JOURNAL OF VIROLOGY, Dec. 1995, p. 7877–7887 Vol. 69, No. 12
0022-538X/95/$04.00+0

Distribution of Baboon Endogenous Virus among Species of African Monkeys Suggests Multiple Ancient Cross-Species Transmissions in Shared Habitats

ANTOINETTE C. VAN DER KUYL,* JOHN T. DEKKER, AND JAAP GOUDSMIT

Human Retrovirus Laboratory, Academic Medical Centre, 1105 AZ Amsterdam, The Netherlands

Received 16 May 1995/Accepted 21 September 1995

PCR amplification of baboon endogenous virus (BaEV) long terminal repeat, reverse transcriptase gene, and *env* fragments from 24 different species of African monkeys indicates that BaEV is less widespread than was formerly thought. Instead of being present in every species of African primates, BaEV can be found only in baboons, geladas, and mangabeys (all belonging to the *Papionini* tribe) and in African green monkey (*Cercopithecus aethiops*) subspecies. BaEV, which can be activated from baboon and gelada tissues, was most likely introduced in the germ line only recently (less than a few million years ago) and has not been inherited from a common ancestor of all extant African monkeys. Neighbor-joining and maximum-likelihood analyses of the sequences obtained showed that two distinct virus clusters can be distinguished: the first containing baboon, gelada, and African green monkey BaEV sequences and the second consisting of mandrill and mangabey BaEV sequences. This viral evolutionary tree does not follow host phylogeny, indicating that cross-species transmissions and multiple germ line fixations of the virus must have occurred in the past. BaEV sequences are found in monkeys inhabiting savannas (baboons, geladas, and African green monkeys) as well as forests (mangabeys and mandrills) and cluster according to the habitats of their hosts, providing evidence for cross-species transmission in shared habitats.

The family *Retroviridae* has recently been subdivided into seven genera (8), of which the spumavirus, lentivirus, and human T-cell leukemia virus type 1 (HTLV-1)-like virus groups show no evidence that their members ever entered the germ lines of mammal species. Consequently, these viruses must be regarded as true exogenous retroviruses. Members of the other four groups (formerly mainly classified as oncoviruses), however, have been found stably integrated in the cellular genome and being inherited in a Mendelian fashion. Many of these endogenous retroviruses have left their tracks in the mammalian genome, including humans. Endogenous proviral genomes are often defective as a result of residing in the genome for a long period of time, although it is possible that de novo-integrated genomes are already defective. Sometimes the division between exogenous and endogenous retroviruses is not so clear, since nondefective endogenous virus genomes can be expressed and particles can be formed in many host species. This is true for endogenous murine leukemia virus (MuLV) in mice (see reference 32 and references therein), avian endogenous retroviruses (9), RD114 in cats (13), jaagsiekte retrovirus in sheep (35), type C viruses from macaques and *Colobus polykomos* (25, 29, 33), endogenous type D viruses in simians (4), baboon endogenous virus (BaEV) in baboons and geladas (1, 34), and human endogenous retrovirus K in humans (6). Integrated genomes of these viruses may not have yet been inactivated by the host cell because of their recent history in their respective hosts. For MuLV it has been found that new germ line insertions are taking place in several species of wild and inbred mice (18, 26).

More is known about the evolution of exogenous primate retroviruses than about their endogenous counterparts. Primate T-cell leukemia/lymphotropic virus type 1 (PTLV-1) is widespread and can be isolated from many species of both Old and New World monkeys and apes, prosimians, and humans (21, 28). One of the main routes of infection of PTLV-1 is vertical transmission by breast-feeding (5). This way of transmission resembles the inheritance of endogenous viruses. Although PTLV-1 is supposed to be very old, different isolates exhibit little sequence divergence. Phylogenetic analysis of sequences obtained by PCR amplification showed that interspecies transmissions have occurred in the past and new transmissions are probably still occurring between primate species at present.

Less is known about the evolution of (intact) endogenous viruses, especially at the sequence level. Generally it is assumed that endogenous virus sequences are inherited from an ancestor in which the oocytes or early embryo was infected by an exogenous retrovirus. The viral sequences are then supposed to be subjected to the cellular mutation rate for pseudogenes, although (limited) virus replication and reintegration events will also play a role in determining overall patterns. Probably there is only a small chance of a retrovirus becoming fixed in a certain population in this way.

BaEV offers one of the best opportunities to gain insight in endogenous virus evolution, because intact virus genomes are present in baboons and geladas (suggesting that the viral sequences are not very ancient) and because it has been reported to be widespread in all species of African primates and many species of Asian primates (3, 30) and even in Mediterranean cat species (2), suggesting that the virus is ancient. The viral sequences are supposed to be inherited only from a common ancestor, and analysis of viral genomes obtained from different species could teach us more about sequence variation in an endogenous virus and coevolution of these viruses with their host species and possibly solve the apparent paradox regarding the ages of BaEV sequences in primates. Recently, we have analyzed BaEV proviral structures in the genomic DNA of the baboon (*Papio cynocephalus*) and found no indications for

* Corresponding author. Mailing address: Human Retrovirus Laboratory, Academic Medical Center, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands. Phone: 31 20 566 4522. Fax: 31 20 691 6531.

large deletions in the clones studied (36). The finding that many integrated BaEV genomes in the baboon are probably intact suggests a relatively short history of the virus in baboon species and makes comparison with exogenous viruses most appropriate. Sequence variation within BaEV reverse transcriptase (RT) and *env* gene fragments recovered from different subspecies of baboons and the gelada was found to be low (36). A separate study of mitochondrial 12S rRNA gene variation in 26, mainly African, primate species provided us with a phylogenetic tree of the BaEV host species based upon molecular evidence (37). Amplification of long terminal repeat (LTR), RT gene, and *env* fragments from total DNA of 24 species of African monkeys, one species of Asian macaque, and a domestic cat sample was attempted using BaEV-specific primers, and positive signals were sequenced. Sequences obtained were analyzed by the neighbor-joining (NJ) and maximum-likelihood methods.

MATERIALS AND METHODS

Monkey samples. Samples from 24 species of the following African monkeys were obtained: *Macaca sylvanus* (barbary monkey), *Mandrillus sphinx* (mandrill), *P. cynocephalus cynocephalus* (yellow baboon), *P. cynocephalus anubis* (olive baboon), *P. cynocephalus ursinus* (chacma baboon), *Papio hamadryas hamadryas* (sacred baboon), *Theropithecus gelada* (gelada), *Colobus guereza* (Abyssinian black-and-white colobus), *Cercocebus torquatus atys* (sooty mangabey), *Cercocebus aterrimus* (black mangabey), *Cercocebus galeritus* (Tana mangabey), *Cercopithecus aethiops aethiops* (grivet), *C. aethiops pygerythrus* (vervet), *C. aethiops tantalus* (tantalus monkey), *C. aethiops sabaeus* (green monkey), *Cercopithecus diana roloway* (diana monkey), *Cercopithecus mitis* (blue monkey), *Cercopithecus mona pogonias* (mona monkey), *Cercopithecus patas* (patas monkey), *Cercopithecus ascanius* (redtail monkey), *Cercopithecus nictitans* (spot-nosed guenon), *Cercopithecus cephus* (moustached monkey), *Cercopithecus neglectus* (De Brazza's guenon), and *Miopithecus talapoin* (talapoin monkey). The origin of the samples (serum, plasma, or blood cells) was as published before (37). Spleen tissue from two *C. guereza* monkeys was kindly donated by Vincent Hervé (Institut Pasteur, Bangui, Central African Republic). Sera from Asian pig-tailed macaques (*Macaca nemestrina*) were a gift from the Stichting AAP (Amstelveen, The Netherlands).

A sample of genomic DNA was obtained from a domestic cat (*Felis catus*).

DNA extraction, amplification, and sequencing. Total DNA was extracted from monkey serum, plasma, blood cells, or tissue by a procedure using silica and guanidium thiocyanate (7). PCR amplifications were performed with three sets (LTR1-LTR2, RT1-RT2, and ENV1-ENV4) of BaEV-specific primers (36), and products were cloned with the TA Cloning system from Invitrogen (San Diego, Calif.). Primers LTR1 and LTR2 amplify 467 bp of the BaEV LTR (nucleotides [nt] 63 to 529 and 8015 to 8482 of the published BaEV sequence [19]), primers RT1 and RT2 amplify 327 bp of the BaEV RT gene (nt 3525 to 3851), and primers ENV1 and ENV4 amplify 377 bp of the *env* gene (nt 6235 to 6610).

PCR amplifications were performed using the following protocol: denaturation for 5 min at 95°C and amplification for 35 cycles of 1 min at 95°C, 1 min at 55°C, and 2 min at 72°C, followed by an extension of 10 min at 72°C. The same protocol was used to generate a product from the cat DNA with the RT1-RT2 primer set. The annealing temperature of 55°C was lowered to 40°C in the case of the barbary monkey sample to obtain a PCR product with RT1-RT2. For the talapoin monkey, no amplification was observed with the RT1-RT2 primer set, even at the lower annealing temperature. However, a 91-bp product was generated with the RT primer pair described by Shih et al. (30), in which case the upstream primer is combined with a different downstream primer. This indicates that type C RT sequences can be amplified from talapoin monkey DNA.

At least two clones from single individuals were sequenced in both directions with an Applied Biosystems 373A automated sequencer by following the manufacturer's protocols. A total of 10 RT clones from single *C. ascanius* and *C. guereza* were sequenced to estimate type C RT sequence variation in a species.

A total of 96 RT clones derived from 30 individuals were sequenced and compared, while 61 *env* clones from 20 individuals and 18 LTR clones (derived from 11 individuals) were analyzed. Sometimes PCR artifacts (amplified sequences that clearly did not belong to RT genes) were encountered with the RT1-RT2 primer set, especially in the cat sample and at the lower annealing temperature used for the barbary monkey DNA.

Some nucleotide substitutions observed could have been generated by the *Taq* polymerase used in the PCRs, which has an estimated level of nucleotide misincorporation of approximately 0.1 to 0.2%.

Sequence analysis. Alignment of the sequences was done by using Clustal (16) and corrected by hand. The phylogenetic analyses were done using the NJ method (27), as implemented in the MEGA package (22). Distances were estimated by Kimura's two-parameter method (20), and 100 bootstrap replicates were analyzed. Additional analyses were done using the maximum-likelihood method (program DNAML in the PHYLIP package [12]), with default parameter settings. Gaps introduced for optimal alignment (especially with the LTR sequences) were not considered informative and were not included in the analyses.

Nucleotide sequence accession numbers. GenBank accession numbers for RT sequences used in the analyses were as follows: M26927 (gibbon ape leukemia virus [GALV]), J02255 (Moloney MuLV [MoMLV]), M18247 (feline leukemia virus [FeLV]), and X51929 (*Felis sylvestris* ECE1). The sequences reported in this paper have been deposited in the GenBank database (accession numbers are forthcoming).

TABLE 1. Amplification of African monkey DNA with BaEV-specific primers

Primate species	LTR	RT	*env*
Papio cynocephalus anubis (olive baboon)	+	+	+
Papio cynocephalus cynocephalus (yellow baboon)	+	+	+
Papio hamadryas hamadryas (hamadryas or sacred baboon)	+	+	+
Papio cynocephalus ursinus (chacma baboon)	+	+	+
Theropithecus gelada (gelada)	+	+	+
Mandrillus sphinx (mandrill)	+	+	+
Cercocebus torquatus atys (sooty mangabey)	+	+	+
Cercocebus aterrimus (black mangabey)	+	+	+
Cercocebus galeritus (Tana mangabey)	+	+	+
Macaca sylvanus (barbary macaque)	ND[a]	+	−
Macaca nemestrina (Asian pig-tailed macaque)	ND	+	ND
Cercopithecus aethiops aethiops (grivet)	+	+	+
Cercopithecus aethiops pygerythrus (vervet)	+	+	+
Cercopithecus aethiops tantalus (tantalus monkey)	+	+	+
Cercopithecus aethiops sabaeus (green monkey)	+	+	+
Cercopithecus mitis (blue monkey)	−	+	−
Cercopithecus patas (patas monkey)	−	+	−
Cercopithecus diana roloway (diana monkey)	−	+	−
Cercopithecus mona pogonias (mona monkey)	−	+	−
Cercopithecus ascanius (redtail monkey)	−	+	−
Cercopithecus nictitans (spot-nosed guenon)	−	+	−
Cercopithecus cephus (moustached monkey)	−	+	−
Cercopithecus neglectus (De Brazza's guenon)	−	+	−
Miopithecus talapoin (talapoin monkey)	ND	+[b]	−
Colobus guereza (Abyssinian black-and-white colobus)	−	+	−

[a] ND, not determined.
[b] Only a smaller RT fragment could be amplified.

RESULTS

Amplification and analysis of RT sequences. PCR results obtained with primer set RT1-RT2 are summarized in Table 1 for 25 species of monkeys. In short, correctly sized products were generated with every sample (including domestic cat DNA [not shown]), except for *Miopithecus talapoin*, in which case only smaller, but clearly related, sequences could be amplified with a different downstream primer. Derived amino acid sequences are shown in Fig. 1 for all species sequenced.

Sequencing of RT clones from four *Papio* subspecies showed that there is very little BaEV RT nucleotide heterogeneity in these animals, resulting in extremely conserved protein sequences (Fig. 1). Other monkeys belonging to the *Papionini* tribe are geladas, mandrills, and mangabeys. All these monkey species showed the same level of BaEV nucleotide and amino acid conservation as observed for the *Papio* species (Fig. 1). Defective clones were encountered in a single mandrill. Three clones were found to miss two A nucleotides in a stretch of five. This could, however, have been an error generated once by the *Taq* polymerase and amplified several times. In other clones

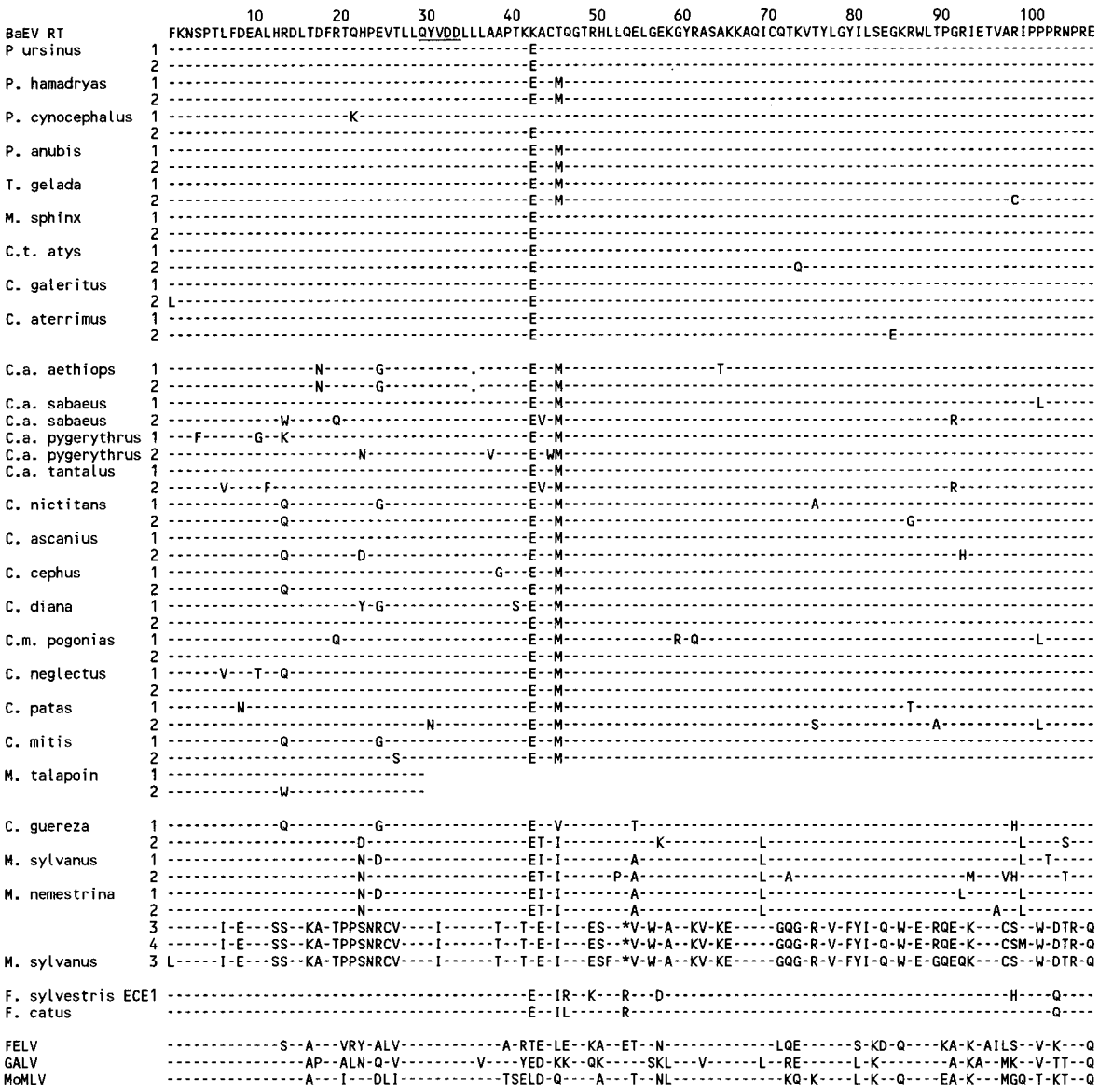

FIG. 1. Alignment of amino acid sequences obtained with BaEV-specific RT primers (RT1 and RT2) from 24 species of African monkeys, an Asian macaque, and a domestic cat. Sequences have been compared with the BaEV sequence of Kato et al. (19), which is shown in the top line. The RT highly conserved QYVDD motif is underlined. Two independent clones per species are shown. For *C. a. sabaeus* and *C. a. pygerythrus* the two sequences were obtained from two different individuals. Sequences taken from GenBank used in the phylogenetic analysis (the exogenous viruses FeLV, GALV, and MoMLV and the endogenous ECE1 sequence from the European wild cat [*F. sylvestris*]) were also aligned. Identical amino acids are indicated by dashes; deletions are indicated by dots. Stop codons present in the nucleotide sequence are indicated with asterisk.

and in other mandrills, this deletion was not seen. In one gelada clone, a single nucleotide deletion was observed.

Thirty-three RT sequences from 12 other species of *Cercopithecus* monkeys, commonly called guenons, including four subspecies of *C. aethiops*, showed more nucleotide variation than RT fragments from the *Papionini* tribe, resulting in random amino acid substitutions (Fig. 1). Ten clones were analyzed from a single redtail monkey (*C. ascanius*) (result not shown) to establish the amount of RT sequence divergence in a single individual. The amount of variation between the 10 clones was the same as the variation between clones from different guenon species.

Although all *Cercopithecus* sequences cluster together in an NJ analysis (Fig. 2), multiple clones from a certain species do not group together (analysis not shown). This finding indicates that slightly different RT genes are randomly present in guenons, in probably more than two copies. Defective clones were also encountered. Four times a sequence contained a 1- or 2-base insertion or deletion, and a larger deletion (19 nt) was present in a *C. patas* clone, while in a *C. aethiops pygerythrus* RT sequence a stop codon had been introduced.

No 327-nt PCR amplification products were generated with talapoin monkey (*Miopithecus talapoin*) serum as a DNA source. However, by changing the downstream primer (see Materials and Methods), a 91-nt fragment could be obtained, indicating that RT sequences homologous to the BaEV sequence are present in this species. Several clones were sequenced, and nucleotide (and amino acid) variation was found

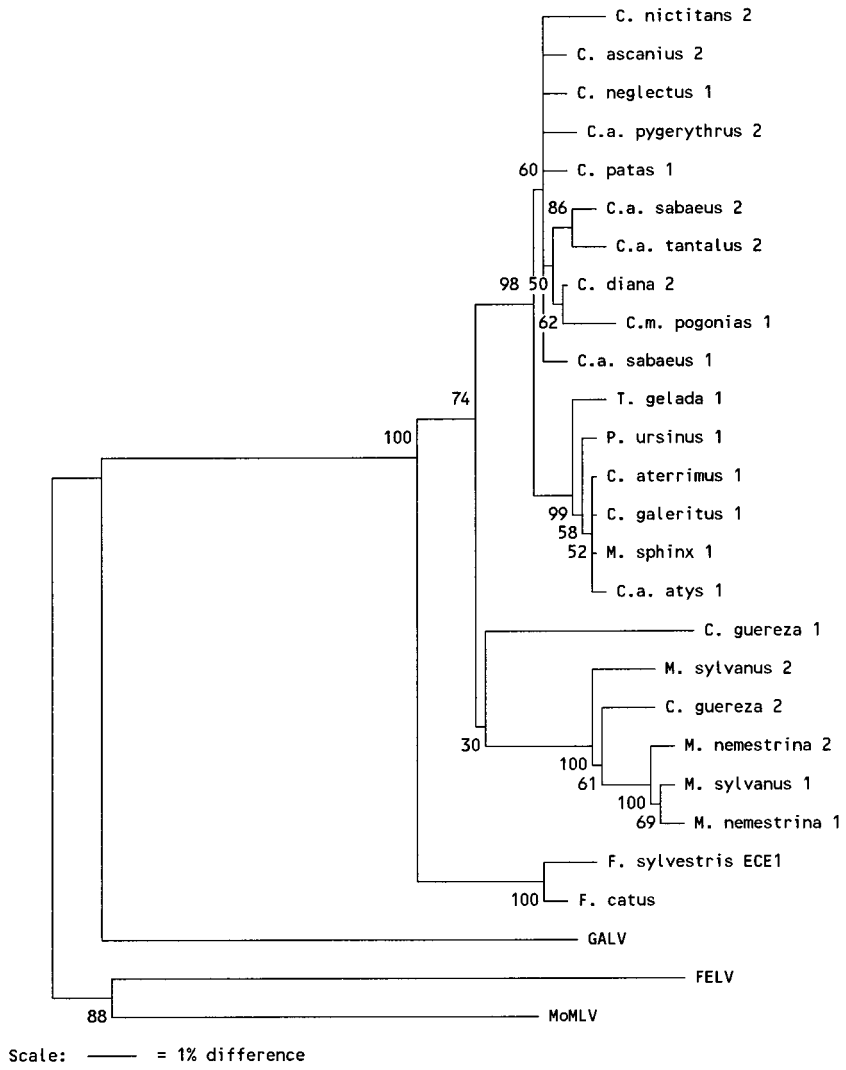

FIG. 2. NJ tree based upon 327-nt RT fragment alignment (Fig. 1) of GALV, MoMLV, FeLV, and BaEV-like RT sequences of different species of African and Asian monkeys and two species of cats. The more divergent RT sequences amplified from macaques have not been incorporated in the analysis. As sequences for the *Papionini* group were in many cases completely identical, only a selected few were analyzed and included in the figure. Because of space limitations, not every *Cercopithecus* sequence is included in the tree. Bootstrap values for 100 replicated trees are indicated.

to be at the same level as in the guenon group, although this estimation is based on a considerably shorter fragment.

Colobine monkeys constitute a distinct cluster in African monkey phylogenetics (37). According to the fossil record, they are probably one of the oldest African primate species existing today. Ten clones from an individual *C. guereza* were analyzed (result not shown); the two most representative of these are shown in Fig. 1. An interesting finding is that 4 of the 10 clones differ at least at the same 11 amino acid positions (and multiple nucleotide positions) from the other 6 clones, suggesting that two distinct type C RT genes are integrated in the colobus monkey genome. Both sequences do not cluster with BaEV in an NJ analysis (Fig. 2), making it more likely that they belong to other viruses. One of these RT sequences (*C. guereza* 2) is homologous to RT sequences amplified from macaques (*M. sylvanus* 1 and 2; *M. nemestrina* 1 and 2) and clusters strongly with them in an NJ analysis (Fig. 2). Two additional clones were found to contain a 1-nt insertion and a 1-nt deletion, respectively, thus destroying the reading frame.

The barbary monkey (*M. sylvanus*) is the only macaque living in Africa today, while all other macaque species inhabit Asia. PCR amplifications with the 327-bp RT fragment primers generated a product with DNA extracted from serum only after the annealing temperature had been lowered to 40°C. This result suggests that type C-like RT sequences in the barbary monkey have considerably diverged from the baboon sequence. Indeed, the nucleotide sequence (result not shown) differed at approximately 40 positions compared to the reference sequence, resulting in 9 to 12 amino acid replacements (Fig. 1). A sequence very homologous to this barbary macaque sequence could be amplified from the Asian *M. nemestrina*, even without lowering the annealing temperature of the PCR. Besides, the RT1-RT2 primer set was able to coamplify a second 327-nt RT-like fragment from both species of macaques (Fig. 1). The second fragments differed in approximately 165 of 327 nucleotide positions from the BaEV RT sequence (≈50% nucleotide homology). Interestingly, in these sequences from both geographically widely separated species, a stop codon is present in an identical position.

BaEV-like sequences have been reported in the domestic cat by using DNA hybridization techniques. Therefore, we used cat (*F. catus*) DNA as input in our PCR. Among several non-

specific fragments, a 327-nt fragment with homology to BaEV was obtained. Translation of the nucleotide sequence showed that most nucleotide substitutions (34 of 39) resulted in silent replacements with respect to the baboon virus gene. This result indicates that there is indeed a gene present in cat genomic DNA with exceptional protein sequence conservation relative to that for BaEV RT. A GenBank database search and NJ analysis of all sequences (Fig. 2) indicated that this domestic cat fragment has high homology to the *F. sylvestris* ECE1 RT sequence, an endogenous virus-derived sequence with no further apparent homology to BaEV.

An NJ analysis was performed with the obtained RT nucleotide sequences. Sequences from defective clones were not used in the phylogenetic analyses. The first objective was to test if all RT sequences do belong to the same virus group. Obtained RT sequences could easily be aligned by hand with corresponding RT fragments of GALV, MoMLV, and FeLV, all exogenous type C retroviruses. Alignment with other retroviruses (endogenous or exogenous) was more difficult. A homologous endogenous RT sequence from *F. sylvestris* which differs by 10 nt (and 4 amino acids) from our domestic cat fragment was taken from GenBank. Figure 2 shows that three obvious endogenous clusters can be recognized, while all exogenous RT sequences are more distantly related (bootstrap value separating the clusters is 100). The BaEV fragments from *Papionini* monkeys and guenons form the first endogenous RT cluster with a high bootstrap value (bootstrap value = 95). These sequences do not form a monophyletic cluster, but *Papionini* sequences are clearly separated (bootstrap value = 100) from the guenon sequences, which form a less stable group with a rather low bootstrap value. Because of the low level of sequence variation between sequences obtained from *Papionini* and guenon monkeys, not every sequence could be included in the analysis. Sequences from macaque and colobus monkeys are more distantly related in a second cluster. One of the two different colobus sequences (*C. guereza* 2) clustered closely with the macaque RTs; the other (*C. guereza* 1) was more distantly related. The less homologous RT fragments obtained from macaque monkeys (*M. nemestrina* 3, *M. nemestrina* 4, and *M. sylvanus* 3) were not included in the analysis shown but were the most distinct sequences in an earlier phylogenetic analysis. The cat sequences form a separate, third cluster (bootstrap value = 100).

Amplification and analysis of *env* sequences. Amplification of BaEV *env* sequences using primer set ENV1-ENV4 was attempted for 24 species of African monkeys. If a species tested negative in this PCR, a nested PCR using primer set ENV2-ENV3 was performed (for a description of the primers, see reference 36). However, every species found negative in the first PCR was also negative in the second PCR. Using input DNA isolated from different individuals of a species did not alter the results. PCR results are summarized in Table 1.

env fragments were obtained for most species of *Papionini* monkeys, except macaques, and the four subspecies of *C. aethiops*. Derived amino acid sequences are shown in Fig. 3. Generally, three clones were sequenced for every species. Defective clones were never observed. For *Mandrillus sphinx* only a single clone was sequenced, as it was difficult to obtain large *env* fragments from this species, although every mandrill tested gave positive results in the nested ENV2-ENV3 PCR. Possibly this difficulty can be attributed to heterogeneity with the primer sequences. In some cases, clones from more than one individual of a (sub)species were sequenced to estimate the amount of sequence variation. In the *Papionini* tribe, sequence divergence (at both the nucleotide and amino acid levels) was very low, while more variation was observed in subspecies of *C. aethiops*, even between individuals of a subspecies. It must be noted that the two monkeys from which sequences *C. a. aethiops* 1 to 3 and 4 to 6 were obtained are probably mother and son, respectively, and show signs of being hybrid animals with *C. a. pygerythrus* characteristics. An NJ analysis of the most representative BaEV *env* sequences is presented in Fig. 4. Because of the low level of nucleotide divergence in this fragment, the tree obtained shows low bootstrap values for almost all clusters, except the one containing mandrill and mangabey sequences (bootstrap value = 93). *env* sequences from other (sub)species did not form clusters with significant bootstrap values. A maximum-likelihood analysis of the *env* sequences gave an identical result (not shown).

Amplification and analysis of LTR sequences. As the amplified RT gene fragments did not prove to be a good determinant for the presence of BaEV and both the RT gene and *env* fragments contain few phylogenetically informative sites, we amplified a third BaEV fragment from African monkey species, consisting of a large part of the viral LTR. Although intact integrated retroviruses contain two LTRs and these should be identical in an active virus, the amount of sequence difference between the two LTR sequences of an endogenous retrovirus may well vary. So, when amplifying LTR sequences, it is not certain whether two clones originate from the same locus in different species or from the same virus in an individual monkey. Still, we estimated that random amplification of BaEV LTR sequences might give phylogenetic information. DNA amplification with primer set LTR1-LTR2 was performed for both *env*-negative and *env*-positive species. PCR results are summarized in Table 1. In short, the *env*-negative species were also negative for BaEV LTR sequences, confirming the absence of the virus, while from the positive species LTR sequences could be amplified without any difficulty. LTR fragments were cloned and sequenced (Fig. 5). There was considerable divergence in sequence length (ranging from 466 to 499 nt) between species. This could mainly be attributed to a difference in the length of a CT repeat and the amplification of *C. a. pygerythrus* LTR sequences containing inserts not observed in other species. Clones amplified from a single individual of a species were very much alike, although they were not identical.

Phylogenetic analysis of LTR sequences by the NJ method (Fig. 6) shows two distinct clusters (bootstrap value separating them = 100), the first consisting of mandrill and mangabey sequences and the second containing all other sequences. The latter cluster is not monophyletic but is separated into two distinct clusters: one containing all *C. aethiops* sequences (bootstrap value = 83) and the other consisting of baboon and gelada LTR sequences (bootstrap value = 86). A maximum-likelihood analysis of the LTR sequences gave an identical result (not shown).

DISCUSSION

It is generally assumed that as long as a microorganism is not pathogenic, there is coevolution between host and pathogen. Endogenous retroviruses are expected to be nonpathogenic, although it is possible that some (unknown) pressure acts upon an intact retrovirus sequence. Nucleotide sequence variation can be quite extensive in exogenous retroviruses (e.g., in the lentivirus group), as in other RNA viruses. This high mutation rate has been attributed to the low fidelity and lack of proofreading of RNA polymerases (11). However, in other classes of retroviruses, sequence variation is less remarkable, e.g., in PTLV-1 and primate foamy viruses, especially between viruses isolated from related species (15, 21, 28). Not much is currently

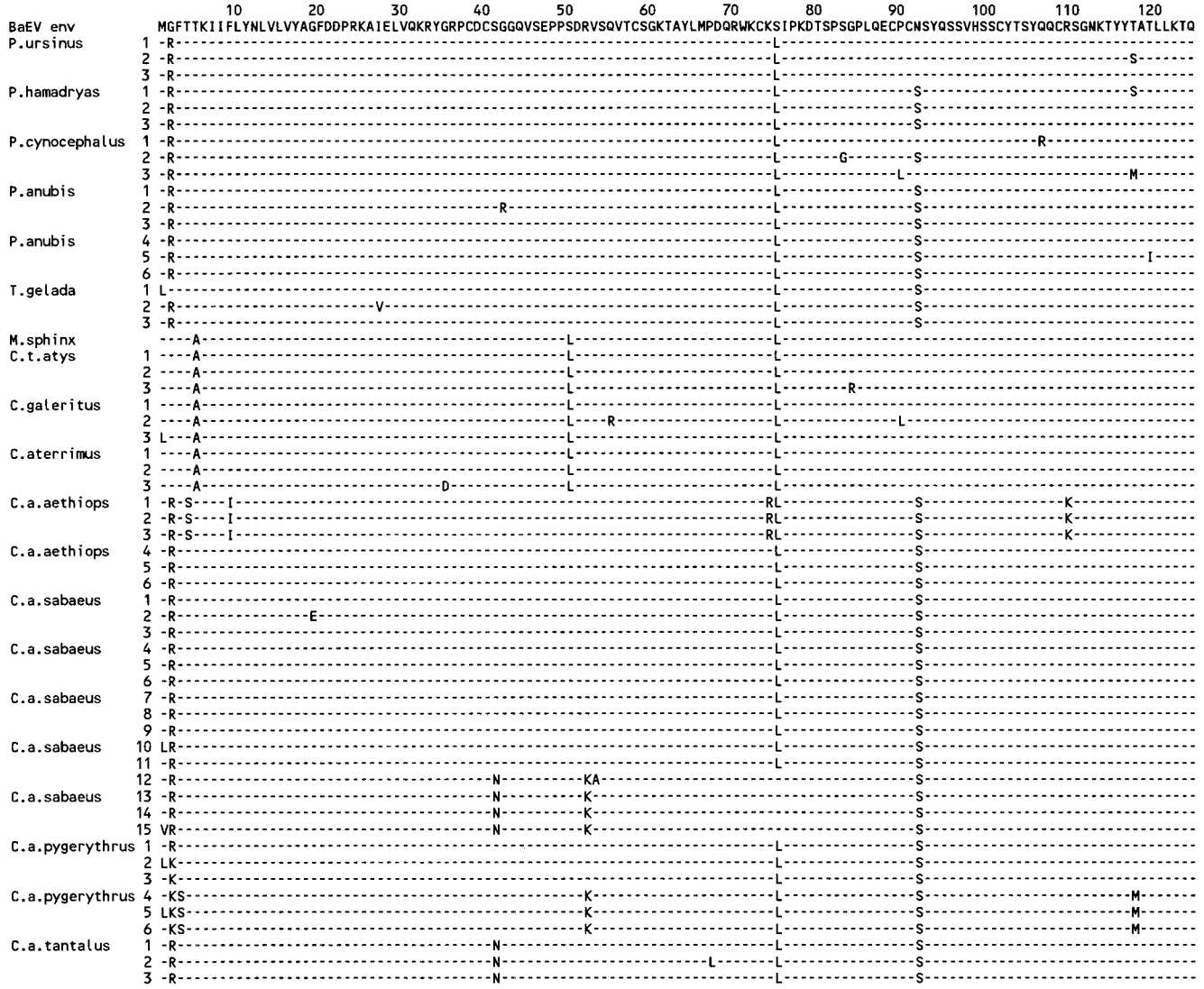

FIG. 3. Alignment of amino acid sequences obtained with BaEV-specific *env* primers (ENV1 and ENV4) from 13 species of African monkeys. Sequences have been compared with the BaEV sequence of Kato et al. (19), which is shown in the top line. Three independent clones per species are shown, except for *Mandrillus sphinx*. For some subspecies of *C. aethiops*, sequences were obtained from more than one individual to investigate sequence variation. Identical amino acids are indicated by dashes.

known about sequence variation of an exogenous retrovirus after fixation in the germ line, e.g., in an endogenous phase. Endogenous retroviral sequences, integrated in the nuclear genome, will be subjected mainly to the cellular replication machinery. Differences in error rates could thus account for variation between endogenous and exogenous sequences. In this light, it is surprising that familial relationships between both classes of viruses can easily be detected.

BaEV is a complete provirus, integrated in the nuclear DNA of Old World monkeys, which can sometimes be reactivated in the baboon and the gelada. In this paper we present the first extensive study on nucleotide variation in the endogenous retrovirus BaEV, both in and between species. Recently, we performed a phylogenetic analysis for African monkey species using sequences of part of the mitochondrial 12S rRNA gene (37), making it possible to test the coevolution hypothesis. Figure 7 shows the mitochondrial DNA-based NJ tree, with species found positive for BaEV LTR, RT gene, and *env* sequences underlined. Interestingly, the viral sequences are not as widespread as was formerly assumed (3, 30) and are probably not inherited from a common ancestor of all African monkeys, as it is difficult to explain why endogenous viral sequences were lost in most species yet are present and extremely conserved in others, although it cannot be excluded that there is sequence heterogeneity with the PCR primers. A similar situation exists for avian endogenous Rous associated virus-0, in which viral sequences do not follow the host phylogeny, suggesting multiple germ line fixations (14).

NJ analysis of an RT gene fragment amplified from African monkeys with primers apparently conserved in type C *pol* genes already indicated that some RT gene fragments do not belong to BaEV, e.g., those from *C. guereza* and *Macaca* species. Also the cat RT gene fragments originate from another virus. The further inability to amplify BaEV LTR and *env* fragments from these species confirmed this conclusion. However, RT gene fragments clustering with the BaEV RT gene could be amplified from all species of *Cercopithecus* monkeys, while corresponding LTR and *env* fragments were obtained

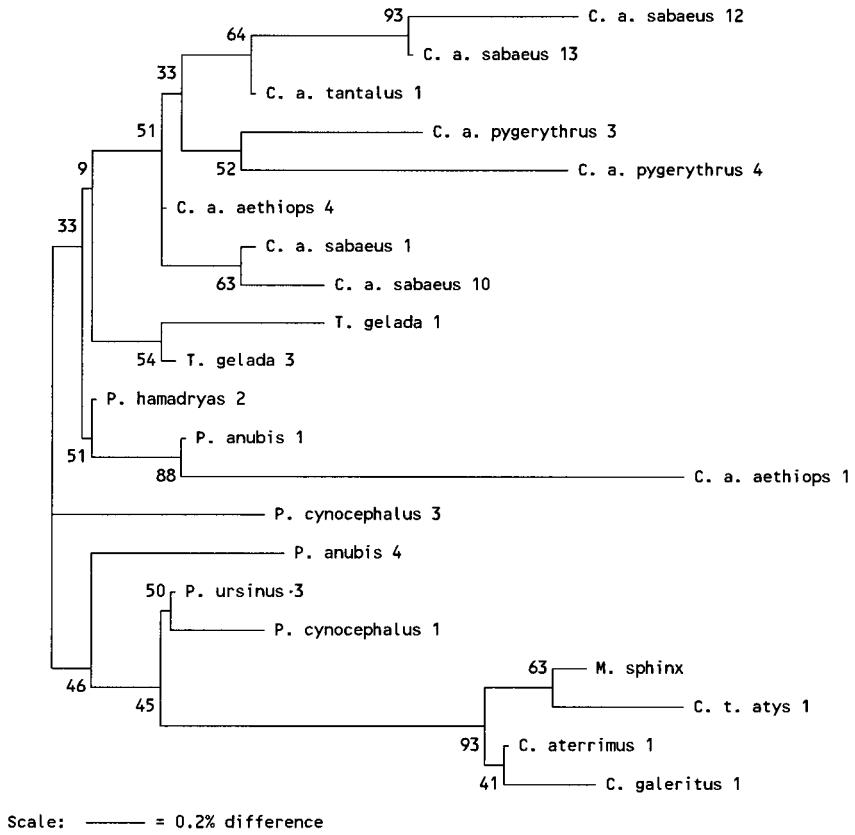

FIG. 4. NJ tree based upon a 377-nt fragment from the BaEV *env* gene sequences shown in Fig. 3. As comparatively few phylogenetically informative sites are present in the sequences obtained, only the most divergent were included in the analyses. Because of the low level of variation, branch lengths do not accurately reflect sequence divergence. Bootstrap values for 100 replicated trees are indicated.

only from *C. aethiops*. It is possible that present in the other *Cercopithecus* species is another virus with similarity to BaEV, e.g., BaEV II, from which only the most conserved genes (like the RT gene) can be amplified with the primer sets used. It must be noted that we were able to generate once with the LTR1-LTR2 primer set a product from *C. patas* which turned out to be not related to BaEV (result not shown).

So, baboon endogenous virus is present as an (probably) intact provirus in *Papionini* monkeys, except the macaques, and in the four subspecies of *C. aethiops*. The phylogenetic analyses of both viral and host DNA fragments (Fig. 2, 4, 6, and 7) indicate that coevolution can only partly explain the results, suggesting that interspecies transmissions have occurred in the past. In this way, the evolutionary history of an endogenous virus is reminiscent of the evolution of exogenous viruses. The viral LTR and *env* trees show that BaEV from mandrills and mangabeys form a separate cluster, while viruses from baboons, geladas, and *C. aethiops* are more related to each other. The host tree and other evolutionary evidence suggests that baboons, geladas, mandrills, and mangabeys should be grouped together, while *C. aethiops* forms a distant cluster together with other *Cercopithecus* species. Besides, according to the mitochondrial sequences and other evidence (references 10 and 37 and references therein), two separate genetic clusters can be distinguished in the *Papionini* group: a baboon (*Papio* species), gelada (*T. gelada*), and black mangabey (*Cercocebus aterrimus*) cluster and a second cluster consisting of mandrills (*Mandrillus sphinx*), sooty mangabeys (*Cercocebus torquatus atys*), and Tana mangabeys (*Cercocebus galeritus*). In contrast, viral sequences group all mangabeys together. A third cluster, albeit more distantly related, is formed by macaque monkeys (*M. sylvanus*), which are supposed to have originated in Africa approximately 6 million years ago, moving out of Africa into Asia much later, as fossil evidence of macaques in Asia is more recent. However, macaque genomes do not contain BaEV sequences, although they probably harbor other endogenous type C viruses. Most likely, all four subspecies of *C. aethiops*, which are estimated to have radiated not longer than 1 million years ago, did inherit the virus from a common ancestor. The same is probably true for the baboon subspecies.

If certain species were infected by BaEV by interspecies transmissions, the virus must have been fixed in the primate germ line more than once. New germ line insertions have often been observed for MuLV, in both laboratory and wild mice (18, 26). It is likely that at least certain types of retroviruses preferentially infect oocytes or the early embryo. This has been shown to be the case for MuLV in mice (24). If the same is true for BaEV, the virus should be able to infect autologous cells. Although BaEV has been described as a xenotropic virus, there are reports showing that it is able to infect baboon cells, albeit with lower efficiency and with some sort of restriction (23). This could, however, have been different in the past or in other species.

It is unlikely that a certain environmental pressure is acting upon the viral sequences, as the monkeys found positive are widely distributed all over Africa, inhabiting both savannas (baboons, geladas, and African green monkeys) and forests

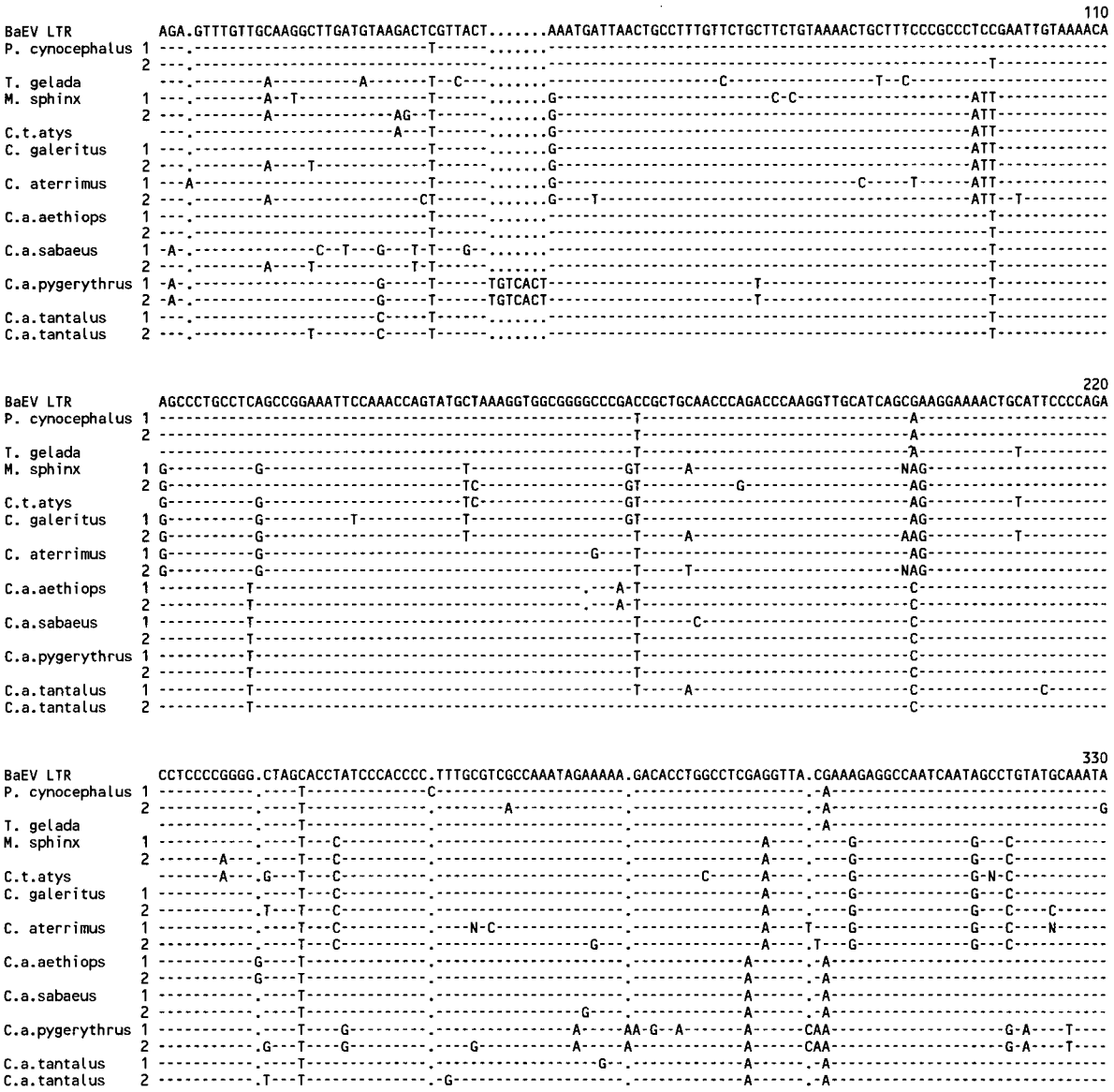

FIG. 5. Alignment of nucleotide sequences obtained with BaEV-specific LTR primers (LTR1 and LTR2) from 10 species of African monkeys. Sequences have been compared with the BaEV sequence of Kato et al. (19), which is shown in the top line. Sequences from *C. a. tantalus* were obtained from two different individuals. Identical nucleotides are indicated by dashes; deletions are indicated by dots. Alignment of the LTR sequences is sometimes controversial, especially in repeat regions.

(mangabeys and mandrills). However, the viral sequences can be divided into two clusters corresponding to the habitats of their hosts, providing evidence for cross-species transmissions in shared habitats. At this time many monkey (sub)species concerned are geographically separated, although subspecies of baboons and African green monkeys can be found sharing the same ranges and it is possible that there is contact between species along geographical boundaries. Monkey species and habitats were probably more widespread in the past, suggesting that at that moment, geography and ecology were better connected.

We name the BaEV strain found in baboons, geladas, and African green monkeys $BaEV_{savanna}$ ($BaEV_{sav}$), and the strain of mandrills and mangabeys is called $BaEV_{forest}$ ($BaEV_{for}$); it is most likely that both strains were already separated from each other during infection of the primate species. A possible common ancestor ($BaEV_{ancestor}$) of the two virus strains has not been detected so far. It is possible that the original host was not a primate but that BaEV has been acquired by one or more interspecies transmissions from other animals to monkeys. It is also possible that other primate species were infected by BaEV in the past but that the virus did not become endogenous in these species. More information about the history of BaEV could be obtained by analyzing (common) integration sites and copy numbers in different monkey species. The separation between $BaEV_{sav}$ and $BaEV_{for}$ is estimated to have occurred between approximately 24,000 and 400,000 years ago when we use evolution rates calculated earlier for simian T-cell leukemia virus type 1 and HTLV-1 (retroviruses with comparable ages, hosts, and interstrain nucleotide differences), which range from 0.4×10^{-7} to 6.8×10^{-7} substitution per site per year (31), and the percent nucleotide difference observed be-

```
                                                                                                                             440
BaEV LTR          AAGCCTAAACGCTATAAAAGAGATGTATCGCCACAAATCGGGGCTCTC......TCTCTCTCTACCACATTCCTT...............AG.AGGGAGGGCCCTGGTGC
P. cynocephalus 1 ------------------------------------------------......----......--T--------C...C.........AG-.T----------------
                2 ---------------------------------------A--------......------....--T--------C...C.........A--.T-A--------------
T. gelada         ---------T--------------------------------------......--------..--T--------T...C.........A--.T-A--------------
M. sphinx       1 ------------------------------------G-----------.CT....--------CT-T-TC-A-C-CATTCCT.....TTG--GT----------------
                2 ------------------------------------G-----------TC.....--------CT-T-TC---C-CATTCCT.....TTG--GT----------------
C.t.atys          ----------------------------------G-G-----------TCTC.C---------CT-T-TC---C-CATTCCT.....TTG--GT----------------
C. galeritus    1 ------------------------------------G-----------.........------CT-T-TC-A-C-CATTCCT.....TTG--GT----------------
                2 ------------------------------------G-----------..........-----CT-T-TC-A-C-CATTCCT.....TTG--GT----------------
C. aterrimus    1 ------------------------------------G-----------..............-CT-T-TC-A-C-CATTCCT.....TAG--GT--------------T-
                2 ------------------------------------G-----------..............-CT-T-TC-A-C-CATTCCT.....TAG--GT--------------T-
C.a.aethiops    1 ----------A----------------A--------------------......--........--T--------CTTGC.........GG-GT----------------
                2 ----------A----------------A--------------------......--........--T--------CTTGC.........GG-GT----------------
C.a.sabaeus     1 ---------------------------A--------------------......----......--T--------CTTGC.........G--GT----------------
                2 ----------------------------T-------------------......--........--T------A-CTTGC.........G--GT----------------
C.a.pygerythrus 1 --------T------------------A--------------------TCTCTC--------..--T-G------CTTGCGAGGTGAAGG-AT--A-G------------
                2 --------T------------------A--------------------TCTCTC---------C--T-G------CTTGCGAGGTGAAGG-AT--A-G------------
C.a.tantalus    1 ---------------------------A--------------------......----......--T--------CTTGC.........G--GT----------------
C.a.tantalus    2 ---------------------------A--------------------......------....--T--------CTTGC.........G--GT----------------
```

```
                                                                             502
BaEV LTR          ACCAGTAAACGACTTTCTGCCGAAATTTGTGTGTTGGGTGGTTCCTCGCGCCGACTCT.AAA
P. cynocephalus 1 ---------T------------------------------------------------C---
                2 ---------T------------------------------------------------C---
T. gelada         ---------T----------------C-------------------------------C--C
M. sphinx       1 -------------------------------------A--------------------C---
                2 -------------------------------------A--------------------C---
C.t.atys          ----------A--------------------------A--------------------C---
C. galeritus    1 -------------------------------------A--------------------C---
                2 --------------------------C----------A--------------------C---
C. aterrimus    1 ------------------A------A-------A---A--------------------C---
                2 -------------------------------------A--------------------C---
C.a.aethiops    1 ---------T------------------------------------------------C---
                2 ---------T------------------------------------------------C---
C.a. sabaeus    1 ---------T----------------------T--A----------------------C---
                2 ---------A-------------------------A---A------------A-----C---
C.a.pygerythrus 1 ------------------T------------------------------A--------C---
                2 ------------------T----------------------------C----------C---
C.a.tantalus    1 ---------T------------------------------------------------C---
C.a.tantalus    2 ---------T------------------------------------------------C---
```

FIG. 5—*Continued.*

tween *env* gene fragments of the two BaEV strains. Although the primate fossil record is not very clear over this period, it is generally assumed that by that time all extant primate species have radiated (although subspecies may still have been evolving), again suggesting that the primate species have been independently infected with BaEV, including the two main clusters of mangabeys.

Sooty mangabeys and mandrills apparently do not cross the Dahomey Gap (Republic of Benin), a large area of dry savanna vegetation in the West African lowland forest, suggesting that

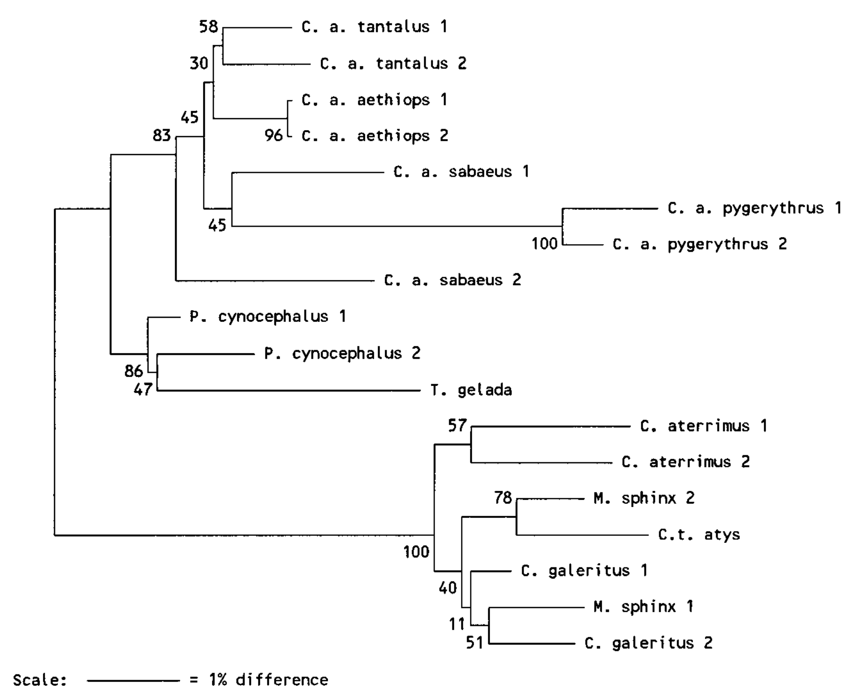

FIG. 6. NJ tree based upon an approximately 467-nt fragment from the BaEV LTR sequences aligned in Fig. 5. All sequences shown in the alignment were analyzed. Bootstrap values for 100 replicated trees are indicated.

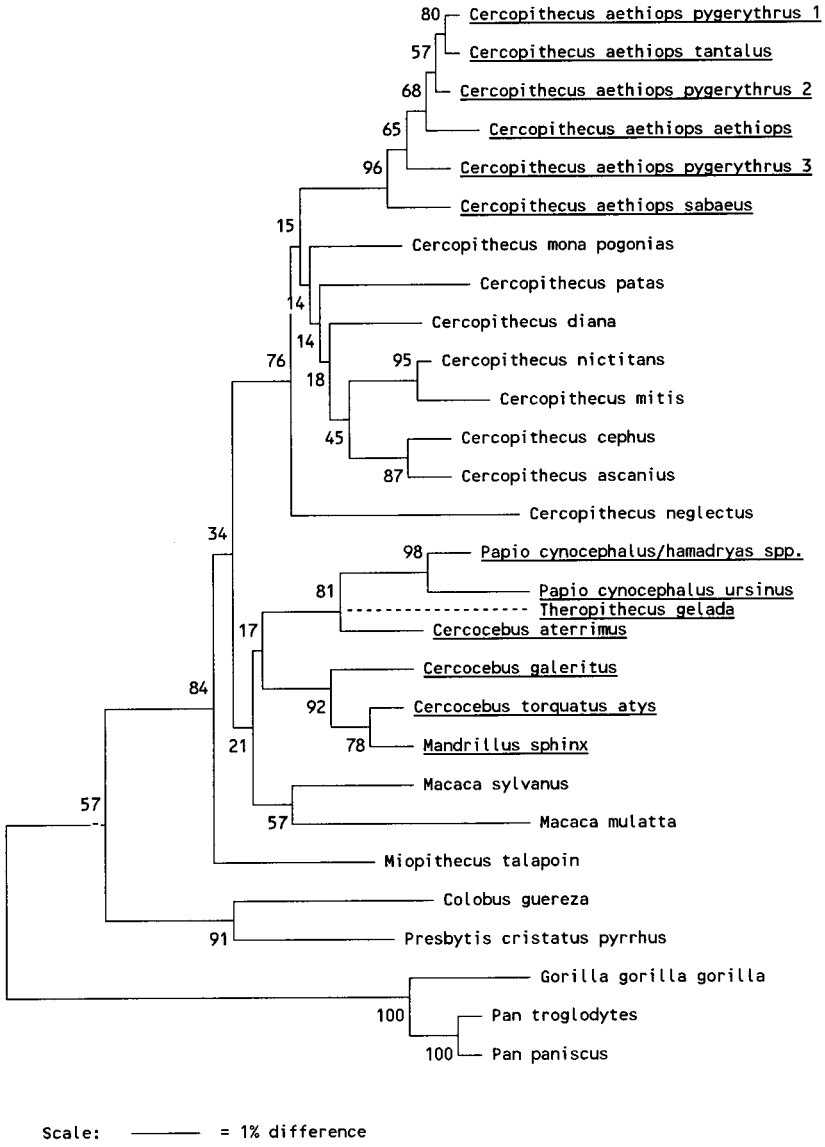

FIG. 7. NJ tree based upon a 386- to 393-nt fragment of the mitochondrial 12S rRNA gene of 24 species of African primates and 2 species of Asian monkeys (37). As sequences derived for groups of vervet (*C. a. pygerythrus*) monkeys acquired from different locations differed substantially, several were included in the analysis. The African apes (chimpanzees [*Pan troglodytes* and *Pan paniscus*] and gorilla [*Gorilla gorilla gorilla*]) were designated outgroups in this analysis. Underlined are species which are positive for BaEV LTR, RT gene, and *env* sequences by both PCR and sequence analysis. The putative phylogenetic placement of *T. gelada*, which was also positive for BaEV, was taken from another NJ tree based upon mitochondrial DNA analysis (10), as this species was originally not included in the tree shown (branch indicated by a dashed line). Bootstrap values (node reproduction frequencies out of 100 trees) are represented from bootstrap analysis. *Presbytis cristatus pyrrhus* is the Asian silvered leaf monkey, which belongs to the family *Colobinae*, and *Macaca mulatta* is the Asian rhesus macaque.

this gap acts as a natural geographic barrier separating the two species. Eventual transmission of BaEV between mandrills and sooty mangabeys should then have occurred before the establishment of this zone. Other monkey species, however, are able to cross the Dahomey Gap.

The presence of an intact endogenous virus in certain species of African monkeys raises questions with regard to its function. It has been speculated that expression of endogenous virus genes protects the individual from superinfection with a related exogenous virus, as has been observed for the *Fv-4* resistance gene in mice (17). Another possibility is that it protects the germ line from reinfection with the same endogenous virus and thus from provirus-associated mutations. In this light, it is interesting to note that expression of endogenous virus genes has often been observed in placental tissue (1, 6).

In conclusion, evidence for coevolution of BaEV with its primate host is weak, since many primate species lack BaEV altogether and, in the positive primate species, BaEV sequences cluster not according to host speciation but according to habitat.

ACKNOWLEDGMENTS

We thank Carla Kuiken for help with the preparation of the final version of the manuscript.

This study is funded in part by the Institute of Virus Evolution and the Environment.

REFERENCES

1. **Benveniste, R. E., M. M. Lieber, D. M. Livingston, C. J. Sherr, and G. J. Todaro.** 1974. Infectious C-type virus isolated from a baboon placenta. Nature (London) **248:**17–20.

2. **Benveniste, R. E., and G. J. Todaro.** 1974. Evolution of C-type viral genes: inheritance of exogenously acquired viral genes. Nature (London) **252:**456–459.
3. **Benveniste, R. E., and G. J. Todaro.** 1976. Evolution of type C viral genes: evidence for an Asian origin of man. Nature (London) **261:**101–108.
4. **Benveniste, R. E., and G. J. Todaro.** 1977. Evolution of primate oncornaviruses: an endogenous virus from langurs (*Presbytis* spp.) with related virogene sequences in other Old World monkeys. Proc. Natl. Acad. Sci. USA **74:**4557–4571.
5. **Blattner, W. A.** 1989. Retroviruses in viral infection of human epidemiology and control, p. 545–592. *In* A. S. Evans (ed.), Retroviruses. Plenum Press, New York.
6. **Boller, K., H. König, M. Sauter, N. Mueller-Lantzsch, R. Löwer, J. Löwer, and R. Kurth.** 1993. Evidence that HERV-K is the endogenous retrovirus sequence that codes for the human teratocarcinoma-derived retrovirus HTDV. Virology **196:**349–353.
7. **Boom, R., C. J. A. Sol, M. M. M. Salimans, C. L. Jansen, P. M. E. Wertheim-van Dillen, and J. van der Noordaa.** 1990. Rapid and simple method for the purification of nucleic acids. J. Clin. Microbiol. **28:**495–503.
8. **Coffin, J. M.** 1992. Structure and classification of retroviruses, p. 19–49. *In* J. A. Levy (ed.), The Retroviridae, vol. I. Plenum Press, New York.
9. **Crittenden, L. B.** 1981. Exogenous and endogenous leukosis virus genes—a review. Avian Pathol. **10:**101–112.
10. **Disotell, T. R., R. L. Honeycutt, and M. Ruvolo.** 1992. Mitochondrial DNA phylogeny of the Old-World monkey tribe Papionini. Mol. Biol. Evol. **9:**1–13.
11. **Domingo, E., and J. J. Holland.** 1988. High error rates, population equilibrium and evolution of RNA replication systems, p. 3–36. *In* E. Domingo, J. J. Holland, and P. Ahlquist (ed.), RNA genetics. CRC Press, Inc., Boca Raton, Fla.
12. **Felsenstein, J.** 1993. PHYLIP manual version 3.5. University Herbarium of the University of California, Berkeley.
13. **Fischinger, P. J., P. T. Peebles, S. Nomura, and D. K. Haapala.** 1973. Isolation of an RD-114-like oncornavirus from a cat cell line. J. Virol. **11:**978–985.
14. **Frisby, D. P., R. A. Weiss, M. Roussel, and D. Stehelin.** 1979. The distribution of endogenous chicken retrovirus sequences in the DNA of galliform birds does not coincide with avian phylogenetic relationships. Cell **17:**623–634.
15. **Herchenröder, O., R. Renne, D. Loncar, E. K. Cobb, K. K. Murthy, J. Schneider, A. Mergia, and P. A. Luciw.** 1994. Isolation, cloning, and sequencing of simian foamy viruses from chimpanzees (SFVcpz): high homology to human foamy virus (HFV). Virology **201:**187–199.
16. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene **73:**237–244.
17. **Ikeda, H., and H. Sugimura.** 1989. *Fv-4* resistance gene: a truncated endogenous murine leukemia virus with ecotropic interference properties. J. Virol. **63:**5405–5412.
18. **Inaguma, Y., N. Miyashita, K. Moriwaki, W. C. Huai, J. Mei-Lei, H. Xinqiao, and H. Ikeda.** 1991. Acquisition of two endogenous ecotropic murine leukemia viruses in distinct Asian wild mouse populations. J. Virol. **65:**1796–1802.
19. **Kato, S., K. Matsuo, N. Nishimura, N. Takahashi, and T. Takano.** 1987. The entire nucleotide sequence of baboon endogenous virus DNA: a chimeric genome structure of murine type C and simian type D retroviruses. Jpn. J. Genet. **62:**127–137.
20. **Kimura, M. A.** 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. **16:**111–120.
21. **Koralnik, I. J., E. Boeri, W. C. Saxinger, A. Lo Monico, J. Fullen, A. Gessain, H.-G. Guo, R. C. Gallo, P. Markham, V. Kalyanaraman, V. Hirsch, J. Allan, K. Murthy, P. Alford, J. Pecon Slattery, S. J. O'Brien, and G. Franchini.** 1994. Phylogenetic analysis of human and simian T-cell leukemia/lymphotropic virus type I strains: evidence for interspecies transmissions. J. Virol. **68:**2693–2707.
22. **Kumar, S., K. Tamura, and M. Nei.** 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. The Pennsylvania State University, University Park.
23. **Lavelle, G., L. Foote, R. L. Heberling, and S. S. Kalter.** 1979. Expression of baboon endogenous virus in exogenously infected baboon cells. J. Virol. **30:**390–393.
24. **Lock, L. F., E. Keshet, D. J. Gilbert, N. A. Jenkins, and N. G. Copeland.** 1988. Studies of the mechanism of spontaneous germline ecotropic provirus acquisition in mice. EMBO J. **7:**4169–4177.
25. **Rabin, H., C. V. Benton, M. A. Tainsky, N. R. Rice, and R. V. Gilden.** 1979. Isolation and characterization of an endogenous type C virus of rhesus monkeys. Science **204:**841–842.
26. **Rowe, W. P., and C. A. Kozak.** 1980. Germ-line reinsertions of AKR murine leukemia virus genomes in *Akv-1* congenic mice. Proc. Natl. Acad. Sci. USA **77:**4871–4874.
27. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
28. **Saksena, N. K., V. Herve, J. P. Durand, B. Leguenno, O. M. Diop, J. P. Digoutte, C. Mathiot, M. C. Muller, J. L. Love, S. Dube, M. P. Sherman, P. M. Benz, S. Erensoy, A. Galat-Luong, G. Galat, B. Paul, D. K. Dube, F. Barre-Sinoussi, and B. J. Poiesz.** 1994. Seroepidemiologic, molecular, and phylogenetic analyses of simian T-cell leukemia viruses (STLV-I) from various naturally infected monkey species from central and western Africa. Virology **198:**297–310.
29. **Sherwin, S. A., and G. J. Todaro.** 1979. A new endogenous primate type C virus isolated from the Old World monkey *Colobus polykomos*. Proc. Natl. Acad. Sci. USA **76:**5041–5045.
30. **Shih, A., E. E. Coutavas, and M. G. Rush.** 1991. Evolutionary implications of primate endogenous retroviruses. Virology **182:**495–502.
31. **Song, K.-J., V. R. Nerurkar, N. Saitou, A. Lazo, J. R. Blakeslee, I. Miyoshi, and R. Yanagihara.** 1994. Genetic analysis and molecular phylogeny of simian T-cell lymphotropic virus type I: evidence for independent virus evolution in Asia and Africa. Virology **199:**56–66.
32. **Stoye, J. P., and J. M. Coffin.** 1987. The four classes of endogenous murine leukemia virus: structural relationships and potential for recombination. J. Virol. **61:**2659–2669.
33. **Todaro, G. J., R. E. Benveniste, S. A. Sherwin, and C. J. Sherr.** 1978. MAC-1, a new genetically transmitted type C virus of primates: "low frequency" activation from stumptail monkey cell cultures. Cell **13:**775–782.
34. **Todaro, G. J., C. J. Sherr, and R. E. Benveniste.** 1976. Baboons and their close relatives are unusual among primates in their ability to release nondefective endogenous type C viruses. Virology **72:**278–282.
35. **York, D. F., R. Vigne, D. W. Verwoerd, and G. Querat.** 1992. Nucleotide sequence of the Jaagsiekte retrovirus, an exogenous and endogenous type D and B retrovirus of sheep and goats. J. Virol. **66:**4930–4939.
36. **van der Kuyl, A. C., J. T. Dekker, and J. Goudsmit.** 1995. Full-length proviruses of baboon endogenous virus (BaEV) and dispersed BaEV reverse transcriptase retroelements in the genome of baboon species. J. Virol. **69:**5917–5924.
37. **van der Kuyl, A. C., C. L. Kuiken, J. T. Dekker, and J. Goudsmit.** 1995. Phylogeny of African monkeys based upon the mitochondrial 12S rRNA gene. J. Mol. Evol. **40:**173–180.