analysis of the tax region**

The tax gene is one of the most stable genes of PTLV. It has maintained sufficient sequence similarity between PTLV types that homologous nucleotides can be aligned unambiguously. Information on divergent strains within each type is also available. The unrooted phylogenetic analysis in Fig. 3 gives an impression of the relationship between PTLV-I, PTLV-II and PTLV-L. From the tree, it is clear that both PTLV-II and PTLV-L have their origins in Africa. PTLV-I has only one representative strain, an African STLV-L. PTLV-II includes both STLV-II and HTLV-II strains that clearly separate according to species of origin (bonobos and humans, respectively). STLV-II is enzootic in the African population, and the deep phylogenetic analysis has shown that the time-frame for the human T-cell lymphotropic virus (HTLV) spreading in the world. Genetic and environmental factors are still being investigated in an attempt to solve this paradox.

**Questions for future research**

- In countries where human T-cell lymphotropic virus (HTLV) screening is mandatory, individuals with indeterminate sera are excluded from blood donation. How many are infected with HTLV-I or other retroviruses? Should the donors be notified about the indeterminate results, and how should they be counseled?
- Can HTLV seronegatives harbor divergent unknown HTLV types?
- How many other simian-to-human transmissions have occurred in the past or will occur in the future? Do these transmissions only occur in tropical areas?
- Will the use of non-human primate organs for xenotransplantation into humans give rise to more pathogenic HTLV viruses?
- Where is the missing link between the African origin of primate T-cell lymphotropic viruses (PTLVs) and the Asian origin of PTLV-I? How did the Asian PTLV-I lineage appear to be the Asian PTLV-I lineage appear to evolve?
- Why is the evolutionary rate of HTLV so slow? Is the evolutionary rate dependent on the mode of transmission?
have migrated to Africa, giving rise to the African explosion of simian and human strains (Figs 1, 3).

Conclusions

The current data support the notion that HTLV-I has arisen several times from STLV-I strains by geographically separate interspecies transmissions, whereas all HTLV-II strains have a common ancient human ancestor virus. This phenomenon might partly explain why HTLV-I is more pathogenic than HTLV-II.

Based on the current sampling of primate species and virus strains, the most likely origin of PTLV is non-human primates in Africa, possibly Eastern Africa. An early lineage of this STLV evolved into STLV-I in Asia (Fig. 4). The Asian STLV-I jumped to humans along their migratory pathway to Melanesia/Australia, resulting in a separate Asian/Australian HTLV-Ic subtype. The STLV that remained in Africa split into PTLV-II (STLV-II in bonobos and HTLV-II in humans) and STLV-I (in baboons). Much later, the Asian STLV-I lineage returned to Africa and spread rapidly in the susceptible non-human primates. Several separate interspecies transmissions resulted in five African HTLV-I subtypes, one of which (HTLV-Ia) spread throughout the world, owing to the increased post-Columbian mobility of its human host (Fig. 4). The close clustering of at least four of these human HTLV-I subtypes (Ib, Ie, Id and If) with simian strains suggests recent or ongoing interspecies transmissions between simians and humans. Characterization of these simian retroviruses is also important in view of the growing interest in xenotransplantation; an interspecies jump of STLV-related viruses might lead to increased pathogenicity in the new host.

Of the six HTLV-I subtypes, three (Id, Ie and If) were discovered during 1997, and within three subtypes (Ib, Ie and If) individuals with indeterminate western blot serology have been found. During screening for HTLV antibodies, indeterminate results are often found. It might be worth checking for the presence of divergent viruses in these cases. A better knowledge of the simian viruses would contribute to the development of better diagnostic tools for detecting the presence of related viruses in blood donations.

As PTLV-I has been found in humans and in many simian species, it is difficult to trace its origins. In this context, it would be interesting to screen Barbary macaques (Macaca sylvanus) for the presence of STLV-related viruses, as most divergent Asian STLV-I strains have been found in macaques. Barbary macaques are the only African/European macaques and are believed to be the remnant of the earliest migration wave of macaques from Africa over Europe to Asia, before the further speciation of this genus in Asia53.

Acknowledgements

The Rega Institute belongs to the HERN concerted action supported by the Biomed programme of the European Commission. We are most grateful to Patrick Goubau, Marianne Van Brussel and Sonia Van Dooren for their significant contribution to this paper. We thank Jill Esnouf for editorial help.

Fig. 3. Unrooted phylogenetic analysis of 30 human and simian T-cell lymphotropic virus type I (HTLV-I and STLV-I, respectively) strains, using a 401-nt fragment of the pX region, covering part of the tax/rex open reading frames (nt 7359–7759 of HTLV-I strain ATK1), using a maximum likelihood method with a transition/transversion bias of 2.0, constructed as described in Fig. 1. Statistical evaluation of branch lengths results in some clades being better supported than others (**p < 0.01). The branch lengths are proportional to the evolutionary distance (scale bar) between the taxa. All available HTLV-I, HTLV-II, STLV, STLV-II and STLV-L strains were used. The human clades are circled. The strains used in this analysis are described in Refs 12, 26, 31, 50, 52. Abbreviation: PTLV, primate T-cell lymphotropic virus.
Fig. 4. Hypothesis for the origin and global dissemination of human and simian T-cell lymphotropic virus type I (HTLV-I and STLV-I, respectively), together called primate T-cell lymphotropic virus type I (PTLV-I). Migrations of humans carrying HTLV-I strains are indicated by filled arrowheads; migrations of simians carrying STLV-I strains are indicated by open arrowheads. ‘Early’ represents migrations that occurred before the period of post-Columbian or documented human migrations.

References

TJM on the Web
Searching for previous articles of interest in TIM? Listings of TIM articles from the past 4 years, together with subscription details, instructions to Authors, aims and scope of the journal, and editorial board information, can be found at: http://www.elsevier.nl:80/inca/publications/store/4/2/4/1/0/0/