Simian T-cell lymphotropic virus (STLV) has been discovered in a broad spectrum of Old World monkey species and great apes and has been classified into STLV types 1–3. The human variant, human T-cell lymphotropic virus (HTLV), has been discovered in endemic populations, sometimes living in remote areas, but has also become disseminated worldwide. Both simian and human viral strains are generally referred to as primate T-cell lymphotropic virus (PTLV). STLV type 1 (STLV-1) strain MarB43 in *Macaca arctoides* has been discovered in more than 23 different species of the families Cercopithecidae (Pecon Slattery et al., 1999; Courgaud et al., 2004) and Pongidae (Ibuki et al., 1997; Voevodin et al., 1997; Meertens et al., 2001; Nerrienet et al., 2004) and seems to be scattered not only over the African continent, but also among macaque and great ape species living in Asia. HTLV-1, which is classified into six different subtypes, has one cosmopolitan subtype (HTLV-1a; Yamashita et al., 1996), four subtypes restricted to Africa (HTLV-1b, -d, -e and -f; Nerurkar et al., 1993; Liu et al., 1994; Mahieux et al., 1997a; Salemi et al., 1998) and one subtype isolated only in descendants of the first settlers of Melanesia and Australia (HTLV-1c; Gessain et al., 1991). PTLV-1 phylogeny indicates a clear separation between the Asian–Melanesian and African STLV-1/HTLV-1 strains (Gessain et al., 1993; Nerurkar et al., 1993; Song et al., 1994). This distant, continental, phylogenetic relationship of PTLV-1 strains is reflected in a comparative difference in evolutionary patterns, although this might be partially biased by the fragmentary sampling in Asia. Asian STLV-1 strains cluster according to host species and form deep branches, indicating that they have undergone a long and independent species and subspecies evolution (Ibrahim et al., 1995). The same phenomenon has been observed to a lesser extent in its Melanesian/Australian human counterpart, HTLV-1c. On the African continent, the pattern of PTLV-1 infection has been a star-like burst followed by successful spread, characterized by a broad range of interspecies transmissions (Vandamme et al., 1998). The relationship between the different African PTLV-1 strains seems to be dependent more on geography than on the primate host species.

Overall, PTLV-1 phylogeny is peculiar in that it suggests an Asian origin for PTLV-1 when phylogenetic trees are rooted with strains of other PTLV types. This observation questions the generally accepted hypothesis that the ancestor of all PTLVs originated in Africa (Salemi et al., 1999), based on the common occurrence of all different PTLV types on the African continent. Taken together, these facts create doubts about the possible origin of PTLV and the dissemination of PTLV-1 to Asia and Africa.
Full-genome sequencing and analysis of one of the most divergent STLV-1 strains of Asian origin would allow us to unravel a piece of this Asian PTLV-1 dissemination. Mahieux et al. (1997b) demonstrated that six Macaca arctoides of southeast Asian origin housed at the Strasbourg Primatology Center showed a peculiar Western blot profile with an HTLV-2-like reactivity. Sequencing of part of the tax (406 bp) region and part of the env (306 bp) region of one M. arctoides strain, Marc1, showed that this strain was the most divergent among Asian STLV-1 strains. In the present study, we sequenced the STLV-1 gp21 env and long terminal repeat (LTR) regions from four other M. arctoides strains, identified by Mahieux et al. (1997b) [MAC912, MAC757, MAC897 and MAC11 (LTR region only)] according to the methodology described previously (Mahieux et al., 1997a). The STLV-1 full-genome sequence was obtained from (sufficiently available) original lymphocyte DNA of M. arctoides animal B43 (MarB43), as a virus-producing cell line could not be established for Marc1. The primers and PCR conditions for amplification are available at http://www.kuleuven.be/rega/cev/TableMarB43sequencingproject.pdf. The full genome of STLV MarB43 was obtained as shown in the sequencing project represented schematically in Fig. 1(a), after agarose gel purification of the PCR products with a QIAquick gel extraction kit (Qiagen) and direct sequencing of both strands using the BigDye Terminator technology on an ABI 310 Genetic Analyzer (Applied Biosystems). The different sequenced fragments were analysed with Sequencing Analysis software (Applied Biosystems) and assembled by using the software package Geneworks 2.5.1 (Oxford Molecular Systems). The GenBank accession number of the MarB43 complete genome is AY590142, and those of the M. arctoides env and LTR sequences are AY141153–AY141155 and AY141171–AY141174, respectively.

All potential open reading frames (ORFs) of the MarB43 complete sequence were analysed in Geneworks 2.5.1 (Oxford Molecular Systems), verified manually and compared with that of the ATK1 HTLV-1a reference strain. The theoretical splicing pattern was analysed by using different software packages (http://www.biogenio.com/splice/index.html, http://www.lecb.ncifcrf.gov/~toms/spliceanalysis.html, http://l25.itba.mi.cnr.it/~webgene/www.spliceview.html). The genomic organization of MarB43 resembles that of other STLV-1 and HTLV-1 strains (Fig. 1b). The major ORFs were found to be conserved. The 3’ end of the tax gene was comparable with that of HTLV-1 and STLV-1 strains. Based on sequence comparison and theoretical splicing analysis, a few differences in the splicing pattern compared with HTLV-1 could be observed. MarB43 probably uses the HTLV-2/STLV-2 env-tax-rex splice acceptor homologue (GCCCTTCCAG↑GAGG; Fig. 1c) due to a critical point mutation (A→T) in the conventional HTLV-1 splice-acceptor sequence (TATCCGTG↑CTCT). A sequence-comparison analysis of the translation products within the proximal pX region, the pX ORFI, -II and -V, normally obtained by alternative splicing, revealed that they probably cannot be synthesized. The three potential splice-acceptor sites of these ORFs were conserved enough to maintain their function (Fig. 1c). However, identical point mutations within these ORFs, resulting in the absence of initiation codons and/or in early stop codons, were observed in the MarB43 and Marc1 sequences, two variants of this STLV-1 clade sequenced in different laboratories. The only exception was p21 Rex, where the initiation codon remained conserved. Extra stop codons in tol/p30 (pX-Tax-ORFII), rol/p27 (pX-Rex-ORF1) and pX-Rex-ORFV shortly after the splice acceptor sites (after amino acid positions 10, 8 and 31, respectively) would result in truncated proteins. The pX ORF-I and -II contained neither the p12 nor the p13 initiation codons, due to point mutations. On the one hand, the abolishment of the function of these proteins could explain the difficulties encountered during the PCR amplifications and the failure to establish a virus-producing cell line, respectively, as p13 and p30 are possibly involved in the maintenance of high proviral loads and p12 in the infectivity of primary lymphocytes and virus replication (Johnson et al., 2001; Albrecht & Lairmore, 2002). The transmission efficiency of this viral strain, however, has been proven in vivo, where this particular virus has been found circulating among at least six M. arctoides of the same colony (Mahieux et al., 1997b). On the other hand, the latter observation clearly demonstrates the viability of this circulating virus and thus questions the role of p12 (only potentially functional in HTLV-1a strains) in infectivity and virus replication, both essential for the establishment of infection. These possible singly and doubly spliced mRNAs and the alternative splicing in the pX region were further experimentally analysed by RT-PCR on total RNA as described previously (Meertens et al., 2002). MarB43-specific primers were developed to analyse the LTR splice-donor and env and tax-rex splice-acceptor sites, respectively (sense primer 5M386, 5’-TACCTGAGGCGCGCATCCGTG-3’, antisense primers 3M5422, 5’-AAGGGGTFTAGGGCCTGATC-3’, 3M7483, 5’-GAGGCGGCGCTGAGCATGTCCCA-3’, respectively; positions shown in Fig. 1c). Alternative splicing in the pX region was analysed with a primer designed approximately 50 bp upstream of the theoretical tax-rex splice-acceptor site (3M7324, 5’-TTCGCAGATGCTGCTC-3’, position shown in Fig. 1c). However, the different RT-PCRs, following hybridization with specific γ32P-labelled probes in Southern blots, did not identify any cDNA that could correspond to the singly and doubly spliced mRNAs or the pX alternative splicing. This may have been due to either low replication levels or low proviral load.

The phylogenetic relationship between the MarB43 strain and other PTLV strains was investigated through a concatenated gag-pol-env-tax full-genome analysis (with all available PTLV full-genome sequences), and an LTR and gp21 env analysis (with more known Asian STLV-1 strains of different Macaca host species and a substantial number of STLV-1 and HTLV-1 reference strains of African and Melanesian origin from GenBank). Alignments were made
in DAMBE 4.0.75 (Xia & Xie, 2001), guided by their amino acid sequence alignment, and further edited manually in Se-Al (http://evolve.zoo.ox.ac.uk). Neighbour-joining and maximum-likelihood trees were constructed according to methodology described elsewhere (Van Dooren et al., 2004). The full-genome, LTR and env phylogenetic analyses consistently depicted MarB43 as the most divergent STLV-1 strain, clustering closer to the Asian STLV-1 strains and the Melanesian HTLV-1c strain Mel5 than to the other African or cosmopolitan STLV-1/HTLV-1 strains (env and LTR analyses shown in Fig. 2a and b, respectively). The positioning of MarB43 within the PTLV-1 clade was analogous to the positioning of STLV-2 with respect to the HTLV-2 strains (Fig. 2a). The PTLV-1 env and LTR phylogenetic analyses demonstrated clearly that STLV-1 strains in M. arctoides are related closely to each other, but distantly to the STLV-1 strains of other Macaca species. The remaining Asian PTLV-1 strains in the env phylogenetic analysis (Fig. 2a) clustered according to host species in a 'ladder-like' topology, seemingly reflecting the Macaca migration waves. STLV-1 strains in Macaca mulatta and Macaca fascicularis clustered closer to the root of the PTLV-1

**Fig. 1.** (a) MarB43 proviral full-genome sequencing strategy. Four proviral fragments were PCR-amplified and sequenced directly. Three of the four fragments were amplified by using nested PCRs, with one outer (closed arrows) and two to three inner (open arrows) PCRs. (b, c) Schematic representation of the STLV-1 MarB43 proviral genome (b) and the putative resulting mRNAs (c). Nucleotide numbering is according to the MarB43 proviral genome. Putative splice-donor (sd) and splice-acceptor (sa) sites and their positions within the proviral genome are indicated with vertical downward and upward arrows, respectively. The positions of the primers designed to analyse the putative singly and doubly spliced messengers are indicated with small horizontal arrows.
part of the tree. Remarkably, these Macaca species, together with M. arctoides, are all descended from proto-fascicularis macaques, which differentiated approximately 5-5 million years ago from the silenus group (Macaca silenus, Macaca nemestrina and Sulawesi macaques) shortly after their arrival on the Asian continent (Delson, 1980; Morales & Melnick, 1998). M. fascicularis has a current habitat closer to and overlapping with that of the silenus macaques. The other STLV-1 strains located further down the Asian part of the PTLV-1 phylogenetic tree were isolated from Macaca tonkeana and M. nemestrina (in the LTR analysis only; Fig. 2b) of the Indonesian archipelago and in the Japanese Macaca fuscata. The first two Macaca species

Asian northwest-to-central mainland (Delson, 1980; Morales & Melnick, 1998). M. fascicularis has a current habitat closer to and overlapping with that of the silenus macaques. The other STLV-1 strains located further down the Asian part of the PTLV-1 phylogenetic tree were isolated from Macaca tonkeana and M. nemestrina (in the LTR analysis only; Fig. 2b) of the Indonesian archipelago and in the Japanese Macaca fuscata. The first two Macaca species

**Fig. 2.** Maximum-likelihood trees of a 498 bp gp21 env fragment [a, b (inset)] and a 518 bp fragment of the LTR (c) constructed under the Tamura–Nei evolutionary model (tv/ti = 29/7-54 and 1/14-82/7-82, respectively), further modelled with a gamma-distributed rate heterogeneity among sites (x, 0-35 and 0-48, respectively). Numbers on the branches indicate the percentage of neighbour-joining bootstrap samples (of 1000); only values above 75% are shown. P indicates positions in the phylogenetic tree for which the maximum-likelihood zero-branch-length test was not statistically significant. Cae, Cercopithecus aethiops; Cas, Cercopithecus ascanius; Cmi, Cercopithecus mitis; Cpy, Cercopithecus aethiops pygerythrus; Csa, Cercopithecus aethiops sabaeus; Cto, Cercocetus torquatus; Ggo, Gorilla gorilla; Hsa, Homo sapiens; Mar, Macaca arctoides; Mfa, Macaca fascicularis; Mu, Macaca fuscata; Mmu, Macaca mulatta; Mne, Macaca nemestrina; Msp, Mandrillus sphinx; Mta, Miopithecus talapoin; Mto, Macaca tonkeana; Pcy, Papio anubis; Ppa, Papio cynocephalus; Ppa, Papio hamadryas; Ppy, Pongo pygmaeus; Psp, Papio species; Ptr, Pan troglodytes; Tge, Theropithecus gelada. 

1956 Journal of General Virology 86
are descended from the silenus group, which migrated soon after their arrival in Asia across the Sundaland from southern China to what is now known as the Indonesian archipelago (Morales & Melnick, 1998). The Japanese M. fuscata probably differentiated from eastern rhesus macaques in more recent times (0.5 million years ago) (Delson, 1980; Morales & Melnick, 1998). The PTLV-1 topology from root to tip is thus indicative of an Asian host species migration-dependent evolution of the virus, suggestive of virus-host co-evolution, known as vicariance
Moreover, this apparent virus–host co-divergence is inconsistent with molecular timings. The full-genome phylogenetic tree was recalibrated including the new, highly divergent MarB43 strain under a relaxed molecular clock, which accommodates rate variation among lineages, according to methodology described previously (Lemey et al., 2005). One calibration node, used frequently to estimate a timescale for PTLV evolution, appeared to coincide with the earliest human migrations to Melanesia and Australia 60 000–40 000 years ago (Gessain et al., 1993; Nerurkar et al., 1993; Yanagihara et al., 1995; Salemi et al., 2000; Van Dooren et al., 2001; Meertens & Gessain, 2003). Based on this time point, the mean evolutionary rate per gene was estimated to be $7 \times 10^{-7} \times (4 \times 10^{-7} \times 1 \times 10^{-6})$ nucleotide substitutions per site per year. The molecular-clock calculations suggest that STLV-1 originated in Asia at least 269 000–156 000 years ago, using MarB43 as the earliest divergence within the PTLV-1 lineage, whereas PTLV-1 started to diverge in Africa only 23 000–18 000 years ago. The archaic presence of STLV-1 on the Asian continent is much more recent than the Macaca host invasion of Asia. The assumption of virus–host co-evolution is further violated by the occurrence of STLV-1 infection in orangutans (Fig. 2b) (Ibuki et al., 1997; Verschoor et al., 1998), suggesting STLV-1 cross-species transmissions between distantly related primate species. This apparent vicariance and virus–host phylogenetic incongruence has also been observed for a related primate retrovirus, simian immuno-deficiency virus. Charleston & Robertson (2002) demonstrated by simulation analysis that this might be explained by differential success in viral host-switching related to host (sub)species similarity and geographical proximity. A non-negligible difference, however, is the long branching pattern of the Asian host subspecies-specific PTLV-1 monophyletic clusters, suggesting that this preferential host subspecies switching mainly occurred early in STLV-1 evolution and that the virus then further evolved within its specific host.

The current PTLV env phylogeny data certainly corroborate the enormous diversity of PTLV-1 observed in Asia. The STLV-1 biodiversity found to date within macaques suggests that they could have formed the reservoir species for the founding PTLV-1 in Asia. To unravel PTLV evolution further, in particular in Asia, and its relation to Africa, more PTLV strains from primates in the wild should be identified and analysed genetically, especially from Asia, to encompass a possible bias of an incomplete STLV-1 sampling pattern.

Acknowledgements

This work was supported by the Fonds voor Wetenschappelijk Onderzoek (Grant 0288.01). L. M. is supported by the CANAM. We are grateful to N. Herrenschmidt and E. André from the Strasbourg Primatology Center for the generous gift of the biological materials of the M. arctoides troop.

References


