JOURNAL OF VIROLOGY, Sept. 1995, p. 5890–5897 Vol. 69, No. 9
0022-538X/95/$04.00+0

Isolation of Novel Human Endogenous Retrovirus-Like Elements with Foamy Virus-Related *pol* Sequence

AGNÈS CORDONNIER, JEAN-FRANÇOIS CASELLA, AND THIERRY HEIDMANN*

Unité de Physicochimie et Pharmacologie des Macromolécules Biologiques, Centre National de la Recherche Scientifique URA 147, Institut Gustave Roussy, 94805 Villejuif Cedex, France

Received 5 April 1995/Accepted 13 June 1995

A new class of reverse transcriptase coding sequences was detected in reverse-transcribed RNAs from human placenta by polymerase chain amplification with primers in highly conserved regions of the *pol* gene of mammalian retroviruses and retrotransposons. Using one of these novel sequences as a probe to screen a human genomic library, we isolated retrovirus-like elements bordered by long terminal repeats and having a potential leucine tRNA primer-binding site. Determination of the complete nucleotide sequence (6,591 bp) of one of these elements, termed HERV-L (for human endogenous retrovirus with leucine tRNA primer), revealed domains of amino acid similarities to retroviral reverse transcriptase and integrase proteins. In addition, a region with homologies to dUTPase proteins was found unexpectedly downstream from the integrase domain. Amino acid sequence and phylogenetic analyses indicate that the HERV-L *pol* gene is related to that of foamy retroviruses. HERV-L-related sequences are detected in several mammalian species and have expanded in primate and mouse genomes up to 100 to 200 copies.

Endogenous retroviruses compose 0.1% of the human genome and can be divided into several distinct families with copy numbers of 1 to 1,000 per haploid genome (reviewed in references 26 and 56). Generally, they have been detected in human DNA by low-stringency screening of genomic libraries with either DNA or oligonucleotide probes from known retroviruses (4, 5, 11, 15, 25, 27, 31, 34, 40–42). Some endogenous retroviruses were discovered incidentally in the course of DNA sequence analyses (29, 30). Retroviral particles were observed by electron microscopy in human placentas (19) and teratocarcinoma cell lines (3, 28), indicating that at least some endogenous retroviral sequences are functional. Because of the possible biological role of such sequences and their potential pathogenic effect, many attempts were made to generate probes homologous to expressed endogenous retroviruses and to identify functional sequences. We describe here a successful application of a different approach, first described by Shih et al. (50), to detect novel reverse transcriptase coding sequences in human nucleic acids. This method is based on polymerase chain amplifications using universal primers within the best-conserved amino acid domains (L-P-Q-G and Y-X-D-D boxes) of reverse transcriptases from retroviruses and retrotransposons (58). Polymerase chain amplification of reverse-transcribed mRNA from human placenta with degenerate primers (Fig. 1A) shorter and less specific than those previously described (2, 12) allowed us to detect still-uncharacterized nucleic acid sequences; this sensitive approach might be used on substrate RNA isolated from retrovirus-like particles as well.

Total RNA was extracted from full-term human placenta tissue by the guanidium-CsCl method. Poly(A)$^+$ RNAs were selected by oligo(dT)-cellulose chromatography and treated with RNase-free DNase. PCR was performed after reverse transcription under standard conditions except that high primer concentrations (4 μM) were used to compensate for the degeneracy of the primers. PCRs were performed at low annealing temperatures (10 cycles at 37°C followed by 30 cycles at 40°C) and resulted in many nonspecific products visible on ethidium bromide-stained acrylamide gels (data not shown). DNA of the expected size (approximately 130 bp) was eluted from the gel and reamplified with the same primers. DNA was extracted with phenol-chloroform, cloned into pBluescript vector (Stratagene, La Jolla, Calif.), and sequenced by the Sanger dideoxynucleotide method (47). In spite of the size selection, a large fraction of the sequences (70%) had no detectable homology to reverse transcriptases. Among the others, we identified some sequences which were related to reverse transcriptases from previously identified human endogenous retroviruses (HERV; these were not analyzed further) and new sequences with no homology to previously characterized proviral DNA (except the conserved S-P box, 3 amino acids after the L-P-Q-G box, which critically identifies reverse transcriptases). Amino acid comparison of these elements shows that they constitute a class of related sequences of 29 amino acids (instead of 30 in the majority of known reverse transcriptases) characterized by conserved motifs (Fig. 1B).

Cloning and DNA sequence analysis of HERV-L. To characterize further this new class of reverse transcriptase sequences and to determine whether they are parts of endogenous proviral genomes, we used the DNA from clone 4 (Fig. 1B) as a probe to screen a human genomic library (a gift from A. Dejean, Pasteur Institute). From 10^5 phage plaques screened under moderately stringent conditions (wash in 0.5× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate]–0.1% sodium dodecyl sulfate for 15 min at 65°C), 10 positive clones were obtained. Cross-hybridization analysis of subcloned restriction fragments and partial sequencing revealed that the clones contain related sequences (80% homologous) with different restriction maps, indicating that they correspond to different genomic locations. Phage clone 10 contained a complete element with two long terminal repeats (LTRs) flanking an internal sequence without any repetitive DNA, and this element was colinear to those from the other clones. DNA from this representative clone was entirely sequenced (6,591 bp [Fig. 2]) by the dideoxy chain termination method (47). The presence of an imperfect cellular sequence duplication border-

* Corresponding author. Mailing address: Unité de Physicochimie et Pharmacologie des Macromolécules Biologiques, CNRS URA 147, Institut Gustave Roussy, 39 rue Camille Desmoulins, 94805 Villejuif Cedex, France. Phone: 33/1 45 59 49 70. Fax: 33/1 46 78 41 20.

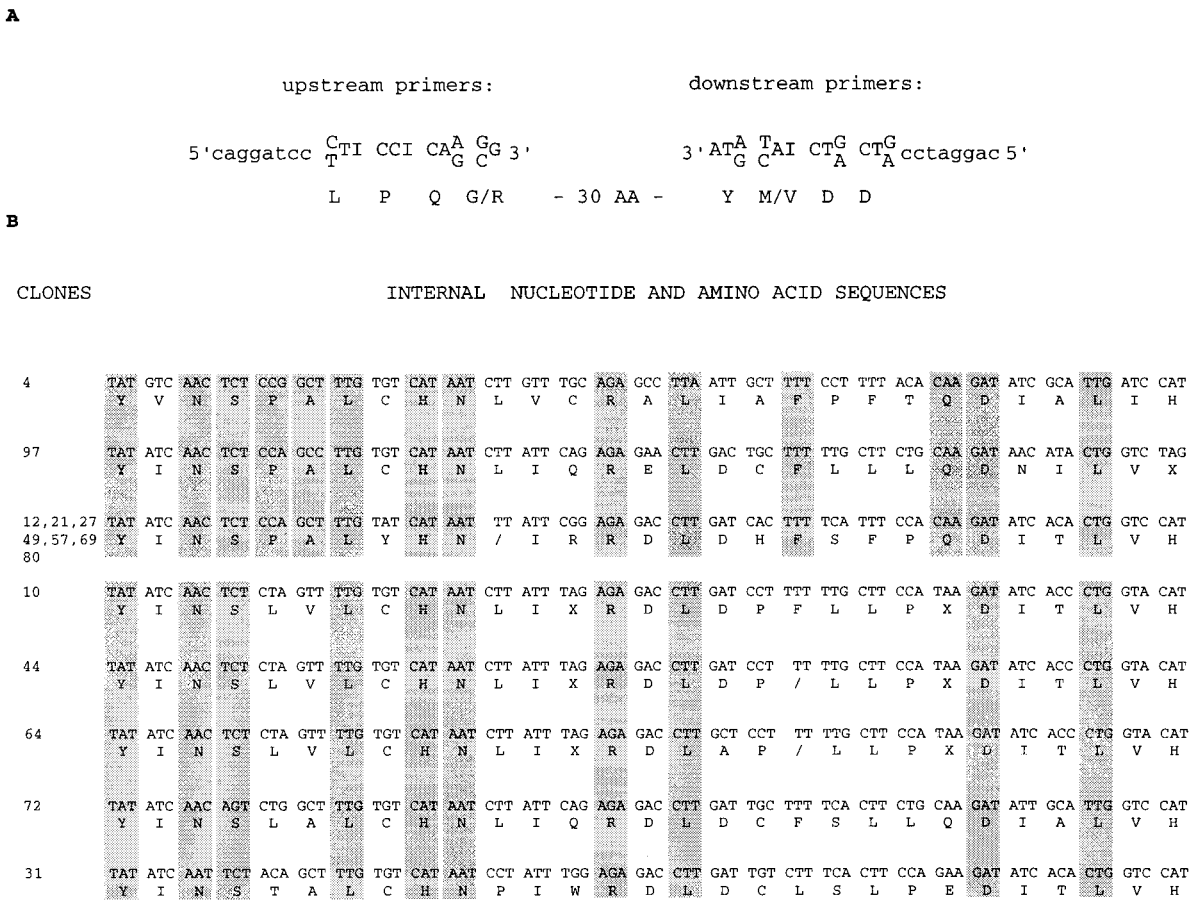

FIG. 1. Identification of *pol*-related sequences by polymerase chain amplification of reverse-transcribed RNA from human placenta tissue. (A) The primers derived from amino acid motifs highly conserved within *pol* regions of retroviruses are indicated. For clarity, the downstream primers are represented 3′ to 5′ rather than by the usual convention. Primer nucleotides in lowercase letters represent noncomplementary 5′ extensions that contain recognition sequences for *Bam*HI. I, inosine. (B) Comparison of nucleotide and predicted amino acid internal sequences of the *pol*-related PCR-amplified DNA fragments (clone numbers are indicated on the left). Shaded letters identify conserved amino acid motifs. Sequences in the bottom group are related sequences that do not contain the typical S-P box (positions 4 and 5). Frameshifts in the amino acid sequence are shown with a slash, and stop codons are indicated with an X.

ing the 6,591-bp proviral element (TATAT in front of the 5′ LTR and CATAT following the 3′ LTR) indicates that this element actually originates from an integration event.

The 5′ and 3′ LTRs are 82% identical over 462 bp. They present the usual features of retroviral LTRs: they are bordered by short inverted repeats (TGT. . .ACA), and they contain a CAT box, 43 bp upstream from a presumptive TATA box, and a polyadenylation signal. The presence of the TATA box was ascertained by comparing the sequence of another cloned LTR that actually contained a typical TATAAA signal (instead of AATAAA in clone 10) at the same position. Screening of the LTR for known transcription factor binding sites reveals the presence of one consensus sequence for AP-1 binding (Fig. 2) (33). A pentanucleotide (AATTT, also found in two other partially sequenced clones) separates the 5′ LTR from a putative tRNA primer-binding site, which was found to be most closely related to the complementary sequence of the 3′ end of a mouse leucine tRNA (CAG anticodon) (45). The sequence corresponding to this tRNA is indicated under the primer-binding site in Fig. 2 and discloses a 2-bp mismatch and a 2-bp deletion (mismatches and/or deletions are also observed in other retroviruses or retrovirus-like elements). The human homolog of this tRNA has not yet been sequenced but should not significantly differ, as it is conserved among mice, rats, and *Drosophila melanogaster* (45). A small purine stretch (13 residues interrupted by two pyrimidines) is present close to the 3′ LTR, at the expected position for the polypurine track in all known retroviruses and retrovirus-like elements. This proviral sequence will now be tentatively referred to as HERV-L for HER with leucine tRNA as the most likely primer for reverse transcription.

Coding regions. Computer-assisted translation of the total nucleotide sequence into amino acid sequence revealed the presence of many stop codons, indicating that the cloned HERV-L element could not code for functional gene products. With the BLASTX program (University of Wisconsin Genetics Computer Group), which translates both strands of a query sequence in all six reading frames and identifies a protein coding region by data similarity search, no evidence for homology of HERV-L to any known Gag or Env sequences was revealed. However, by using the FASTA computer program (which permits the introduction of gaps in the search for similarity), a small region of 35 amino acids (indicated as box A in Fig. 2) was found to be 37% identical to a peptide from the end of the human foamy virus (HFV) Gag (32). This peptide includes a motif rich in glycine and arginine residues possibly involved in interactions with nucleic acids. As for the foamy viruses, HERV-L lacks the typical cysteine-and-histidine motif

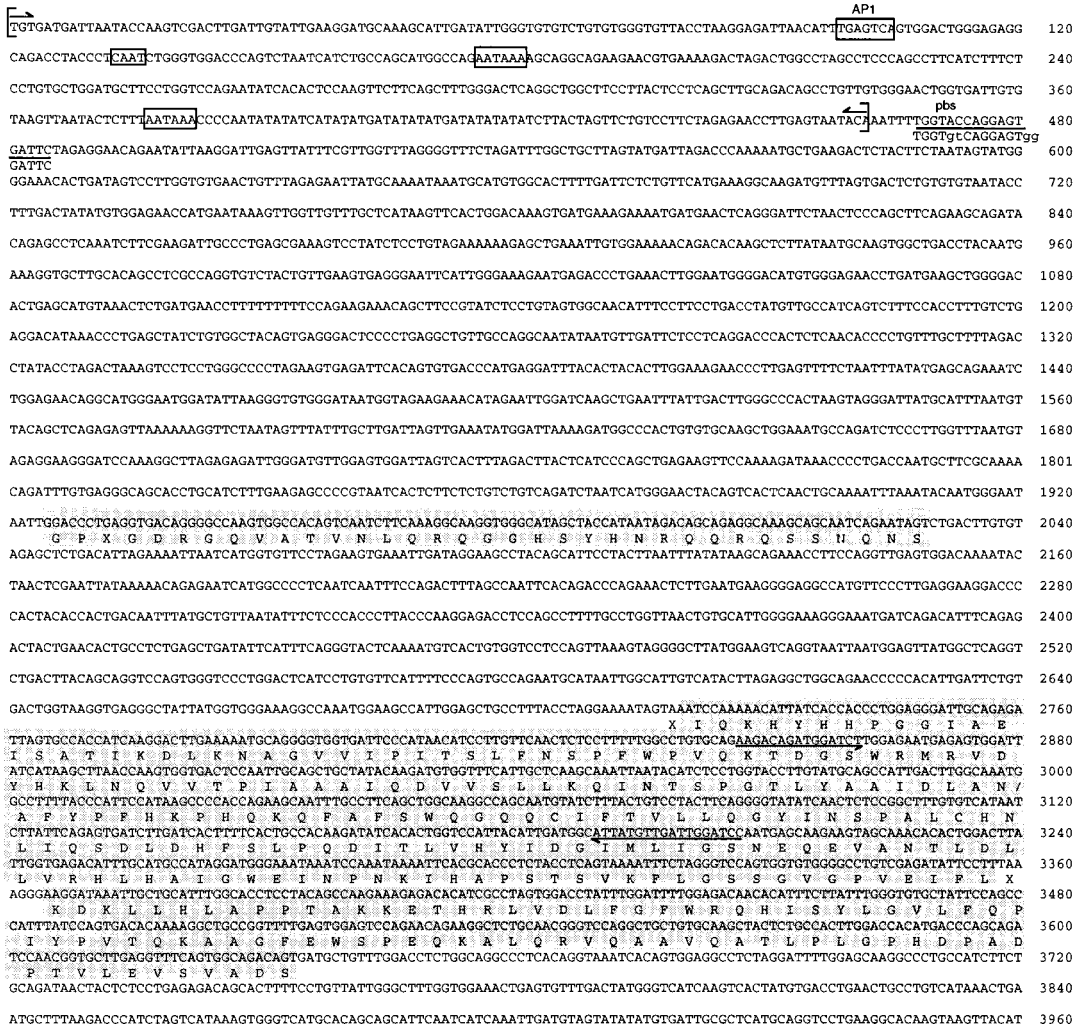
AP1
pbs
Box A
Box B
Box C
Box D
Box E
Box F
ppt
AP1

FIG. 2. Nucleotide sequence of HERV-L proviral DNA. LTRs are enclosed by square brackets, and the small inverted termini TGT and ACA are overlined with arrows. The transcriptional regulatory sequences in the LTRs, i.e., the AP-1 site, the CAAT and TATA boxes, and the polyadenylation signal, are boxed. The primer-binding site (pbs) and polypurine tract (ppt) are underlined. Sequence complementary to the 3′ end of mouse leucine tRNA is shown under the pbs sequence, with lowercase letters for mismatched nucleotides. The nucleotides underlined with arrows in box B correspond to the primers used for PCR. Translated amino acid sequences with homologies to those of other retroviruses are given under the nucleotide sequences in the shaded boxes A to F. Frameshifts in the amino acid sequence are indicated with a slash, and stop codons are indicated with an X.

normally found in the nucleic acid-binding domains of all other retroviral Gag proteins.

In contrast to Gag and Env, several open reading frames with homology to Pol retroviral proteins were detected in HERV-L by using the BLASTX program. Coding domains homologous to the reverse transcriptases of various retroviruses were found, with the following top scores: 40% identity with HFV for amino acids encoded by nucleotides 2748 to 2954 (frame 3); 62% identity with HFV, simian foamy virus type 1 (24), and simian foamy virus type 3 (43) for those encoded by nucleotides 3040 to 3111 (frame 1); and 36% identity with Mason-Pfizer monkey virus (52) for the portion encoded by nucleotides 3169 to 3291 (frame 1). These coding domains are maintained on the same reading frame, provided that a one-guanosine-residue deletion is introduced at position 3000 in phage 10, as actually observed in the nucleotide sequences of DNAs from four other phage clones. The predicted sequence of the protein encoded by the complete open reading frame encompassing the segments described above is indicated below the nucleotide sequence in Fig. 2 (box B). The sequence similarities between HERV-L amino acids in box B and the corresponding regions of other retroviruses range from 20% identity for human immunodeficiency virus type 1 (55) to 33% for HFV. The homology with reverse transcriptase extends into another region 3′ to this box (box C, nucleotides 3368 to 3635; frame 2), disclosing 32% amino acid identity with HFV Pol.

Analysis of the sequence 3′ to the reverse transcriptase domain identified the F-T-D-G-S motif conserved in previously described retroviral RNase H (18). Alignment of translated amino acids in this domain (box D) shows additional residues shared by a large fraction of retroviral RNase H. The position of this box is consistent with the presence of the tether region that separates reverse transcriptase and RNase H domains in retroviral *pol* genes.

Two overlapping regions of HERV-L amino acid sequence (box E, nucleotides 4615 to 5262) were found to be homologous to integrase proteins, with the following score: 25% identity with HFV integrase for the first 179 amino acids and 29% identity with bovine leukemia virus integrase (9) for the last 204 amino acids. This protein sequence includes the highly conserved D,D(35)E motif shown to be critical for integrative recombination of retroviruses and transposable elements (23). The N substitution for the second D residue in HERV-L is probably the consequence of a single base mutation (GAC to AAC at position 4927), since the correct D codon was found in another sequenced genomic clone. N terminal to this central catalytic domain, two cysteine residues can be aligned with those found in the potential zinc-binding motif ($H\text{-}X_3\text{-}H\text{-}X_{22–32}\text{-}C\text{-}X_2\text{-}C$) observed in the retroviral integrases (18).

Finally, HERV-L contains a distinct region (nucleotides 5166 to 5587) disclosing 53% DNA homology to a mouse mammary tumor virus (39) retroviral sequence, which has been identified as a dUTPase on the basis of both sequence similarity to the dUTPase family (36) and enzymatic activity (1, 22). A dUTPase sequence is found in some retroviruses (in type B and D oncoviruses and in nonprimate lentiviruses), in poxviruses, and in herpesviruses (36). Comparison of the amino acids encoded by HERV-L (from nucleotide 5266 to 5545; box F) with dUTPases of various origins (Fig. 3) using the CLUSTALV program (16) showed the presence of highly conserved motifs, strongly suggesting that these sequences are related. Sequences with maximum homology were from related retroviruses, thus confirming their evolutionary relationships, with 66% amino acid identity between caprine arthritis encephalitis virus (46) and visna lentivirus (53) and 56% identity between mouse mammary tumor virus and Mason-Pfizer monkey virus. However, the HERV-L dUTPase is not significantly closer to the oncovirus family (39 and 32% amino acid identity with mouse mammary tumor virus and Mason-Pfizer monkey virus, respectively) than to the nonprimate lentivirus family (34% amino acid identity with feline immunodeficiency virus [54]), therefore suggesting that it belongs to a distinct branch. Most importantly, the genomic location of this sequence, overlapping the 3′ end of the integrase domain in HERV-L, is different from that in the other retroviral groups (adjacent to the protease in type B and D oncoviruses and between RNase H and integrase in nonprimate lentiviruses). This demonstrates that dUTPase sequences have been acquired independently in these lineages. No definite conclusion concerning the origin of this gene could be derived from the analysis of the percentages of amino acid identity among the various dUTPases: horizontal transfer from an ancestral retrovirus or a DNA virus as well as capture from the cellular genome are both plausible (8).

Phylogenetic analysis. To determine the relationship between HERV-L and other retroelements, the major part of the protein sequence in box B was tentatively aligned with reverse

```
HERV-L  TLCSAGLEVL SQMEECLPRG D-TTTIPLKW KLRLPPGHFG LLLPLSQQAK KGV-TVLAGV IDQDCQEEIN LLLHNGGKEE YEWNPGDPLV HLLIL
MMTV    TPGSAGLDLS SQKDLILSLE DGVSLVPTLV KGTLPEGTTG LIIGRSSNYK KGL-EVLPGV IDSDFQGEIK -VMVKAAKNA VIIHKGERIA QLLLL
MPMV    TPGSAGLDLC STSHTVLTPE MGPQALSTGI YGPLPPNTFG LILGRSSITM KGL-QVYPGV IDNDYTGEIK -IMAKAVNNI VTVSQGNRIA QLILL
FIV     RSEDAGYDLL AAKEIHLLPG E-VKVIPTGV KLMLPKGHWG LIIGKSSIGS KGLD-VLGGV IDEGYRGEIG VIMINVSRKS ITLMERQKIA QLIIL
EIAV    RDEDAGFDLC VPYDIMIPVS D-TKIIPTDV KIQVPPNSFG WVTGKSSMAK QGLL-INGGI IDEGYTGEIQ VICTNIGKSN IKLIEGQKFA QLIIL
VISNA   RAEDAGYDLI CPQEISIPAG Q-VKRIAIDL KINLKKDQWA MIGTKSSFAN KGVF-VQGGI IDSGYQGTIQ VVIYNSNNKE VVIPQGRKFA QLILM
CAEV    REEDAGYDLI CPEEVTIEPG Q-VKCIPIEL RLNLKKSQWA MIATKSSMAA KGVF-TQGGI IDSGYQGQIQ VIMYNSNKIA VVIPQGRKFA QLILM
Human   SARAAGYDLY SAYDYTIPPM E-KAVVKTDI QIALPSGCYG RVAPRSGLAA KHFIDVGAGV IDEDYRGNVG VVLFNFGKEK FEVKKGDRIA QLICE
VV      SPGAAGYDLY SAYDYTIPPG E-RQLIKTDI SMSMPKICYG RIAPRSGLSL KG-IDIGGGV IDEDYRGNIG VILINNGKYT FNVNTGDRIA QLIYQ
HSV-1   SPGSAGFDLS VLEDREFIRG -CHYRLPTGL AIAVPRGYVG IITPRSSQAK NFV-ST--GI IDSDFRGHIH -IMVSAIADF -SVKKNQRIA QLVVT
```

FIG. 3. Amino acid homologies between HERV-L dUTPase and various viral or cellular dUTPase sequences. Amino acids are shaded only when identical to those in HERV-L. The dUTPase domain in HERV-L corresponds to nucleotides 5266 to 5545. Abbreviations and sources for the other sequences are as follows: MPMV, Mason-Pfizer monkey virus (52); MMTV, mouse mammary tumor virus (39); FIV, feline immunodeficiency virus (54); EIAV, equine infectious anemia virus (21); VISNA, visna lentivirus (53); CAEV, caprine arthritis encephalitis virus (46); Human, human dUTPase (37); VV, vaccinia virus (14); and HSV-1, herpes simplex virus type 1 (10).

A

```
                    * *       *       *   * * *
HERV-L   KNAGVVIP ITSLFNSPFW PVQKTDGS-W RMRVDYHKLN QVVTPIAAAI
HFV      LKQGVLTP QNSTMNTPVY PVPKPDGR-W RMVLDYREVN KTIPLTAAQN
SFV-1    LKQGVLIQ QNSTMNTPVY PVPKPDGK-W RMVLDYREVN KTIPLIAAQN
HERV-E   RTFRIIVP CQSPWNTPLL PVPKPGTKDY RPVQDLRLVN QATVTLHPTV
MoMLV    LDQGILVP CQSPWNTPLL PVKKPGTNDY RPVQDLREVN KRVEDIHPTV
MPMV     LEAGHITE SSSPWNTPIF VIKKK-SGKW RLLQDLRAVN ATMVLMGALQ
MMTV     LQLGHLEE SNSPWNTPVF VIKKK-SGKW RLLQDLRAVN ATMHDMGALQ
HERV-K   LEKGHIEP SFSPWNSPVF VIQKK-SGKW HTLTDLRAVN AVIQPMGPLQ
RSV      LQLGHIEP SLSCWNTPVF VIRKA-SGSY RLLHDLRAVN AKLVPFGAVQ
HTLV-1   LEAGHIEP YTGPGNNPVF PVKKA-NGTW RFIHDLRATN SLTIDLSSSS
HIV-1    EGKISKIG PENPYNTPVF AIKKKDSTKW RKLVDFRELN KRTQDFWEVQ

                          *                   **
         QDVVSLLKQI NTSPGTLYAA IDLANAFY-- PFHKPHQKQF AFSWQGQ---
         QHSAGILATI -VR-QKYKTT LDLANGFWAH PITPESYWLT AFTWQGK---
         QHSAGILSSI -YR-GKYKTT LDLTNGFWAH PITPESYWLT AFTWQGK---
         PNLYTLLGLL PAE-DSWFTC LDLKDAFFSI RLAPERQKLF AFQWEDPE-S
         PNPYNLLSGL PPS-HQWYTV LDLKDAFFCL RLHPTSQPLF AFEWRDPE-M
         PGLPSPVAI- -PQGYLKI-I IDLKDCFFSI PLHPSDQKRF AFSLPSTNFK
         PGLPSPVAV- -PKGWEII-I IDLQDCFFNI KLHPEDCKRF AFSVPSPNFK
         PGLPSPAMI- -PKDWPLI-I IDLKDCFFTI PLAEQDCEKF AFTIPAINNK
         QGAPVLSAL- -PRGWPLM-V LDLKDCFFSI PLAEQDREAF AFTLPSVNNQ
         PGPPDLSSL- -PTTLAHLQT IDLRDAFFQI PLPKQFQPYF AFTVPQQCNY
         LGIPHPAGL- -KKKKS-VTV LDVGDAYFSV PLDEDFRKYT AFTIPSINNE
                     ***   **                          * **
         ----QCIFTV LPQGYINSPA LCHNLIQSDL DHFSLPQ-DI TLVHYIDDIM
         ----QYCWTR LPQGFLNSPA LFTADVVDLL KEIPNVQV-- ----YVDDIY
         ----QYCWTR LPQGFLNSPA LFTADVVDLL KEIPNVQA-- ----YVDDIY
         GVTTQYTWTQ LPQRFKNSPT IFGEALARDL QKFPTRDLGC VLLQYVDDLL
         GISGQLTWTR LPQGFKNSPT LFDEALHRDL ADFRIQHPDL ILLQYVDDLL
         EPMQRFQWKV LPQGMANSPT LCQKYVATAI HKVRHAWKQM YIIHYMDDIL
         RPYQRFQWKV LPQGMKNSPT LCQKFVDKAI LTVRDKYQDS YIVHYMDDIL
         EPATRFQWKV LPQGMLNSPT ICQTFVGRAL QPVREKFSDC YIIHYIDDIL
         APARRFQWKV LPQGMTCSPT ICQLVVGQVL EPLRLKHPSL CMLHYMDDLL
         GPGTRYAWKV LPQGFKNSPT LFEMQLAHIL QPIRQAFPQC TILQYMDDIL
         TPGIRYQYNV LPQGWKGSPA IFQSSMTKIL EPFRKQNPDI VIYQYMDDLY

                             *       *       **
         LIGSNEQEVA NT-LDLLVRH LHAIGWEINP NKIHAPSTSV KFLGSSG
         LSHDDPKEHV QQ-LEKVFQI LLQAGYVVSL KKSEIGQKTV EFLGFNI
         ISHDDPQEHL EQ-LEKIFSI LLNAGYVVSL KKSEIAQREV EFLGFNI
         LGHPTAVGWP RE-QMLYSGT WRTVGIRCPR KKAQICRQQV CYLGFTI
         LAATSELDCQ QG-TRALLQT LGNLGYRASA KKAQICQKQV KYLGYLL
         IAGK-DGQQV LQCFDQLKQE LTAAGLHIAP EKVQLQDP-Y TYLGFEL
         LAHP-SRSIV DEILTSMIQA LNKHGLVVST EKIQKYDN-L KYLGTHI
         CAAE-TKDKL IDCYTFLQAE VANAGLAIAS DKIQTSTP-F HYLGMQI
         LAAS-SHDGL EAAGEEVIST LERAGFTISP DKVQREPG-V QYLGYKL
         LASP-SHEDL LLLSEATMAS LISHGLPVSE NKTQQTPGTI KFLGQII
         VGSDLEIGQH RTKIEELRQH LLRWGLTTPD KKHQKEPPFL -WMGYEL
```

B

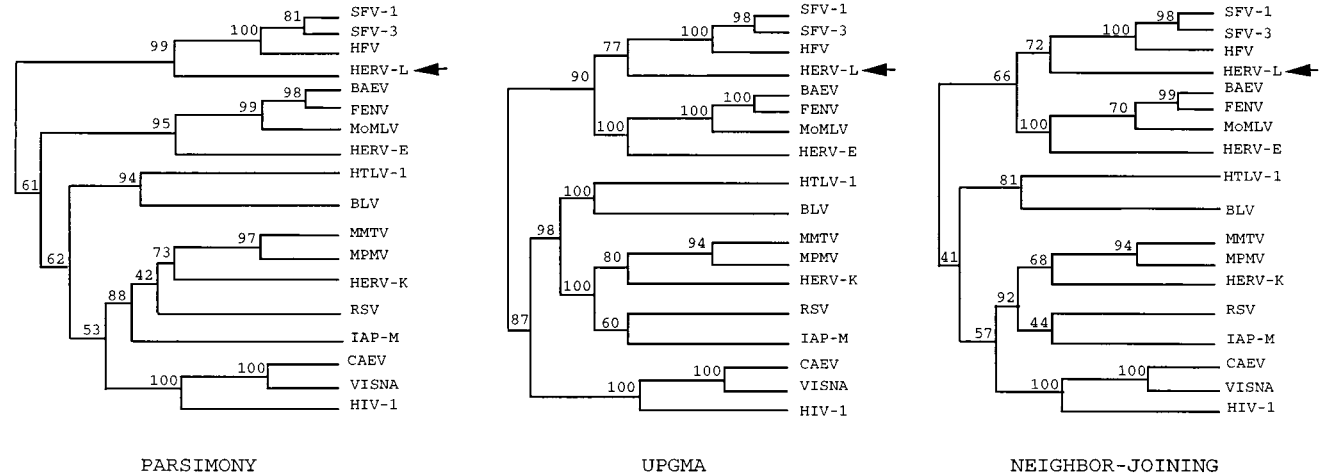

FIG. 4. Homologies between reverse transcriptases of HERV-L and other retroviral sequences. Phylogenetic analyses. (A) Alignment of reverse transcriptase amino acids (shaded only when identical to those in HERV-L); the asterisks at the top of the alignment indicate residues unvariant among retroviruses (57). The HERV-L reverse transcriptase domain corresponds to nucleotides 2784 to 3333, with two single-amino-acid modifications: L and G in clone 10 were changed into P (position 110) and D (position 146), respectively, as systematically found in five other sequenced clones. (B) Consensus phylogenetic trees were obtained from the alignment shown in panel A (with additional sequences) by the maximum parsimony method (left), the unweighted-pair-group method (UPGMA; middle), and the neighbor-joining method (right). The values at the branch points indicate the percentage of bootstrapped trees supporting each node; branch lengths are arbitrary. Abbreviations and sources of the sequences are as follows: SFV-1, simian foamy virus type 1 (24); MoMLV, Moloney murine leukemia virus (51); FENV, feline endogenous virus ECE1 (Swiss-Prot database, accession no. P31792); BAEV, baboon endogenous virus (20); MPMV, Mason-Pfizer monkey virus (52); MMTV, mouse mammary tumor virus (39); RSV, Rous sarcoma virus (48); IAP-M, mouse intracisternal A particle (38); HTLV-1, human T-cell leukemia virus type 1 (49); BLV, bovine leukemia virus (9); HIV-1, human immunodeficiency virus type 1 (55); VISNA, visna lentivirus (53); CAEV, caprine arthritis encephalitis virus (46); HERV-E, human endogenous retrovirus E (44); and HERV-K, human endogenous retrovirus K-10 (42). The HFV sequence is from reference 32, and the SFV-3 sequence is from reference 43.

transcriptase sequences from other endogenous and exogenous retroviruses (Fig. 4A) by using the CLUSTALV program. As illustrated in Fig. 4A, most of the highly conserved residues in retroelements (denoted with asterisks) are indeed present in HERV-L. Interestingly, optimum alignment of the HERV-L sequence required the introduction of a gap of 7 residues at positions 96 to 102, exactly (and exclusively) as found for the foamy viruses. Phylogenetic trees based on this multiple alignment (including additional sequences) were constructed by either "distance" methods, such as the unweighted-pair-group and neighbor-joining methods, or the maximum parsimony method, from the PHYLIP package, version 3.52c (13). With the two latter methods, we used as an outgroup the Ty1 element (7) that belongs to the most distantly related group of retroid elements. The trees had the same overall topology with parsimony and distance analyses (Fig. 4B) and were consistent with those already described (35, 58). One difference between the two types of methods concerns the location of foamy viruses, which are among the most distantly related elements within the phylogenetic trees, being placed either as the most distant member of the type C group (distance methods) or as a separate branch of retroviruses (parsimony method). An interesting outcome of the phylogenetic analysis, then, is that whatever the method, HERV-L is found to be associated with the foamy viruses, with bootstrap scores strongly supporting this relationship (Fig. 4B). This suggests that despite important differences—implicating gain and/or loss of specific genes—HERV-L and foamy viruses might have a common evolutionary history.

Distribution of HERV-L-like sequences among eukaryotic species. To elucidate the organization of HERV-L elements in the genomes of different species, *Eco*RI-digested cellular DNAs of various origins (zoo-blot; Clontech) were probed with a 360-bp *pol* sequence from HERV-L. As illustrated in Fig. 5, hybridization with HERV-L sequences could be detected under moderately stringent conditions with all mammals tested (including the rabbit, with longer exposure), whereas no HERV-L *pol*-related sequence was detected in the DNA of chicken cells and *Saccharomyces cerevisiae*. This result might suggest that HERV-L-related sequences were present early in the divergence of the mammalian branch. This is consistent with the results of PCR analyses using primers (underlined in Fig. 2, box B) selected from regions of the reverse transcriptase sequence that are conserved among four genomic phage DNAs analyzed: salmon, drosophila, and yeast DNAs were negative by this test, whereas human, simian, murine, and feline DNAs were positive (data not shown). The hybridization signals shown in Fig. 5 are much more intense for human, monkey, and mouse DNAs than for those of other species. This difference might reflect the degree of sequence divergence or differences in copy number. From dot blot hybridizations (data not shown), the number of gene copies hybridizable with the *pol* gene was estimated to be 200 per haploid human genome (as measured in both human peripheral lymphocytes and HeLa cells) and at least 100 per haploid mouse genome (3T3 cells and BALB/c mouse genomic DNA).

The pattern of hybridization of mouse cellular DNA resolves into a small number of cross-hybridizing bands with a major 2.3-kb *Eco*RI fragment, which further persists after a high-stringency wash (0.1× SSC–0.1% sodium dodecyl sulfate, 30 min, 65°C). The same pattern was obtained with cellular DNA of all laboratory mouse strains tested, implying that the HERV-L-related sequences in these species might represent a family of relatively homogeneous, well-conserved units (whether the same applies to wild mice remains to be determined). The occurrence of a high copy number of HERV-L related

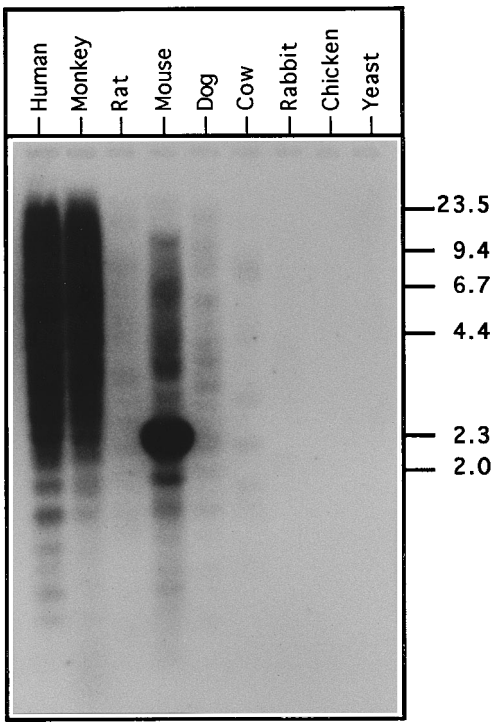

FIG. 5. Southern blot analysis of HERV-L *pol*-related sequences in various eukaryotic species. A blot containing *Eco*RI-digested DNA from the indicated species (5 μg per lane; Clontech) was hybridized under standard conditions (6) with a ^{32}P-labeled nick-translated 360-bp DNA fragment from phage 10 (obtained by PCR using the primers indicated in Fig. 2). The filter was washed under moderately stringent conditions (15 min, 0.5× SSC–0.1% sodium dodecyl sulfate, 65°C), and exposed for 24 h; positions of standard size markers (in kilobases) are indicated. Monkey, rhesus monkey; Rat, Sprague-Dawley rat; Yeast, *S. cerevisiae.*

sequences in the mouse lineage—not observed in the rat—raises interesting questions relative to the history of the HERV-L elements. Amplification of HERV-L in the mouse genome should actually have occurred after the "recent" divergence between rats and mice (<35 million years ago [17]), a result not simply compatible with the occurrence of HERV-L sequences in primates. An investigation of HERV-L-related sequences in the murine genome—which already disclosed >75% homology between human and murine sequences within a 360-bp PCR-amplified *pol* domain—is currently being conducted to gain insight into the mechanism of invasion: retrotransposition of a competent HERV-L element could have occurred independently in the mouse and primate genomes, from an ancestral sequence common to both phyla, or alternatively amplification could have occurred after an interspecies horizontal transfer (in that case possibly from primates to mice). In this respect, the identification of competent HERV-L-like elements in the murine genome could be easier than the identification of such elements in the human genome, all the more so as the former discloses a rather homogeneous pattern in Southern blot analysis.

In conclusion, we have identified a new family of endogenous retrovirus-like elements—the HERV-L elements—which are widespread within the human genome (approximately 200 copies). Amino acid sequences and phylogenetic analyses based on *pol* genes indicate that HERV-L is most closely related to the foamy retroviruses. As such, this retrovirus-like endogenous element might be an ancestor for the present-day infectious mammalian foamy viruses. Interestingly, and as ob-

served for some infectious retroviruses, HERV-L has acquired a dUTPase gene, but at a distinct location. The presence of HERV-L-related elements at a high copy number (at least 100 copies) within the murine genome and the analysis of their evolutionary relationships could finally provide insights into the interspecies horizontal transfers—or alternatively into the intraspecies independent amplifications—of retroviruses and retrovirus-like elements.

Nucleotide sequence accession number. The HERV-L sequence has been entered in the EMBL database under the number X89211.

We are grateful to L. Chakrabarti, F. Tekaïa, and P. Dessen for helpful discussions and computer assistance; to A. Dejean for providing the human genomic library; and to C. Lavialle for critical reading of the manuscript.

This work was supported by the CNRS and INSERM and by grants from the ARC (contract 6552) and Rhone-Poulenc Rorer ("Bioavenir" contract to T.H. and to J.-F.C.).

REFERENCES

1. **Bergman, A. C., O. Björnberg, J. Nord, P. O. Nyman, and A. M. Rosengren.** 1994. The protein p30, encoded at the gag-pro junction of mouse mammary tumor virus, is a dUTPase fused with a nucleocapsid protein. Virology **204:**420–424.
2. **Bohannon, R. C., L. A. Donehower, and R. J. Ford.** 1991. Isolation of a type D retrovirus from B-cell lymphomas of a patient with AIDS. J. Virol. **65:** 5663–5672.
3. **Boller, K., H. Frank, J. Löwer, R. Löwer, and R. Kurth.** 1983. Structural organisation of unique retrovirus-like particles budding from human teratocarcinoma cell lines. J. Gen. Virol. **64:**2549–2554.
4. **Bonner, T. I., C. O'Connell, and M. Cohen.** 1982. Cloned endogenous retroviral sequences from human DNA. Proc. Natl. Acad. Sci. USA **79:**4709–4713.
5. **Callahan, R., W. Drohan, S. Tronick, and J. Schlom.** 1982. Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. Proc. Natl. Acad. Sci. USA **79:**5503–5507.
6. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:**1991–1995.
7. **Clare, J., and P. Farabaugh.** 1985. Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. Proc. Natl. Acad. Sci. USA **82:**2829–2833.
8. **Coffin, J. M.** 1993. Reverse transcription and evolution, p. 445–479. In A. M. Skalka and S. P. Goff (ed.), Reverse transcriptase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
9. **Coulston, J., H. Naif, R. Brandon, S. Kumar, S. Khan, R. C. W. Daniel, and M. F. Lavin.** 1990. Molecular cloning and sequencing of an Australian isolate of proviral bovine leukaemia virus DNA: comparison with other isolates. J. Gen. Virol. **71:**1737–1746.
10. **Davison, A. J.** 1992. Channel catfish virus: a new type of herpesvirus. Virology **186:**9–14.
11. **Deen, K. C., and R. W. Sweet.** 1986. Murine mammary tumor virus *pol*-related sequences in human DNA: characterization and sequence comparison with the complete murine mammary tumor virus *pol* gene. J. Virol. **57:**422–432.
12. **Donehower, L. A., R. C. Bohannon, R. J. Ford, and R. A. Gibbs.** 1990. The use of primers from highly conserved *pol* regions to identify uncharacterized retrovirus by the polymerase chain reaction. J. Virol. Methods **28:**33–46.
13. **Felsenstein, J.** 1989. PHYLIP—phylogeny inference package. Cladistics **5:**164–166.
14. **Goebel, S. J., G. P. Johnson, M. E. Perkus, S. W. Davis, J. P. Winslow, and E. Paoletti.** 1990. The complete DNA sequence of vaccinia virus. Virology **179:**247–266.
15. **Harada, F., N. Tsukada, and N. Kato.** 1987. Isolation of three kinds of human endogenous retrovirus-like sequences using tRNA (Pro) as a probe. Nucleic Acids Res. **15:**9153–9162.
16. **Higgins, D. G., A. J. Bleasby, and R. Fuchs.** 1992. ClustalV: improved software for multiple sequence alignment. Comput. Appl. Biosci. **8:**189–191.
17. **Janke, A., G. Feldmaier-Fuchs, W. K. Thomas, A. von Haeseler, and S. Paabo.** 1994. The marsupial mitochondrial genome and the evolution of placental mammals. Genetics **137:**243–256.
18. **Johnson, M. S., M. A. McClure, D.-F. Feng, J. Gray, and R. F. Doolittle.** 1986. Computer analysis of retroviral *pol* genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. Proc. Natl. Acad. Sci. USA **83:**7648–7652.
19. **Kalter, S. S., R. J. Helmke, R. L. Heberling, M. Panigel, A. K. Fowler, J. E. Strickland, and A. Hellman.** 1973. C-type particles in normal human placentas. J. Natl. Cancer Inst. **50:**1081–1084.
20. **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.
21. **Kawakami, T., L. Sherman, J. Dahlberg, A. Gazit, A. Yaniv, S. R. Tronick, and S. A. Aaronson.** 1987. Nucleotide sequence analysis of equine infectious anemia virus proviral DNA. Virology **158:**300–312.
22. **Köppe, B., L. Menendez-Arias, and S. Oroszlan.** 1994. Expression and purification of the mouse mammary tumor virus *gag-pro* transframe protein p30 and characterization of its dUTPase activity. J. Virol. **68:**2313–2319.
23. **Kulkoski, J., K. S. Jones, R. A. Katz, J. P. G. Mack, and A. M. Skalka.** 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. **12:**2331–2338.
24. **Kupiec, J.-J., J. A. Kay, M. Hayat, R. Ravier, G. Peries, and F. Galibert.** 1991. Sequence analysis of the simian foamy virus type 1 genome. Gene **101:**185–194.
25. **La Mantia, G., G. Pengue, D. Maglione, A. Pannuti, A. Pascucci, and L. Lania.** 1989. Identification of new human repetitive sequences: characterization of the corresponding cDNAs and their expression in embryonal carcinoma cells. Nucleic Acids Res. **17:**5913–5922.
26. **Larsson, E., N. Kato, and M. Cohen.** 1989. Human endogenous proviruses. Curr. Top. Microbiol. Immunol. **148:**115–132.
27. **Leib-Mösch, C., R. Brack, T. Werner, V. Erfel, and R. Hehlmann.** 1986. Isolation of an SSAV-related endogenous sequence from human DNA. Virology **155:**666–677.
28. **Löwer, R., K. Boller, B. Hasenmaier, C. Korbmacher, N. Mueller-Lantzsch, J. Löwer, and R. Kurth.** 1993. Identification of human endogenous retroviruses with complex mRNA expression and particle formation. Proc. Natl. Acad. Sci. USA **90:**4480–4484.
29. **Maeda, N.** 1985. Nucleotide sequence of the haptoglobin and haptoglobin-related gene pair: the haptoglobin-related gene contains a retrovirus-like element. J. Biol. Chem. **260:**6698–6709.
30. **Mager, D. L., and P. S. Henthorn.** 1984. Identification of a retrovirus-like repetitive element in human DNA. Proc. Natl. Acad. Sci. USA **81:**7510–7514.
31. **Martin, M. A., T. Bryan, S. Rasheed, and A. S. Khan.** 1981. Identification and cloning of endogenous retroviral sequences present in human DNA. Proc. Natl. Acad. Sci. USA **78:**4892–4896.
32. **Maurer, B., H. Bannert, G. Darai, and R. M. Flügel.** 1988. Analysis of the primary structure of the long terminal repeat and the *gag* and *pol* genes of the human spumaretrovirus. J. Virol. **62:**1590–1597.
33. **Maurer, B., E. Serfling, V. Meulen, and A. Rethwilm.** 1991. Transcription factor AP-1 modulates the activity of the human foamy virus long terminal repeat. J. Virol. **65:**6353–6357.
34. **May, F. E. B., B. R. Westley, H. Rochefort, E. Buetti, and H. Diggelmann.** 1983. Mouse mammary tumor virus related sequences are present in human DNA. Nucleic Acids Res. **11:**4127–4139.
35. **McClure, M. A., M. S. Johnson, D. F. Feng, and R. F. Doolittle.** 1988. Sequence comparisons of retroviral proteins: relative rates of change and general phylogeny. Proc. Natl. Acad. Sci. USA **85:**2469–2473.
36. **McGeoch, D. C.** 1990. Protein sequence comparisons show that the "pseudoproteases" encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. Nucleic Acids Res. **18:**4105–4110.
37. **McIntosh, E. M., D. D. Ager, M. H. Gadsen, and R. H. Haynes.** 1992. Human dUTP pyrophosphatase: cDNA sequence and potential biological importance of the enzyme. Proc. Natl. Acad. Sci. USA **89:**8020–8024.
38. **Mietz, J. A., Z. Grossman, K. K. Lueders, and E. L. Kuff.** 1987. Nucleotide sequence of a complete mouse intracisternal A-particle genome: relationship to known aspects of particle assembly and function. J. Virol. **61:**3020–3029.
39. **Moore, R., M. Dixon, R. Smith, G. Peters, and C. Dickson.** 1987. Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events are required for translation of *gag* and *pol*. J. Virol. **61:**480–490.
40. **O'Brien, S. J., T. I. Bonner, M. I. Cohen, C. O'Connell, and W. Nash.** 1983. Mapping of an endogenous retroviral sequence to human chromosome 18. Nature (London) **303:**74–77.
41. **O'Connell, C., S. J. O'Brien, W. G. Nash, and M. Cohen.** 1984. ERV-3, a full-length human endogenous provirus: chromosomal localization and evolutionary relationships. Virology **138:**225–235.
42. **Ono, M., M. T. Yasunaga, T. Miyata, and H. Ushikubo.** 1986. Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. J. Virol. **60:**589–598.
43. **Renne, R., E. Friedl, M. Schweizer, U. Fleps, R. Turek, and D. Neumann-Haefelin.** 1992. Genomic organisation and expression of simian foamy virus type 3 (SFV-3). Virology **186:**597–608.
44. **Repaske, R., P. Steele, R. O'Neill, A. Rabson, and M. Martin.** 1985. Nucleotide sequence of a full-length human endogenous retroviral segment. J. Virol. **54:**764–772.
45. **Ross, B. M., J. E. Looney, and J. D. Harding.** 1986. Nucleotide sequence of a mouse tRNA Leu gene. Nucleic Acids Res. **13:**5567.
46. **Saltarelli, M., G. Querat, D. A. M. Konings, R. Vigne, and J. E. Clements.**

1990. Nucleotide sequence and transcriptional analysis of molecular clones of CAEV which generate infectious virus. Virology **179:**347–364.
47. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
48. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32:**853–869.
49. **Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida.** 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA **80:**3618–3622.
50. **Shih, A., R. Misra, and M. G. Rush.** 1989. Detection of multiple, novel reverse transcriptase coding sequences in human nucleic acids: relation to primate retroviruses. J. Virol. **63:**64–75.
51. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293:**543–548.
52. **Sonigo, P., C. Barker, E. Hunter, and S. Wain-Hobson.** 1986. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. Cell **45:**375–385.
53. **Staskus, K. A., E. F. Retzel, E. D. Lewis, J. L. Silsby, S. St. Cyr, J. M. Rank, S. W. Wietgrefe, A. T. Haase, R. Cook, D. Fast, P. T. Geiser, J. T. Harty, S. H. Kong, C. J. Lahti, T. P. Neufeld, T. E. Porter, E. Shoop, and K. R. Zachow.** 1991. Isolation of replication-competent molecular clones of visna virus. Virology **181:**228–240.
54. **Talbott, R. L., E. E. Sparger, K. M. Lovelace, W. M. Fitch, N. C. Pedersen, P. A. Luciw, and J. H. Elder.** 1989. Nucleotide sequence and genomic organization of feline immunodeficiency virus. Proc. Natl. Acad. Sci. USA **86:**5743–5747.
55. **Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon.** 1985. Nucleotide sequence of the AIDS virus, LAV. Cell **40:**9–17.
56. **Wilkinson, D. A., D. L. Mager, and J. A. C. Leong.** 1994. Endogenous human retroviruses, p. 465–535. *In* J. A. Levy (ed.), The Retroviridae. Plenum Press, New York.
57. **Xiong, Y., and T. H. Eickbush.** 1988. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol. Biol. Evol. **5:**675–690.
58. **Xiong, Y., and T. H. Eickbush.** 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9:**3353–3362.