JOURNAL OF VIROLOGY, Sept. 1995, p. 5917–5924 Vol. 69, No. 9
0022-538X/95/$04.00+0

Full-Length Proviruses of Baboon Endogenous Virus (BaEV) and Dispersed BaEV Reverse Transcriptase Retroelements in the Genome of Baboon Species

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

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

Received 6 March 1995/Accepted 19 June 1995

Baboon endogenous virus (BaEV) is present in multiple copies in many Old World monkey species. BaEV proviruses may contain open reading frames for all major genes, as is indicated by the rescue of infectious virus particles from baboon and gelada tissues after cocultivation with permissive cells. We have analyzed full-length BaEV proviral structures in a baboon (*Papio cynocephalus*) genomic library and found no evidence for the rearrangements or large deletions commonly observed in endogenous virus genomes from other mammalian species. The two proviruses studied were integrated next to or nearby long interspersed repeat sequence (LINE) transposable elements. Additionally, isolated dispersed fragments with 100% and approximately 77% homology, respectively, to part of the BaEV reverse transcriptase gene were detected. These presumed retroelements were present in an ~10-fold excess compared with the full-length proviral genomes. PCR amplification and sequencing of BaEV reverse transcriptase and *env* fragments from the lambda clones and from the genomic DNA of other baboon species showed that there is little sequence variation present in BaEV DNA in the baboon genome.

A substantial part (estimated to be 0.1 to 0.6% in humans) of the eukaryotic genome consists of retroviral and retrovirus-related sequences. Several of these sequences are results of ancient retroviral infections and are inherited as stable Mendelian genes. In contrast to horizontally transmitted exogenous retroviruses, endogenous retroviral elements are present in the genome of every cell of all individuals of a species. Copy numbers, however, sometimes vary between individuals, as was observed for endogenous feline leukemia virus (FeLV)-related sequences in domestic cats and endogenous murine leukemia virus (MuLV) in mice (16, 34). Also, polymorphism of different classes of MuLV proviruses has been detected in mouse strains (37). Occasionally, endogenous viral particles have been observed in primates as well as in rodents (reference 7 and references therein). A possible involvement of endogenous viruses in cancer pathogenesis has been implicated (for a review, see reference 18). Besides, endogenous viruses can act as sources with which exogenous retroviruses can exchange sequence information. In that way, endogenous viruses could contribute substantially to outbreaks of exogenous viral activity, which are observed from time to time in vertebrates.

Type C endogenous oncoviruses have been reported in rodents, reptiles, birds, pigs, cats, and primates (reference 29 and references therein). For primates, four endogenous type C viruses have been isolated from Old World monkeys so far. Two endogenous viruses originate from macaques: MAC-1 from the stump-tailed macaque (*Macaca arctoides*) and MMC-1 from the rhesus monkey (*Macaca mulatta*) (27, 38). MAC-1 and MMC-1 are closely related, and are possibly different isolates of the same virus. A third type C virus, CPC-1, was isolated from a *Colobus polykomos* (30). DNA hybridization showed that CPC-1 sequences can be detected in other Old World monkeys and are present in approximately 50 to 70 copies per diploid genome. The fourth and best-characterized type C virus is baboon endogenous virus (BaEV), which was originally isolated from baboon tissue by cocultivation with permissive cell lines (2, 40). However, BaEV appears to be chimeric, containing a type D *env* gene (17). Normally, BaEV is not expressed in baboons, although viral particles and RNA can be detected in placental tissue (2, 39). The complete nucleotide sequence of the M7 baboon isolate of BaEV has been determined (17). DNA hybridization, Southern blotting, and PCR amplification have shown that BaEV is present in many Old World monkey species (1, 4, 8, 31). Although BaEV genomes are estimated to be present in 50 to 100 copies per cell (31), many of the integrated genomes are probably defective. The baboon embryo fibroblast cell strain BEF-3, enriched in BaEV sequences, releases infectious BaEV particles upon transfection of permissive cells. BEF-3 provirus cloning showed that only one out of four clones was nondefective (13). Restriction-site heterogeneity was observed between different BaEV genomes (11, 13). PCR amplification of a small fragment of the BaEV reverse transcriptase (RT) gene showed that limited sequence variation both in and between species exists (31).

There is no conclusive evidence showing that BaEV integrations in humans and apes exist. A complete BaEV probe gave only a low level of hybridization with human, gibbon, and orangutan genomic DNA (6). More significant values were reached with chimpanzee and gorilla DNA. Using PCR, Shih et al. (31) were not able to detect BaEV RT fragments in human DNA, but they did detect amplification products with chimpanzee and gorilla DNA. Human ERV-1 and ERV-3 sequences have moderate homology with type C viruses, including BaEV, but ERV-1 and ERV-3 are defective single-copy sequences located on chromosomes 18 and 7, respectively (12). BaEV can also not be detected in New World monkeys (1), suggesting that the virus entered the germ line after the Old World-New World monkey split (approximately 30 to 40 million years ago). In cat species originating from the Mediterranean basin, an endogenous virus (RD114) with strong homology to BaEV was detected by DNA hybridization (3). This finding could be explained by the possible infection of a Mediterranean cat ancestor with the primate virus, as cat spe-

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

TABLE 1. Primers used to amplify BaEV sequences

BaEV Primer	Position (nt)[a]	Sequence
LTR1 upstream	38 to 62 and 7990 to 8014	5′ GGTTCTGTTTGTTCGTGGCGCAGGC 3′
LTR2 downstream	530 to 555 and 8483 to 8507	5′ GATGTTAGAATCACGTTCCTTAGGG 3′
RT1 upstream	3505 to 3524	5′ TGGACTCGACTTCCCCAGGG 3′
RT2 downstream	3852 to 3871	5′ CAGTTCCCAGGAATTCACGC 3′
ENV1 upstream	6210 to 6234	5′ GCAGTAACCCCTCATATAGTCCAAG 3′
ENV2 upstream[b]	6282 to 6301	5′ CTACGCGGGGTTTGACGACC 3′
ENV3 downstream[b]	6469 to 6488	5′ GGCCCGCTTGGGGAGGTGTC 3′
ENV4 downstream	6611 to 6634	5′ CTTGTACATCACTGGTGCCCCCAG 3′

[a] Nucleotide numbering according to the published BaEV sequence of Kato et al. (17).
[b] *env* primers ENV2 and ENV3 amplify a 167-bp fragment located inside the 377-bp fragment generated by the primer set ENV1/ENV4.

cies from the New World and Asia and the larger African cats lack homologous virus sequences.

Little is known about the structure and evolution of BaEV in the baboon genome. For the related type C endogenous retroviruses RD114, FeLV, and MuLV, different evolutionary patterns do exist. RD114, present in approximately 20 copies in the cellular genome of the domestic cat, has a relatively conserved *gag-pol* region but has many substitutions and deletions in *env* (28). Another pattern is present in endogenous FeLV, in which integrated copies have large deletions in *gag-pol* and shorter ones in the *env* gene, compared with exogenous FeLV (34). Many endogenous MuLV proviruses have lost all or almost all of their *env* genes, while others contain deletions in *gag* or *pol* or both (36). The present study was designed to gain insight into the molecular genetic organization of BaEV in the baboon genome. A baboon genomic library in the lambda DASH^{R}II vector, constructed from kidney tissue of a normal 18-year-old male *Papio cynocephalus*, was obtained from Stratagene (La Jolla, Calif.) and screened with probes homologous to BaEV long terminal repeat (LTR), RT, and *env* sequences. The products of PCR amplifications of *Papio ursinus* DNA with appropriate sets (LTR1/LTR2, RT1/RT2, and ENV1/ENV4) of BaEV-specific primers (Table 1) were cloned with the TA cloning system from Invitrogen (San Diego, Calif.), and these plasmids were used to generate ^{32}P-labelled specific probes. LTR1/LTR2 amplifies 467 bp of the BaEV LTR, RT1/RT2 amplifies 327 bp of the BaEV RT gene, and ENV1/ENV4 amplifies 377 bp of the *env* gene.

Serum samples were obtained from the following baboon species: *Papio anubis* (olive baboon), *P. cynocephalus* (yellow baboon), and *Papio hamadryas* (hamadryas or sacred baboon). Additionally, blood cells were obtained from *P. ursinus* (chacma baboon). The origins of the samples were as published before (41). DNA from gelada baboons (*Theropithecus gelada*) was a gift from Todd Disotell (University of New York, New York). Total DNA was extracted from the samples by a procedure with silica and guanidinium thiocyanate (9), and RT and *env* fragments were amplified from the DNA (with primer sets RT1/RT2 and ENV1/ENV4), cloned into the TA vector, and sequenced. All PCR amplifications were performed according to the following protocol: denaturation, 5 min 95°C; amplification, 35 cycles of 1 min 95°C, 1 min 55°C, 2 min 72°C; extension, 10 min 72°C. Sequencing was done with an Applied Biosystems 373A automated sequencer, following the manufacturer's protocols.

Initially, approximately 36,000 plaques of the baboon library were screened with a BaEV *env* probe. Five positive plaques were obtained (named 12.1, 12.2, 23.1, 25.1, and 25.2) and purified. After a second screening with BaEV LTR and BaEV RT probes, with duplicate filters containing approximately 26,000 plaques, 10 additional clones were isolated. Only 1 plaque (31.1) of these 10 reacted with both probes; the other 9 hybridized only to the RT probe. Of these 9, 4 (30.1, 33.1, 36.1, and 37.1) were (partly) characterized. Lambda DNA was isolated from purified positive plaques with the Wizard Lambda Preps DNA purification system from Promega (Madison, Wis.). Inserts were first characterized by PCR with BaEV-specific primer sets for the LTR, RT, and *env* gene. Results of the PCRs are shown in Table 2. For three clones (12.1, 12.2, and 31.1), correct PCR fragments were obtained with all three sets (LTR, RT, and the *env* outer primer set ENV1/ENV4), while one clone (25.1) was positive for LTR, RT, and the ENV1/ENV3 primer set. This suggested that clones 12.1, 12.2, and 31.1 possibly contained full-length viruses, although the presence of only one LTR can be confirmed by PCR. A recognition site for the enzyme *Sau*3A (used to construct the genomic library) is present between primers ENV3 and ENV4 in the reference sequence, which could explain the PCR result for clone 25.1, assuming it to be an artificially truncated viral sequence. The enzyme *Not*I was used to determine the size of the lambda clone insert, which was found to be approximately 16 to 22 kb in all clones. Subsequently, the clones were characterized by digestion with the restriction enzymes *Bam*HI, *Eco*RI, *Not*I, *Stu*I, and *Xho*I and several others (all from Boehringer GmbH, Mannheim, Germany); separation on 0.5% agarose gels; Southern blotting (with 20× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate] and positively charged nylon membranes [Boehringer]); and hybridization with biotin-labelled LTR, RT, and *env* probes, with the nonradioactive

TABLE 2. PCR results for lambda clone inserts

Clone	Probe[a]	Primer set[b]		
		LTR	RT	*env*
12.1	*env*	LTR1/LTR2	RT1/RT2	ENV1/ENV4[c]
12.2	*env*	LTR1/LTR2	RT1/RT2	ENV1/ENV4[c]
25.1	*env*	LTR1/LTR2	RT1/RT2	ENV1/ENV3
31.1	RT/LTR	LTR1/LTR2	RT1/RT2	ENV1/ENV4[c]
23.1	*env*	LTR1/LTR2[d]		ENV2/ENV3[e]
25.2	*env*	LTR1/LTR2[d]		ENV2/ENV3[e]
30.1	RT		RT1/RT2	
33.1	RT		RT1/RT2	
36.1	RT		RT1/RT2	
37.1	RT		RT1/RT2	

[a] Probe initially used to screen the baboon library.
[b] Primer sets which gave a positive PCR signal are shown.
[c] Outer primer set.
[d] The fragment obtained was not of the correct size.
[e] Inner primer set.

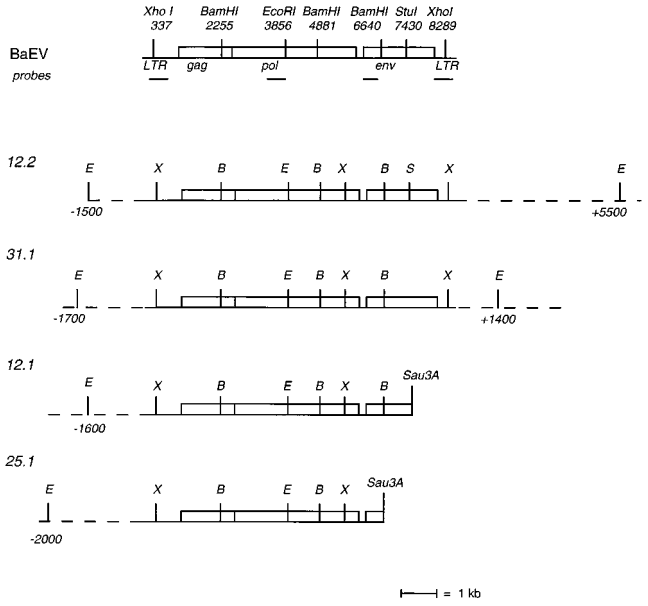

FIG. 1. Characterization of four baboon genomic lambda clones containing BaEV sequences. The complete BaEV genome is 8,507 nt in size (17) and is shown in the upper part of the figure. Viral fragments used for PCR and hybridization are indicated. The *pol* fragment is located within the RT part of the gene. Each individual clone contains an insert of approximately 18 to 22 kb of baboon genomic DNA. Sequences flanking the viral information are indicated by dashed lines, and the approximate positions of *Eco*RI (*E*) sites relative to the BaEV sequences are shown. Each viral clone contains an additional *Xho*I (*X*) site at the 3′ end of the *pol* gene relative to the reference sequence. For clone 12.2, the utmost left *Eco*RI site shown is the one present in the lambda Dash^{R}II vector, and the orientation of the insert has been determined to be 3′ to 5′ with respect to the vector sequences. *B*, *Bam*HI; *S*, *Stu*I.

detection system from Boehringer. The probes were prepared by PCR amplification (with biotin-dCTP) of plasmids containing appropriate BaEV fragments, which had been used previously to screen the genomic library. These procedures confirmed that clones 12.2 and 31.1 contain full-length viruses with no apparent deletions or insertions, while the BaEV sequence of clone 25.1 had been artificially terminated in the *env* gene downstream of the probe used. This was also the case for clone 12.1, although for this clone a *Sau*3A site more downstream in the *env* gene was used (probably at position 7125 or at position 7397, as the *Stu*I site at position 7430 was no longer present). A schematic representation of the obtained BaEV clones is shown in Fig. 1. Sequencing of the RT and *env* fragments amplified from the four lambda clones (Fig. 2 and 3) showed that they were highly homologous to each other and to the BaEV reference sequence. Because there is no 3′ integration site present in two of the four obtained BaEV clones because of *Sau*3A digestion of the proviral DNA, we have investigated only the 5′ viral integration site sequences in all four clones. For clones 12.1 and 25.1, each containing a single LTR, the sequencing was done directly from the lambda DNA with the reverse complement of primer LTR1 (5′ GCCTGCGCCACG AACAAACAGAACC 3′). In the case of clones 12.2 and 31.1, 5′ integration sites were subcloned as *Eco*RI-*Xho*I fragments into plasmid pSP73 (Promega) digested with *Eco*RI and *Xho*I, and the sites were sequenced with the T7 sequencing primer. The sequences adjoining the 5′ LTR are shown in Fig. 4. It is clear that all proviruses obtained have integrated in a different part of the baboon genome, confirming that no two clones are the same. The sequence adjoining the viral 5′ LTR of clone 12.1 was shown to be part of a long interspersed repeat sequence (LINE) L1 repeat (showing 83 to 89% homology with human LINE repeats), while the upstream sequences in the other three clones could not be identified (with EMBL release 37). A second LINE repeat was accidentally subcloned and sequenced from the lambda DNA of clone 12.2 (result not shown), indicating that this provirus is located in the proximity of such an element, or vice versa. LINE copy numbers are sufficiently high in the primate genome to detect these repeats in a certain part of the genome by chance, but it could also be possible that both retroviruses and retroelements prefer the same genomic location, explaining why we found LINE repeats in at least two out of four lambda clones, while it is also believed that LINE repeats can still be inserted de novo (for a review, see reference 33). The integration of retroviral DNA into the host genome has been found to be dependent upon several factors, e.g., the particular target sequence (containing a small [4- to 6-bp] direct repeat), the tertiary structure of the DNA (like that of intrinsically bent DNA), and the presence of DNA binding proteins (14, 21, 26, 32). The viral integration machinery prefers transcribed DNA regions (23) and nucleosomal DNA (26), while acceptor sites are DNase I-hypersensitive regions (42). For BaEV, a highly preferred integration locus, termed BEVI, was found on human chromosome 6 (10, 20). The locus is characterized by the presence of multiple *Pst*I sites, which were found to be present at the 3′ ends of the de novo integrated viruses. At the 5′ end of clone 12.1, we detected a *Pst*I site located at a single-nucleotide distance from the viral LTR, which should be regarded as a mere coincidence, considering the absence of further sequence information on the BEVI locus. Interestingly, in the two full-length clones 12.2 and 31.1, the two 5′ nucleotides of the viral LTR are missing, as expected in retroviral integration (15), but clone 12.1 apparently contains the complete LTR, while three nucleotides are lacking in clone 25.1. Possibly the extra nucleotides observed in clone 12.1 were contributed by the host genome, and the nucleotide apparently missing in clone 25.1 could have been mutated from A to C in the course of time. Another explanation could be that the viral integration machinery is sometimes defective.

Clones 23.1 and 25.2, which did not amplify with RT primers and did not amplify correctly with LTR primers, also did not hybridize to the corresponding probes and hybridized only weakly to the *env* probe on a Southern blot. Sequencing of the fragments amplified from these clones with ENV2/ENV3 showed that although they have strong homology to the *env* gene of BaEV (approximately 80% at the nucleotide and amino acid level), they are quite different from all BaEV sequences obtained so far (Fig. 3). Both sequences are closely related but not identical to each other. It must be noted that although primers LTR1 and LTR2 did not correctly amplify a product from these clones, an approximately 700-nucleotide (nt) fragment was obtained in each case. These PCR products have probably no other homology with BaEV than (part of) the primer sequences, as no hybridization of the digested lambda DNA was observed with a BaEV LTR fragment, even after prolonged exposure. Sequence analysis should clarify whether the 700-nt fragment has LTR-like features. Sequence results obtained with the reverse complement of primer ENV2 showed that upstream of primer ENV2 both clones also have a strong homology to each other, but any homology to BaEV was completely lost, and similarity to any other known (retro) virus could also not be found (unpublished observation). The predicted BaEV signal peptide-GP70 boundary is located in the ENV2 primer sequence, suggesting that a completely different signal peptide is present in these two putative *env* genes. PCR amplification with the primer set ENV2/ENV4 gave a

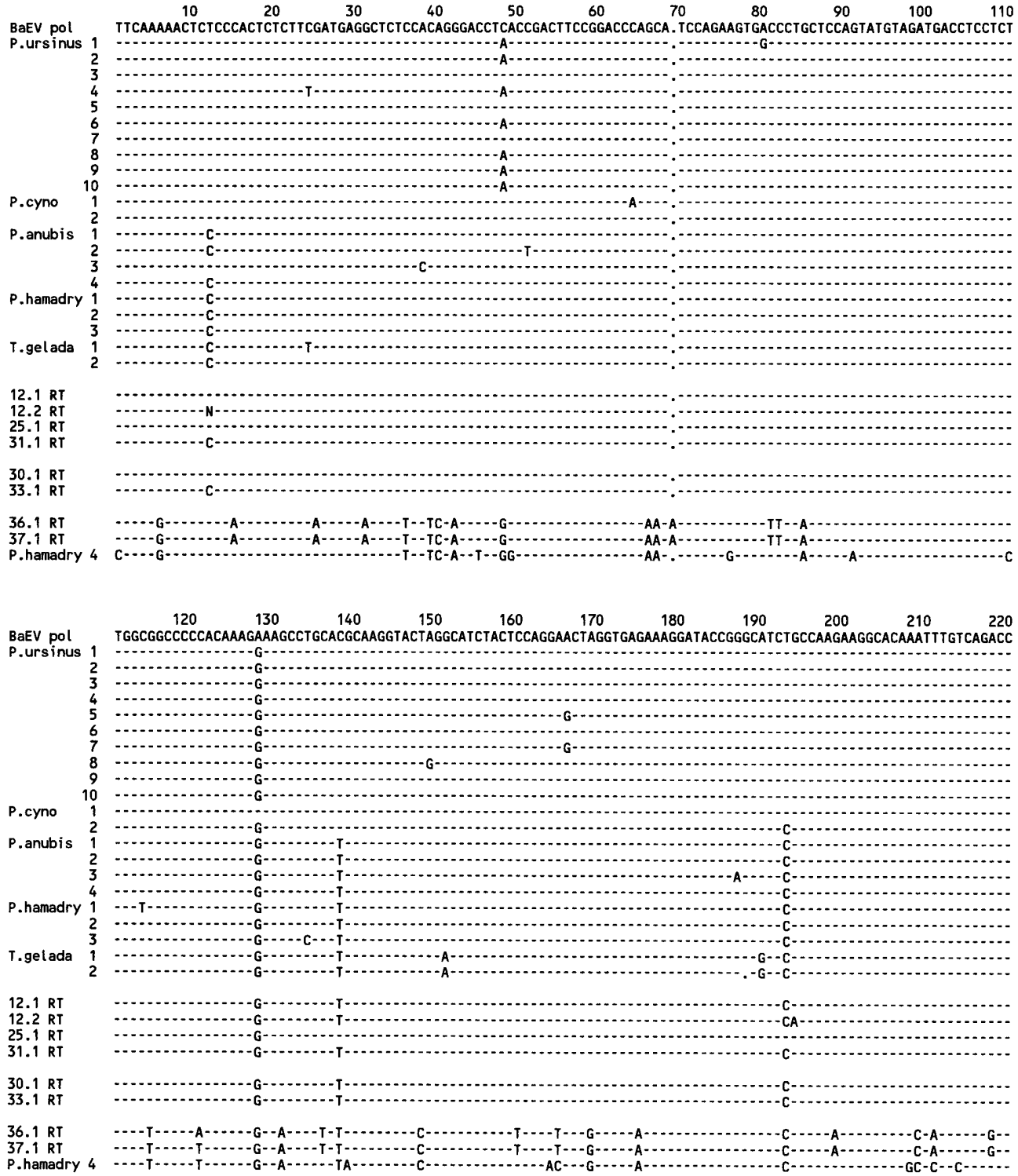

FIG. 2. RT sequences amplified from total DNA of four baboon species (*P. ursinus*, *P. cynocephalus* [P.cyno], *P. anubis*, *P. hamadryas* [P.hamadry]) and the gelada baboon (*T. gelada*), from four BaEV clones (12.1, 12.2, 25.1, and 31.1), and from four lambda clones (30.1, 33.1, 36.1, and 37.1) detected with a BaEV RT probe are shown. The BaEV sequence (nt 3526 to 3849) of Kato et al. (17) is shown in the upper line. Identical nucleotides are indicated by dashes; gaps introduced for optimal alignment are indicated by dots.

positive signal for clone 25.2 but not for 23.1 (result not shown), indicating that the downstream homology to BaEV has not disappeared altogether. Possibly, these two clones represent endogenous retroviral elements belonging to a family of viruses with BaEV-like *env* genes and probably do not represent BaEV genomes with multiple substitutions, as we have sequenced endogenous virus clones from other species of African monkeys with high-level identity to the BaEV *env* frag-

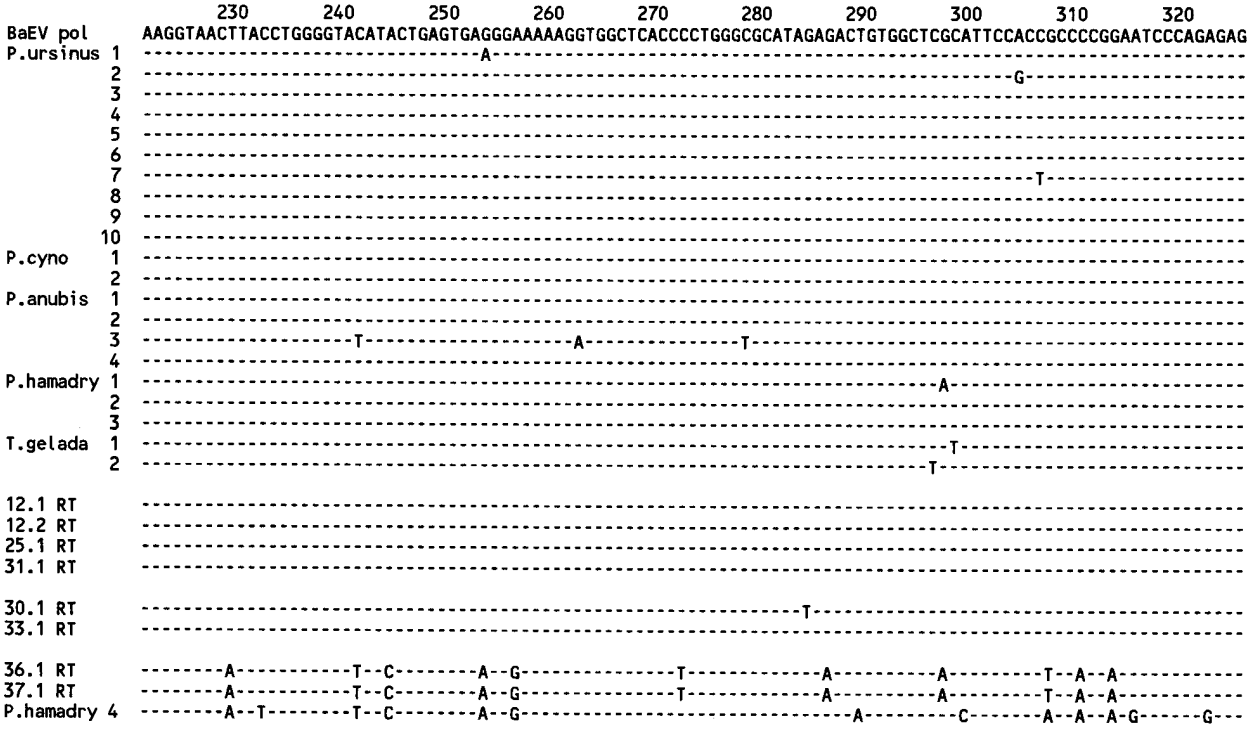

FIG. 2—*Continued.*

ments from baboons (unpublished observation). These species included monkeys of the Papionini tribe (which includes baboons, mandrills, mangabeys, and geladas), but homologous sequences were also amplified from the more distantly related *Cercopithecus aethiops* (African green monkey). The finding of these homologous *env* clones also questions the proposed type C/type D chimeric descent of BaEV. Extensive homology between BaEV, RD114, and the simian type D viruses, which all share the same cellular receptor (35), is predominantly located at the 3′ end of the *env* gene. This part putatively encodes the transmembrane protein P20E in BaEV and probably has the same function in the other viruses. However, until now no virus showed extensive homology with the GP70 *env* region of BaEV.

In an initial filter screening, several clones obtained from the baboon genomic library hybridized only to the RT probe of BaEV and not to an LTR probe. Four of these were isolated and tested by PCR. The results (Table 2) show that the four clones (30.1, 33.1, 36.1, and 37.1) amplify only with RT primers, although very weak large double bands were obtained with the primer set LTR1/LTR2 (not shown). Sequence analysis of individual clones of the RT1/RT2 PCR products produced a surprising result, as two different but highly related RT sequences were present in the four lambda clones (Fig. 2). The RT sequences of two clones (30.1 and 33.1) were identical to BaEV RT (from nt 3505 to 3871), while the other two (36.1 and 37.1) were more divergent (showing approximately 77% homology to BaEV). Sequences 36.1 RT and 37.1 RT were almost identical to each other. Compared with the relative amounts of clones hybridizing to other parts of the BaEV genome (*env* or LTR), the clones exclusively positive for BaEV RT are present in an ~10-fold excess. Restriction analysis of two of these clones (36.1 and 37.1; result not shown) indicated that they are not identical and that each contains at least a small, but possibly the same, part of a BaEV-like RT gene (estimated size between 366 and 945 nt), while the complete *pol* gene of BaEV is 3,590 nt long (protease, nt 2575 to 2958 of the BaEV reference sequence; RT, nt 2959 to 4989; integrase, nt 4990 to 6165). Preliminary sequencing results (not shown) indicate that clone 36.1 contains 620 nt of BaEV-like RT sequence (nt 3311 to 3926), flanked by a mosaic of sequences with homology (ranging from 73 to 92%) to BaEV *gag-pol* fragments. It is known that solitary LTRs derived from the endogenous type B virus HERV-K are dispersed in the human genome in approximately 5,000 to 25,000 copies (19, 24). Homology between these sequences and the putative ancestor virus is 70 to 90%.

To estimate the level of sequence divergence in the BaEV RT gene for individual animals and between different baboon species, RT sequences were PCR amplified from total DNA with the primer set RT1/RT2 and were sequenced (Fig. 2). Very few sequence differences were present for single individuals and between the baboon species examined, including the more distantly related gelada baboon (*T. gelada*). Sequence 25.1 RT was identical to sequence P.ursinus 3, while sequences 12.1 RT, 31.1 RT, P.hamadry 2, P.hamadry 4, and P.anubis 3 were all identical. Sequence 12.2 RT was most closely related to P.hamadry 1 and P.hamadry 4. Sequences differing in a single nucleotide could be amplified from the same allele by *Taq* polymerase. Only once (in a *P. hamadryas* DNA sample) was a sequence (P.hamadry 4) with homology to the divergent sequences 36.1 RT and 37.1 RT found. This suggests that the primer sequences are not completely conserved in this sequence and that the divergent RT is amplified with low efficiency. No substitutions were observed in any sequence in the conserved RT domain YXDD (25), where X is V in BaEV (nt 92-103 in Fig. 2).

To estimate the level of sequence divergence for the BaEV *env* gene for individual animals and between different baboon species, *env* sequences were PCR amplified from total DNA with the primer set ENV1/ENV4 and were sequenced (Fig. 3).

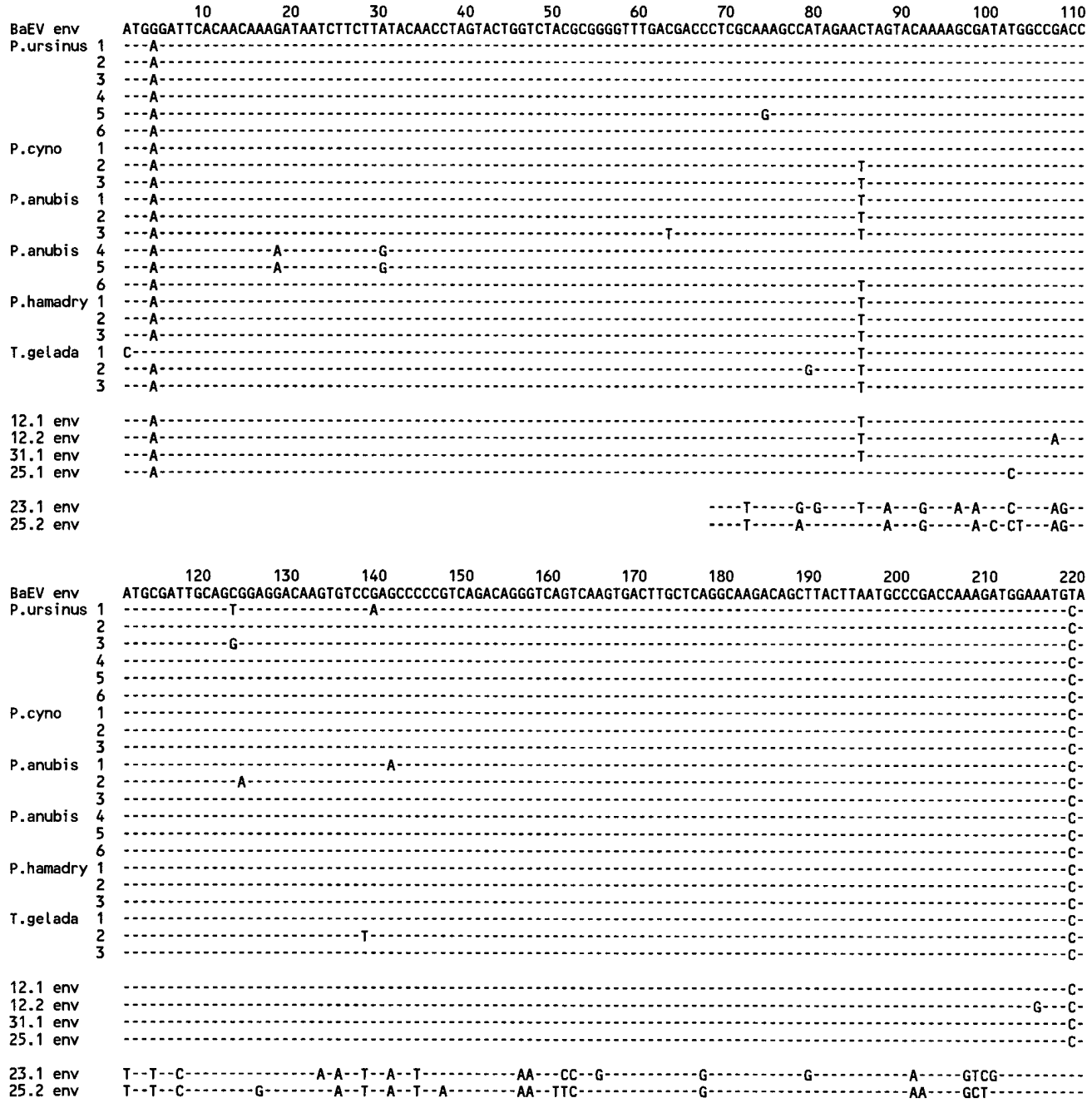

FIG. 3. *env* sequences amplified from total DNA of four baboon species (*P. ursinus*, *P. cynocephalus* [P.cyno], *P. anubis*, and *P. hamadryas* [P.hamadry]) and the gelada baboon (*T. gelada*) and from six lambda clones, obtained after screening a baboon genomic library with a BaEV *env* probe, are shown. The BaEV sequence (nt 6235 to 6609) of Kato et al. (17) is shown in the upper line. *P. anubis* sequences were amplified from two animals. Identical nucleotides are indicated by dashes.

The sequence shown starts with the start codon (ATG) of the *env* gene. The first 60 nt encode a putative signal peptide, while the GP70 coding sequence presumably starts at nt 61. All sequences obtained (including those for the four BaEV genomic clones) were found to differ in three positions from the BaEV reference sequence. The open reading frame was never disturbed by the introduction of stop codons nor by insertions and/or deletions, although in the sequence T.gelada 1 the ATG start codon of *env* was mutated to CTG. Also, in the divergent clones 23.1 and 25.2 the open reading frame was not interrupted in the fragment sequenced. Remarkably few nucleotide differences were present among all BaEV sequences. Sequences 12.1 env and 31.1 env were identical to each other and to sequences P.anubis 6 and P.hamadry 2. Sequences generated from total baboon DNA with heterogeneity only in a single position could have been amplified from the same allele by *Taq* polymerase. There was no clear association of sequence clusters with baboon species.

BaEV proviruses have been calculated to be present in approximately 50 to 100 copies in the haploid baboon genome. This estimation was based upon Southern blotting of baboon genomic DNA and detection of BaEV with an RT fragment as

```
                  230        240        250        260        270        280        290        300        310        320        330
BaEV env          AGTCAATTCCAAAAGACACCTCCCCAAGCGGGCCACTCCAAGAGTGCCCCTGTAATTCTTACCAGTCCTCAGTACACAGTTCTTGTTATACCTCATACCAACAATGCAGA
P.ursinus 1       ---T---------------------T-------------------------------------------------------------------------------------
          2       ---T---------------------T-------------------------------------------------------------------------------------
          3       ---T---------------------T-------------------------------------------------------------------------------------
          4       ---T---------------------T-------------------------------------------------------------------------------------
          5       ---T---------------------T-------------------------------------------------------------------------------------
          6       ---T---------------------T-------------------------------------------------------------------------------------
P.cyno    1       ---T---------------------T-------------------------------------------------------------------------------G-----
          2       --CT---------------------TG------------G------------G-------------------------------------------------------------
          3       ---T---------------------T--------------------T-------------------------------------------------------------------
P.anubis  1       ---T---------------------T--------------------------G-------------------------------------------------------------
          2       ---T---------------------T--------------------------G-------------------------------------------------------------
          3       ---T---------------------T--------------------------G-------------------------------------------------------------
P.anubis  4       ---T---------------------T--------------------------G-------------------------------------------------------------
          5       ---T---------------------T--------------------------G-------------------------------------------------------------
          6       ---T---------------------T--------------------------G-------------------------------------------------------------
P.hamadry 1       ---T---------------------T--------------------------G-------------------------------------------------------------
          2       ---T---------------------T--------------------------G-------------------------------------------------------------
          3       --CT---------------------T--------------------------G-------------------------------------------------------------
T.gelada  1       ---T---------------------T--------------------------G---------------------------------C---------------------------
          2       ---T---------------------T--------------------------G---------------------------------C---------------------------
          3       ---T---------------------T--------------------------G---------------------------------C---------------------------

12.1 env          ---T---------------------T--------------------------G-------------------------------------------------------------
12.2 env          ---T---------------------T--------------------------G-------------------------------------------------------------
31.1 env          ---T---------------------T--------------------------G-------------------------------------------------------------
25.1 env          ---T----------

23.1 env          ----T-CC----G-
25.2 env          ------CC----G-
```

```
                  340        350        360        370
BaEV env          TCAGGCAATAAGACATATTATACGGCTACTCTGCTAAAAACACAA
P.ursinus 1       ---------------------------------------------
          2       ---------------------------------------------
          3       ---------------------------------------------
          4       --------------------T------------------------
          5       ---------------------------------------------
          6       ---------------------------------------------
P.cyno    1       ---------------------------------------------
          2       ---------------------------------------------
          3       ----------------------T----------------------
P.anubis  1       ---------------------------------------------
          2       ---------------------------------------------
          3       ---------------------------------------------
P.anubis  4       ---------------------------------------------
          5       ----------------------------T----------------
          6       ---------------------------------------------
P.hamadry 1       --------------------T------------------------
          2       ---------------------------------------------
          3       ---------------------------------------------
T.gelada  1       ---------------------------------------------
          2       ---------------------------------------------
          3       ---------------------------------------------

12.1 env          ---------------------------------------------
12.2 env          ------------------------T--T-----------------
31.1 env          ---------------------------------------------
```

FIG. 3—*Continued.*

probe (31). We have shown here that this method is unreliable because BaEV-like RT fragments are present as retroelements in multiple copies in the baboon genome. Part of this RT sequence was used as probe by Shih et al. (31). Besides, BaEV RT and BaEV *env* sequences (and probably other parts of the genome as well) hybridize efficiently to genomic DNA sequences more distantly related to the virus, making Southern blot quantification with other probes or quantitative PCR methods difficult. We estimate that the number of BaEV proviral genomes is ~10-fold lower than the number calculated by Shih et al. for baboons, which would be more in accordance with the 5 to 15 integrations per haploid baboon genome presented by Benveniste and Todaro (5).

In conclusion, screening a baboon (*P. cynocephalus*) genomic library with probes specific for BaEV led to the isolation of four BaEV clones, two lambda clones containing apparently full-length proviruses and two others containing provirus clones terminating in the *env* gene as a consequence of the construction method of the library. In contrast to endogenous type C viruses of other mammalian species, e.g., endogenous MuLV and endogenous FeLV and RD114 of cats, there are no strong indications for the existence of defective BaEV genomes in the baboon, although the homologous RT-containing elements could be defined as proviruses with large genomic

12.2 5′ integration site

```
5′  TGTGTGTCTGGATGTGTGGATGTGGATGTGGATGTGGATGTGTGTGGACAAAAAATAAAG
        -120       -110       -100       -90        -80        -70
    GCCCTTATAATAAACCTCNAGGAATTTATTGTCCATTGGCCATTAAATCTGTTTTAATTT
        -60        -50        -40        -30        -20        -10
    CCCTCTATGAAAAGTAAAACTTTTAGGCCCTCCCCGAAAT  3′
         +1         +10        +20        +30
```

31.1 5′ integration site

```
5′  ACCAACAAACCCGTTTGACCAAAAAGCCAAACTGTAAAATATTTAAAGAGGTTTATTCTG
   -150       -140       -130       -120       -110       -100
    AGCTAGTATGAGTGACCATGGCCTGAGGAACAGCCTCAAGAGATTCTAAGAAAATGTGTC
   -90        -80        -70        -60        -50        -40
    CTATGTGGACAGGTTGTAGCTTCGTCTTATATGAAAAGTAAAACTTTTAGGCCCTCCCCG
   -30        -20        -10        +1         +10        +20        +30

    TAAAT  3′
```

12.1 5′ integration site

```
5′  CCCTCACCCCTGGACAGGCCCCAGTGTGTGATGTTCCCCTCCCTGTGTCCATGTGTTCTC
     -300       -290       -280       -270       -260       -250
    ATTGTTCAACTCCCACTTATGACTGAGAATATGCAGTGTTTGGTTTTGTGTTCCCGTGTT
     -240       -230       -220       -210       -200       -190
    AGTTTGCTGAGAATGGTGGTTTCCAGCTTCATCCATGTCCCTGTGAAGGACATGAACTCA
     -180       -170       -160       -150       -140       -130
    TCCTTTTTTACAGCTACATCGTATTCCATGGTGTATATGTGCCACATTTTCTTTATCCTG
     -120       -110       -100       -90        -80        -70
    TTAGTCATTGATGGGCATTTGGGTTGGTTCCAAGTCTTTGCTATTGTGAATAGTGCTGCA
     -60        -50        -40        -30        -20        -10
    GTAAATGAAAAGTAAAACTTTTAGGCCCTCCCCGTAAAT  3′
     +1         +10        +20        +30
```

25.1 5′ integration site

```
5′  CTAACCACTGGGTGGGTGTGGGAAAAGTATGTGTCACTCTTGACCGCCTGTTATCTCATG
    -140       -130       -120       -110       -100       -90
    GCCATTGCCACAGCCTTTTCAGTGCACAAAGCATTCAGACCACTAATTGGAAGAGTCAGC
    -80        -70        -60        -50        -40        -30
    AAAAAGTTGTTGGAGTTGGTCTGAAAAGTAAAACTTTTAGGCCCTCCCCGTAAAT  3′
    -20        -10        +1         +10        +20        +30
```

FIG. 4. Sequences of 5′ integration sites of four obtained BaEV genomic clones. Part of the viral 5′LTR is underlined. A *Pst*I site located next to the viral LTR is indicated in boldface for clone 12.1.

deletions. However, the methods used for identification of the BaEV sequences (Southern blotting and PCR amplification) cannot exclude the existence of small insertions, deletions, or stop codons in the clones obtained. The low amount of sequence variation present between the BaEV clones, as observed in PCR fragments from the RT and *env* genes, was illustrative of BaEV variation in baboons as demonstrated by random amplification of these fragments from total DNA of four different baboon species and the more distantly related gelada baboon. Because all BaEV lambda clones could be amplified with RT and *env* primer sets, we think it is unlikely that we missed BaEV provirus variation because of primer mismatch or the PCR protocol used. Besides, more divergent clones, detected by Southern blotting, were more distantly related to BaEV. This lack of variation is not expected for endogenous viruses, which are supposed to act as pseudogenes, the latter supposedly having a relatively high mutation frequency (5×10^{-9} substitution/site/year [22]), although a similar lack of variation has been observed for endogenous MuLV *env* genes (36). This lack of variation could be attributed to a relatively short history of BaEV in the baboon, which could also explain the existence of active proviruses in this species. The amount of substitutions observed in the RT fragments of clones 36.1 and 37.1 is more in accordance with the mutation rate calculated for pseudogenes than that observed with complete BaEV proviruses.

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

REFERENCES

1. **Benveniste, R. E., R. Heinemann, G. L. Wilson, R. Callahan, and G. J. Todaro.** 1974. Detection of baboon type C viral sequences in various primate tissues by molecular hybridization. J. Virol. **14:**56–67.
2. **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.
3. **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.
4. **Benveniste, R. E., and G. J. Todaro.** 1974. Evolution of type C viral genes. I. Nucleic acid from baboon type C virus as a measure of divergence among primate species. Proc. Natl. Acad. Sci. USA **71:**4513–4518.
5. **Benveniste, R. E., and G. J. Todaro.** 1974. Multiple divergent copies of endogenous C-type virogenes in mammalian cells. Nature (London) **252:**171–173.
6. **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.
7. **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.
8. **Bonner, T. I., and G. J. Todaro.** 1980. The evolution of baboon endogenous type C virus: related sequences in the DNA of distant species. Virology **103:**217–227.
9. **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 purification of nucleic acids. J. Clin. Microbiol. **28:**495–503.
10. **Cohen, J. C., and M. Murphey-Corb.** 1983. Targeted integration of baboon endogenous virus in the BEVI locus on human chromosome 6. Nature (London) **301:**129–132.
11. **Cohen, M., N. Davidson, R. V. Gilden, R. M. McAllister, M. O. Nicolson, and R. M. Stephens.** 1980. The baboon endogenous virus genome. II. Provirus sequence variations in baboon cell DNA. Nucleic Acids Res. **8:**4423–4440.
12. **Cohen, M., M. Powers, C. O'Connell, and N. Kato.** 1985. The nucleotide sequence of the *env* gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA. Virology **147:**449–458.
13. **Cohen, M., A. Rein, R. M. Stephens, C. O'Connell, R. Gilden, M. Shure, M. O. Nicolson, R. M. McAllister, and N. Davidson.** 1981. Baboon endogenous virus genome: molecular cloning and structural characterization of nondefective viral genomes from DNA of a baboon cell strain. Proc. Natl. Acad. Sci. USA **78:**5207–5211.
14. **Hughes, S. H., A. Mutschler, J. M. Bishop, and H. E. Varmus.** 1981. A Rous sarcoma virus provirus is flanked by short direct repeats of a cellular sequence present in only one copy prior to integration. Proc. Natl. Acad. Sci. USA **78:**4299–4303.
15. **Hughes, S. H., P. R. Shank, D. H. Spector, H.-J. Kung, J. M. Bishop, H. E. Varmus, P. R. Vogt, and M. L. Breitman.** 1978. Proviruses of avian sarcoma virus are terminally redundant, co-extensive with unintegrated linear DNA and integrated at many sites. Cell **15:**1397–1410.
16. **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.
17. **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.
18. **Leib-Mösch, C., R. Brack-Werner, T. Werner, M. Bachmann, O. Faff, V. Erfle, and R. Hehlmann.** 1990. Endogenous retroviral elements in human DNA. Cancer Res. **50**(Suppl.)**:**5636s–5642s.
19. **Leib-Mösch, C., M. Haltmeier, T. Werner, E.-M. Geigl, R. Brack-Werner, U. Francke, V. Erfle, and R. Hehlmann.** 1993. Genomic distribution and transcription of solitary HERV-K LTRs. Genomics **18:**261–269.
20. **Lemons, R. S., S. J. O'Brien, and C. J. Sherr.** 1977. A new genetic locus, Bevi, on human chromosome 6 which controls the replication of baboon type C virus in human cells. Cell **12:**251–262.
21. **Milot, E., A. Belmaaza, E. Rassart, and P. Chartrand.** 1994. Association of a host DNA structure with retroviral integration sites in chromosomal DNA. Virology **201:**408–412.
22. **Minghetti, P. P., and A. Dugaiczyk.** 1993. The emergence of new DNA repeats and the divergence of primates. Proc. Natl. Acad. Sci. USA **90:**1872–1876.
23. **Mooslehner, K., U. Karls, and K. Harbers.** 1990. Retroviral integration sites in transgenic Mov mice frequently map in the vicinity of transcribed DNA regions. J. Virol. **64:**3056–3058.
24. **Ono, M., M. Kawakami, and T. Takezawa.** 1987. A novel human nonviral retroposon derived from an endogenous retrovirus. Nucleic Acids Res. **15:**8725–8737.
25. **Poch, O., I. Sauvaget, M. Delarue, and N. Tordo.** 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. **8:**3867–3874.
26. **Pryciak, P. M., and H. E. Varmus.** 1992. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell **69:**769–780.
27. **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.
28. **Reeves, R. H., and S. J. O'Brien.** 1984. Molecular genetic characterization of the RD-114 gene family of endogenous feline retroviral sequences. J. Virol. **52:**164–171.
29. **Sherr, C. J., M. M. Lieber, R. E. Benveniste, and G. J. Todaro.** 1974. Endogenous baboon type C virus (M7): biochemical and immunologic characterization. Virology **58:**492–503.
30. **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.
31. **Shih, A., E. E. Coutavas, and M. G. Rush.** 1991. Evolutionary implications of primate endogenous retroviruses. Virology **182:**495–502.
32. **Shih, C.-C., J. P. Stoye, and J. M. Coffin.** 1988. Highly preferred targets for retrovirus integration. Cell **53:**531–537.
33. **Singer, M. F., and J. Skowronski.** 1985. Making sense out of LINES: long interspersed repeat sequences in mammalian genomes. Trends Biochem. Sci. **10:**119–122.
34. **Soe, L. H., B. G. Devi, J. I. Mullins, and P. Roy-Burman.** 1983. Molecular cloning and characterization of endogenous feline leukemia virus sequences from a cat genomic library. J. Virol. **46:**829–840.
35. **Sommerfelt, M. A., and R. A. Weiss.** 1990. Receptor interference groups of 20 retroviruses plating on human cells. Virology **176:**58–69.
36. **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.
37. **Stoye, J. P., and J. M. Coffin.** 1988. Polymorphism of murine endogenous proviruses revealed by using virus class-specific oligonucleotide probes. J. Virol. **62:**168–175.
38. **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.
39. **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.
40. **Todaro, G. J., C. J. Sherr, R. E. Benveniste, M. M. Lieber, and J. L. Melnick.** 1974. Type C viruses of baboons: isolation from normal cell cultures. Cell **2:**55–61.
41. **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.
42. **Vijaya, S., D. L. Steffen, and H. L. Robinson.** 1986. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. J. Virol. **60:**683–692.