

Detection of Multiple, Novel Reverse Transcriptase Coding Sequences in Human Nucleic Acids: Relation to Primate Retroviruses

ANDY SHIH, RAVI MISRA, AND MARK G. RUSH*

Department of Biochemistry, New York University School of Medicine, New York, New York 10016

Received 30 June 1988/Accepted 8 September 1988

A variety of chemically synthesized oligonucleotides designed on the basis of amino acid and/or nucleotide sequence data were used to detect a large number of novel reverse transcriptase coding sequences in human and mouse DNAs. Procedures involving Southern blotting, library screening, and the polymerase chain reaction were all used to detect such sequences; the polymerase chain reaction was the most rapid and productive approach. In the polymerase chain reaction, oligonucleotide mixtures based on consensus sequence homologies to reverse transcriptase coding sequences and unique oligonucleotides containing perfect homology to the coding sequences of human T-cell leukemia virus types I and II were both effective in amplifying reverse transcriptase-related DNA. It is shown that human DNA contains a wide spectrum of retrovirus-related reverse transcriptase coding sequences, including some that are clearly related to human T-cell leukemia virus types I and II, some that are related to the L-1 family of long interspersed nucleotide sequences, and others that are related to previously described human endogenous proviral DNAs. In addition, human T-cell leukemia virus type I-related sequences appear to be transcribed in both normal human T cells and in a cell line derived from a human teratocarcinoma.

Two major classes of eucaryotic dispersed repetitive sequences, retrotransposons (retroviruslike transposable elements) and retroposons (short and long interspersed nucleotide sequences as well as processed pseudogenes), appear to transpose by a mechanism involving the reverse transcription of an RNA intermediate (4, 47, 65). In retrotransposons (such as vertebrate retrovirus proviral DNAs, mouse intracisternal A particle [IAP] proviral DNAs, *Drosophila copia* and *copialike* DNAs, and yeast Ty DNAs), the reverse transcriptase required for transposition is clearly a transposon-coded function, whereas in most retroposons (such as the primate *Alu* family of short interspersed nucleotide sequences and processed pseudogenes in general) the reverse transcriptase required for transposition is assumed to be supplied from retroviruses and/or retroviruslike transposable elements. However, one class of retroposons, the L-1 family of mammalian long interspersed nucleotide sequences, almost certainly contains some members that possess reverse transcriptase coding capacity (13, 23, 54). The potential importance of retrotransposons and retroposons in regard to the evolution of the mammalian genome and to mutagenesis and oncogenesis (64) has been most thoroughly studied in mice, where the total mass of these repetitive sequences represents at least 10% of the genome (4, 47, 60, 65). Although short interspersed nucleotide sequences (hundreds of thousands of copies per cell) and long interspersed nucleotide sequences (tens of thousands of copies per cell) account for most of this material, endogeneous C, B, IAP, and viruslike 30S-type proviral DNAs (here considered as retrotransposons) are also present at significant levels (thousands of copies per cell). Indeed, it is these proviral DNAs and the viruses that some of them generate that have been the most thoroughly characterized eucaryotic transposable elements at the genetic and biochemical levels, and it is

possible that similar proviral DNAs may be amenable to serious functional investigation in humans.

A few investigators have recently begun to search successfully for retroviruslike and/or proviral DNAs in humans. In essence, these searches have been based primarily on three approaches: (i) the low-stringency screening of human genomic libraries with proviral DNA probes of murine leukemia virus (MLV) (27, 44, 45), murine mammary tumor virus (MMTV) (1, 2), baboon or chimpanzee virus (33, 34), and IAPs (35); (ii) the screening of human genomic libraries with oligonucleotide probes homologous to the primer-binding sites of known mammalian retroviruses (12, 18); and (iii) the fortuitous identification of previously uncharacterized repetitive elements in genomic clones followed by subcloning and sequencing (24, 25, 39). All of these approaches have been successful, and the results may be briefly summarized by stating the following: (i) the elements identified to date are significantly different from each other; (ii) six of the elements that have been extensively or completely sequenced each use a different tRNA for putative priming of reverse transcription; (iii) some of the elements clearly contain all of the information associated with proviral DNA (long terminal repeats and *gag*, *pol*, and *env* regions), whereas others contain extensive deletions; (iv) copy numbers of different elements range from a single copy to tens of thousands; (v) many of the elements are transcribed (7, 9, 14, 36, 39, 42); and (vi) the long terminal repeats of most of the sequenced elements contain all of the regulatory sequences generally associated with their function as both promoters and terminators of transcription, although *gag* and *pol* genes have acquired deletions or accumulated additional termination codons within their predicted functional open reading frames. (We have adopted the terminology of Kröger and Horak [18], in which human retrovirus-related sequences [HuRRS] are classified on the basis of putative tRNA primer-binding sites. Thus HuRRS-E [45] and HuRRS-K

* Corresponding author.

MATERIALS AND METHODS

Standard procedures. Restriction endonuclease digestions, gel electrophoresis, Southern and Northern (RNA) blotting, hybridization of blots to oligonucleotide-primed labeled probes, labeling of synthetic oligonucleotides with polynucleotide kinase, and preparation of high-molecular-weight chromosomal and bacteriophage lambda DNAs were performed as previously described (30) and/or according to standard procedures (26). Sequencing of single-stranded M13 cloned DNAs was performed by the dideoxy-chain terminator method with a modified T7 polymerase enzyme (Sequenase) and [α - 35 S]dATP with the reagents and protocol of the United States Biochemical Corp. All sequences were obtained from multiple gel readings. Experiments involving the probing of blotted gels containing PCR reaction mixtures (or total genomic DNA) always included stringency and copy number controls. Specifically, random oligonucleotide-primed probes synthesized from single-stranded M13 clones were hybridized to both blotted gels and control dots of filter-bound complementary M13 RF, and both blots and dots were washed in 0.03 M NaCl–3 mM trisodium citrate–0.1% sodium dodecyl sulfate at a series of temperatures ranging from 50 to 70°C at 5°C intervals. For the experiments shown in Fig. 6 and 7, blots and controls exhibited comparable losses in signal intensity, indicating that probes were extensively homologous to their complementary blotted DNAs and, in the case of PCR mixtures, not due solely to the 30% expected primer homology. The filters autoradiographed for Fig. 6 were washed at 60°C, and those for Fig. 7 were washed at 55°C.

High-molecular-weight DNAs. High-molecular-weight DNAs isolated from livers of C57L/J and AKR/J mice were obtained from P. D'Eustachio (New York University, Department of Biochemistry), as was DNA isolated from the A9 mouse cell line (derived from strain C3H); DNA isolated from spleens of RF/J mice was from S. Sloan (New York University, Department of Pathology); DNA isolated from the 293 human cell line (11) was from R. Schneider (New York University, Department of Biochemistry); DNA isolated from the CEM human T-cell lymphoblastoid cell line was from J. Krolewski (New York University, Department of Pathology); and DNA isolated from human scalp skin and the human melanoma cell lines BSFM and HM20 were from D. Zouzas (New York University, Department of Dermatology).

Genomic and cDNA libraries. A human genomic placental library was obtained from J. Hall (New York University, Department of Biochemistry), a human teratocarcinoma (TC)-derived cell line cDNA library (54, 55) was from J. Skowronski (Cold Spring Harbor Laboratory), and a human T-cell (PTL)-derived cell line cDNA library was from J. Krolewski. All of these bacteriophage lambda libraries were plated on *Escherichia coli* LE392 and screened with oligonucleotide probes by the hybridization procedure described below.

Total and poly(A)⁺ RNAs. Total and poly(A)⁺ RNAs isolated from the Burkitt lymphoma cell line ST486, were obtained from J. Krolewski. HeLa total poly(A)⁺ RNA was isolated by the guanidinium isothiocyanate procedure of Maniatis et al. (26).

Hybridization of oligonucleotides to filter-immobilized DNAs. A low-stringency hybridization (37°C in 0.9 M NaCl, 0.09 M trisodium citrate, 10% dextran sulfate) followed by high-stringency washes in the presence of 3 M tetramethyl ammonium chloride (a compound that eliminates the effect of base composition on duplex DNA melting temperatures)

was used in all studies exactly as described by Wood et al. (66). In addition, controls of Moloney murine leukemia virus proviral DNA were included in most experiments (one of the mixed probe sequences is about 95% homologous to the known conserved Moloney 27-base-pair region) as a rough measure of probe homology, at least as determined by comparative reductions in signal intensity as a function of washing temperatures (60 to 75°C). Hybridizations were performed in 10-ml solutions containing 200 ng of 5' 32 P-labeled oligonucleotides with specific activities ranging from 2×10^8 to 1×10^9 cpm/ μ g. For the experiments shown in Fig. 2 and 3 the washing temperature was 60°C.

PCR. PCRs were performed essentially as described by Saiki et al. (48) with *Thermus aquaticus* DNA polymerase (Perkin Elmer Cetus) to increase reaction specificity (since chain extension occurs at 72°C) as well as to reduce the overall time of amplification (since this enzyme is stable at a 94°C denaturation temperature). In short, reactions were performed in a DNA Thermal Cycler (Perkin Elmer Cetus) in 100- μ l solutions containing 1 to 2 μ g of DNA, 50 mM KCl, 10 mM Tris (pH 8.4), 2.5 mM MgCl₂, 1 μ M each unique primer or 10 μ M each mixed oligonucleotide primer (MOP), 200 μ M each deoxyribonucleoside triphosphate, 200 μ g of gelatin per ml, and 2 U of polymerase. Samples overlaid with an equal volume of mineral oil were subjected to 30 cycles of amplification by (i) heating from 72 to 94°C (over 30 s) and incubating at 94°C for 2 min (denaturation), (ii) cooling (over 1 min) from 94°C to a primer annealing temperature of between 37 and 60°C (depending on the experiment) and incubating for 3 min, and (iii) heating from the annealing temperature (over 1 min) to a chain extension temperature of 72°C and incubating for 4 min. The final extension was incubated for 7 min, and 15- μ l samples were then resolved on composite gels of 3% NuSieve–1% electrophoresis grade agarose (Schwartz/Mann) containing 0.5 μ g of ethidium bromide per ml. Fragments to be subcloned (20- μ l samples) were first phosphorylated by adding 20 U (2 μ l) T4 polynucleotide kinase and 1 μ l of 50 mM ATP, and incubating for 45 min at 37°C. After ethanol precipitation and suspension in 20 μ l of H₂O, the fragments were cloned by blunt-end ligation (10 U of T4 DNA ligase) into the *Sma*I site of alkaline phosphatase-treated M13mp19.

Comparison of amino acid sequences. Computer analysis of sequence data was performed on a Digital VAX 11/750 microcomputer by using the protein information resource (40, 41) and/or Staden (58, 59) nucleic acid sequence programs.

Synthetic primers and probes. All oligonucleotides were synthesized in the laboratory of B. Goldschmidt (New York University, Department of Environmental Medicine) with an Applied Biosystems 380A Synthesizer.

RESULTS

Synthetic oligonucleotide primers and probes. In the interest of detecting new human endogenous retroviruses and retroviral related sequences that might lack extensive (kb) homologies with known retroviruses, two procedures with reverse transcriptase-specific synthetic oligonucleotides were employed. The first was based on conventional genomic library and Southern blot screenings, whereas the second was based on the use of PCR. As shown below, the PCR procedure represented a far more productive approach to characterizing the multigene reverse transcriptase family than did conventional screening and resulted in the identification of sequences that are apparently related to known

primate retroviruses. Figure 1 presents published nucleotide and amino acid sequences of 18 retroviral or retroviruslike reverse transcriptase genes in a region that contains a variety of particularly well-conserved, so-called diagnostic amino acids (38, 62, 63). Also included in this figure are the sequences of six synthetic oligonucleotides used as primers for PCR analyses of human (and in some cases, mouse) DNAs. One of these oligonucleotides (see below) was also used as a conventional probe for library and blot screening.

The primers represent putative (MOP) or actual (HTLV-I and HTLV-II) coding sequences for exceptionally well conserved amino acid blocks. The rationale for using unique HTLV primers was that human DNA might contain endogenous proviral DNAs related to human retroviruses, whereas the rationale for using MOP was that human DNA is already known to contain a variety of murine and other retrovirus-related sequences. The biases used for determining the sequence of the mixed primers, in which both the upstream and downstream mixtures are composed of 16 distinct oligonucleotides, were based both on theoretical considerations (20) and on the amino acid composition of the conserved reverse transcriptase regions in birds, mice, and humans. The downstream MOP contains 27 nucleotides rather than the 20 present in the other primers, since this mixed oligonucleotide (as well as its complement) was also designed for use as a probe for conventional screening of genomic and cDNA libraries.

The theories of probe selectivity and sensitivity upon which the latter screening was based indicated that rather long unique synthetic oligonucleotides (>25 residues), containing sequences based on primate codon usage and possessing about 85% homology to a target sequence, should be more than adequate for selecting true positive nonrepetitive clones from cDNA libraries. Slightly better homologies and/or longer probes are required for selecting true positive unique clones from genomic libraries. Our mixed 27-nucleotide oligomer represents a compromise between the use of a single long probe with about 85% target site homology (based on codon usage) and the use of a mixture of many smaller probes representing all possible coding possibilities. This compromise was adopted since the amino acid sequence(s) upon which the probe was based is itself a consensus. The expectation that nearly all of our target sequences should be present in repetitive elements, and that the selection of positive clones with at least 90% homology to our probe (estimated by stringency comparisons with a Moloney MLV proviral DNA control), were two additional factors that favored the selection of true positive clones.

Of the two complementary probes, one, 5' TAC $\frac{A}{G}$ TGGATGAC $\frac{A}{C}$ $\frac{C}{G}$ T-CTGCTGGCC $\frac{T}{G}$ GC 3', represents the sequence expected in an mRNA coding for reverse transcriptase and is known as the sense probe, whereas the other, 5' GG $\frac{A}{C}$ GGCCAGCAG $\frac{G}{C}$ $\frac{A}{G}$ T-GTCATCCA $\frac{T}{C}$ GTA 3', represents the complement of this and is known as the antisense probe.

Detection of conserved reverse transcriptase coding sequences by conventional screening. The detection of conserved reverse transcriptase coding sequences in Southern and Northern blots of human DNA and RNA is shown in Fig. 2 and 3. Also shown in Fig. 2 is the detection of such sequences in mouse DNAs. The results of the Southern blot analysis shown in Fig. 2 and of other Southern blot analyses, including additional mouse and human DNAs (data not

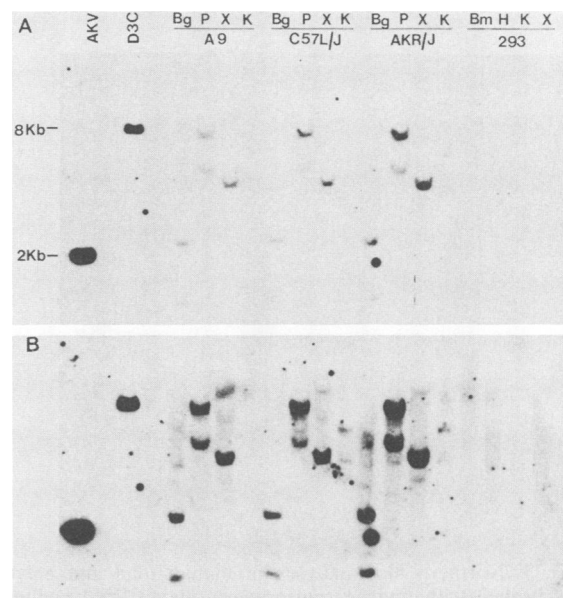


FIG. 2. Southern blot analysis of mouse and human DNAs probed with a radioactively labeled synthetic oligonucleotide mixture homologous to a conserved region of mammalian reverse transcriptase coding sequences. Restriction endonuclease-digested DNA samples were electrophoresed for 11 h in 1% agarose at 2 V/cm, transferred to nitrocellulose, and hybridized to the sense oligonucleotide mixture (200 ng, 10^9 cpm/ μ g) as described in Materials and Methods. Panels A and B represent, respectively, short (16-h) and long (60-h) exposures of the same probed filter. The lanes contain the following (from left to right). AKV, 1 μ g of an *Eco*RI-cleaved mouse cDNA clone of AKV viral RNA in phage lambda. D3C, 1 μ g of an *Eco*RI-cleaved human genomic clone (D3C) in phage lambda. (This clone was selected by using the 27 nucleotide mixed probe; although it contains a reasonably homologous target [see below], its reverse transcriptase context is not clear.) A-9, C57L/J and AKR/J, 10 μ g of high-molecular-weight DNA isolated from mouse cell line A-9 and mouse strains C57L/J and AKR/J digested with the following restriction endonucleases: Bg, *Bg*III; P, *Pst*I; X, *Xba*I; K, *Kpn*I. 293, 10 μ g of high-molecular-weight DNA isolated from the human cell line 293 and digested with the following endonucleases: Bm, *Bam*HI; H, *Hind*III. The approximate sizes of the AKV and D3C inserts homologous to the oligonucleotide probe are indicated on the left. The signal from the 2-kb AKV insert corresponds to about 10,000 copies of the oligonucleotide target sequence per cell. (The target in AKV is about 95% homologous to one member of the mixed 27 nucleotide probe, whereas that in D3C is about 85% homologous to another member.) A 2- μ g sample of *Hind*III-digested phage lambda DNA included for visible (ethidium bromide-stained) size markers was electrophoresed in a lane (not shown) adjacent to D3C.

shown), were as follows: (i) distinct, moderately repetitive bands were readily detected in mouse DNAs after a 16-h blot exposure, whereas lower-copy-number species became visible after 60 h; (ii) the patterns of moderately repetitive mouse DNAs were essentially identical for all three strains, whereas those of the lower-copy-number bands differed; and (iii) human DNA revealed a diffuse pattern of homologous material, with perhaps a few distinct visible bands in the 60-h exposure.

At present it appears that the *pol* gene represents the most highly conserved region among the endogenous proviral DNAs of mice, and the results shown in Fig. 2 strongly suggest that our probe detected many members of different proviral (retrotransposon) families. Unfortunately, it is dif-

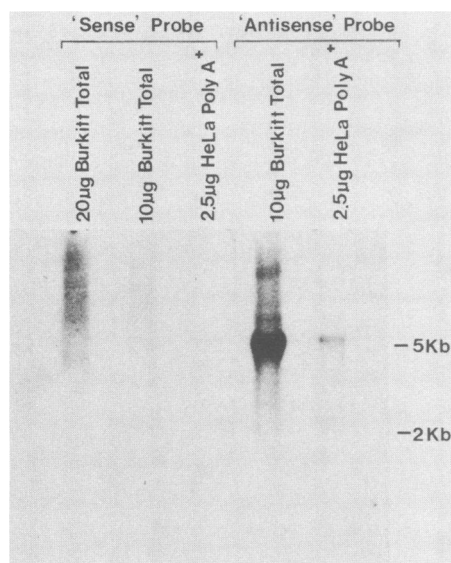


FIG. 3. Northern blot analyses of human total and poly(A)⁺ RNAs probed with reverse transcriptase-related DNA-coding sequences. The indicated amounts of total or poly(A)⁺ RNAs were electrophoresed for 13 h in 1% agarose containing 6.6% formaldehyde at 2.5 V/cm, transferred to nitrocellulose, and hybridized to either a sense or antisense synthetic oligonucleotide mixture as described in Materials and Methods. The positions of 5-kb (28S) and 2-kb (18S) rRNAs were determined from ethidium bromide-stained gels before transfer and are indicated on the right. The sense probe represents an oligonucleotide mixture composed of sequences expected in a reverse transcriptase coding mRNA, whereas the antisense probe represents a mixture of complementary sequences. Thus, it is the antisense probe that is expected to hybridize to mRNAs containing an homologous conserved reverse transcriptase coding sequence. Filters were autoradiographed for 16 h.

difficult to correlate the bands shown in this figure with known endogenous viral loci due to the fact that *gag*- and *env*-specific probes have been routinely used to characterize such loci and by the fact that mouse cells contain thousands of IAPs, hundreds of viruslike 30S, tens of C-type xenotropic and polytropic, a few C-type ecotropic, a few B-type, and various quantities of other endogenous proviral DNAs (60). Since the IAP family seems to be the most abundant and non-strain-specific mouse endogenous proviral DNA, it has been tentatively correlated with the moderately repetitive bands shown in Fig. 2 (29). The less intense bands (in some cases strain specific) have been tentatively correlated with the remaining classes of proviral DNAs.

The failure to detect discrete bands homologous to the 27 nucleotide mixed probe in Southern blots of human DNAs probably reflects the presence of thousands of divergent and dispersed reverse transcriptase sequences rather than their absence (see below).

The results of the Northern blot analysis shown in Fig. 3 and that of other Northern blot analyses were as follows: (i) HeLa poly(A)⁺ RNA and to a lesser extent, poly(A)⁺ RNA isolated from a Burkitt's lymphoma cell line (data not shown) contained an abundant 5- to 6-kb species homologous to the antisense (but not sense) oligonucleotide probe; (ii) Burkitt's lymphoma cell total RNA contained a variety of abundant species homologous to the antisense probe [including the 5- to 6-kb RNA detected in poly(A)⁺ samples] but contained much lower amounts of material (with an apparently different size distribution) homologous to the sense

probe; and (iii) 28S (5-kb) rRNA failed to hybridize to the antisense probe under the conditions used for our screenings (data not shown). Surprisingly, 28S rRNA contains a target sequence with about 70% homology to one member of the probe mixture (10).

The extensive hybridization of total RNA [in contrast to poly(A)⁺ RNA] with the oligonucleotide probe is typical of many highly conserved, repetitive dispersed sequences, whereas the detection of a discrete band of poly(A)⁺ RNA homologous to a highly conserved region of endogenous retroviruses is consistent with the recent identification of a variety of such RNAs in many other cell types (7, 9, 14, 36, 42).

Despite the positive results obtained in the blotting experiments, attempts to isolate reverse transcriptase homologs from human genomic and cDNA libraries by screening with the MOP downstream probe failed. We therefore investigated the use of PCR; the requirement that a template contain sequences homologous to two distinct probes was expected to increase the specificity of the screen, whereas the amplification inherent in the PCR was expected to allow detection of rare sequences.

Detection of conserved reverse transcriptase coding sequences by the PCR. PCR amplification has been shown to be an extremely powerful tool for the analysis of unique or extremely low-copy-number genes (19, 21, 48); it may also represent one of the most rapid and productive approaches for the characterization of multigene families. Specifically, the reverse transcriptase-specific primers presented in Fig. 1 were used to amplify a variety of human (and one mouse) DNAs at annealing temperatures ranging from 37 to 60°C, and the reaction products were analyzed by gel electrophoresis. In successful cases of amplification, the fragment mixtures were cloned by blunt-end ligation into M13mp19, sequenced, and then compared at the amino acid level both with each other and with known reverse transcriptase genes. The results of some gel electrophoresis analyses and of all of the amino acid sequence comparisons are shown in Fig. 4 and Table 1, respectively. The results of the electrophoresis analysis shown in Fig. 4 indicate that (i) the brightest band in any PCR generally corresponded to the size expected (~130 base pairs) for amplification of retroviruslike reverse transcriptases; (ii) the MOP functioned in a specific PCR amplification even though it was relatively ineffective as a probe for screening libraries and Southern genomic blots; (iii) higher annealing temperatures (higher PCR stringencies) generated fewer and more clearly defined amplification products; and (iv) HTLV-I- and HTLV-II-specific primers functioned at annealing temperatures as high as 60°C, even though it is generally assumed that amplification at annealing temperatures of 55°C requires nearly perfect primer-template homology (48). In addition the results of many other electrophoresis analyses (data not shown) indicate the following: (i) all of our primers, including those corresponding to human immunodeficiency virus (HIV), generate a product of the expected size when human genomic DNAs are used as templates at an annealing temperature of 60°C; (ii) only HTLV-I primers generate a product of the expected size when TC cDNA and T-cell (PTL) cDNA are used as templates at an annealing temperature above 55°C; and (iii) the upper limit for effective annealing of our primers is in the range of 65 to 70°C. On the whole, gel analysis of PCR-amplified fragments strongly suggested that conditions of both moderate and very high stringency allowed the synthesis of reverse transcriptase-related sequences. The results of the amino acid sequence comparisons (Table 1) confirmed

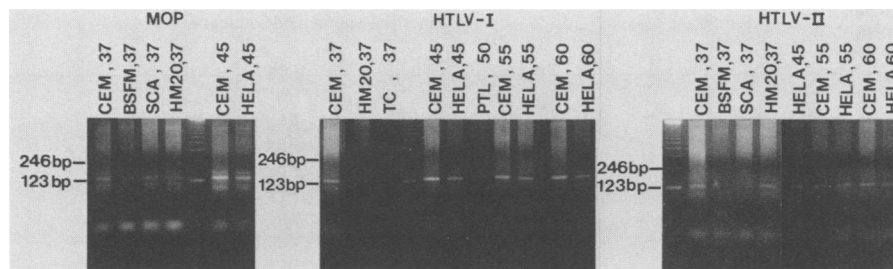


FIG. 4. Electrophoretic analysis of products from 25 polymerase chain reactions with seven human DNAs and three distinct reverse transcriptase-related primer mixtures. Gels were run for 2.5 h at 5 V/cm as described in Materials and Methods. The composite photographs shown here are grouped on the basis of primers (MOP, HTLV-I, and HTLV-II, see Fig. 1), and the template DNA and annealing temperature are indicated above each lane. A marker lane (unlabeled) containing 1.5 μ g of a 123-base-pair ladder (Bethesda Research Laboratories, Inc.) is present in each group, and the positions of the 123- and 246-base-pair fragments are indicated on the left. Template abbreviations: CEM, human T-cell lymphoblastoid cell line DNA; BSFM and HM20, human melanoma cell line DNAs; SCA, human scalp DNA; HELA, human HeLa cell line DNA; TC, human teratocarcinoma cell line cDNA (in phage lambda); PTL, human T-cell cDNA (in phage lambda). The lowest broad band in all of the lanes represents unincorporated primers.

this suggestion. Although some of the group designations and diagnostic amino acid assignments in Table 1 are admittedly a bit arbitrary, the reverse transcriptase contexts of all of the sequences outside of groups C and F are clear. Some of the more relevant technical observations related to Table 1 as well as to the nucleotide sequences from which it is derived are as follows. (i) Blunt-end ligation appeared to result in the cloning of full-length fragments. (ii) All primer sequences in the cloned fragments were exactly as expected, and in the case of MOP primer usage appeared to be random. (iii) Mixed HTLV-I-HTLV-II priming yielded two clones (A.9 and F.6) with only HTLV-I primer ends and two (B.4 and B.5) with a 5' HTLV-I and a 3' HTLV-II primer end. (iv) Fewer than 20% (groups C and F) of the total sequences were not in a reverse transcriptase context, and of these about half were derived from mixed oligonucleotide priming at 37°C (F.1 through F.4). (v) Of the 40 sequences in a reverse transcriptase context, only 2 (A.1 and A.11) contained termination codons. Only one amino acid sequence in the entire table, B.8, was not derived by direct translation of its cloned fragment. The nucleotide sequence of B.8 contained an A after the sixth codon which, when removed, allowed a perfect nucleotide sequence match with clone B.7. (vi) All examples of identical amino acid sequences (except for B.7 and B.8; see above) in cloned fragments, whether these fragments were derived from a single reaction mixture such as the three clones of group C or from different reaction mixtures such as clones A.3/A.4, A.5/A.16, A.9/A.12, and B.3/B6, were due to identical nucleotide sequences. (vii) The reported mutation frequency of the *T. aquaticus* polymerase (48) suggests that one out of three clones generated by amplification of the same template should contain a single nucleotide substitution. Thus, clones differing by a single nucleotide (data not shown) (such as A.2 and A.3, A.6 and A.8, A.12 and A.14, B.4 and B.5, and D.1 and D.2) and also differing by a single amino acid may have been generated from the same site.

In addition to these technical points and the overall efficacy of the PCR procedure, a few additional observations concerning the homologies of some sequences shown in Table 1 should be noted. (i) None of the human sequences was identical to that of any published retrovirus or human endogenous provirus, although many were similar to HuRRS-E and K. B.1 differed by only one amino acid from HuRRS-E. One mouse sequence (D.1) differed by one amino acid from a known IAP, and two other mouse amino acid sequences (D.2 and D.3) were identical to IAP and MMTV,

respectively. The nucleotide sequence comparisons between B.1 and HuRRS-E, D.2 and IAP, and D.3 and MMTV are shown in Fig. 5. Since clones B.1, D.2, and D.3 differ from their known chromosomal homologs at two or four internal (nonprimer) positions, it is likely that these particular homologs did not initiate amplification. (HuRRS-E is part of a rather large human endogenous provirus family, as is the IAP family of mice. Mice may also contain a variety of sequences that are highly homologous to MMTV.) The clustering of mismatches in primer regions suggests that mixed oligonucleotide priming at 37°C can amplify relevant regions even when primers and targets are relatively poorly matched. (ii) Some human sequences, especially those generated by high-stringency PCR with HTLV-I primers, bore limited but striking resemblances to primate-specific reverse transcriptases. Note, for example, group E and members of group A containing the HTLV-I diagnostic FPQC group four amino acids in from their carboxy termini. (iii) Some identical human sequences were present in both genomic and cDNAs (A.5/A.16, A.9/A.12), and many cDNA-related sequences also showed the resemblance to primate-specific reverse transcriptases described above. (iv) The final interesting homology relationship shown in Table 1 is that of group C, in which both the length and amino terminal residues of the amplified clones suggested a similarity to the L-1 family. (The hepatitis B virus homolog is included to show that these amino-terminal residues, or functionally conserved ones, are present in yet another reverse transcriptase gene.) Although the relationships in this group are unclear, the MOP was not expected to detect L-1 family members; it is therefore possible that PCR has amplified an L-1-like element in which the sequences of the MOP-binding sites resemble those of retroviruses.

Distribution of some reverse transcriptase-related, cloned DNA fragments in total PCR products. To investigate the distribution of specific amplified PCR fragments in various reaction mixtures, four reverse transcriptase-related M13 clones were labeled and used to probe blots of gels containing electrophoretically resolved PCR products. An experiment involving blots of the gels shown in Fig. 4 is presented in Fig. 6. The four reverse transcriptase-related cloned probes are as follows (see Table 1): (i) A.12, an HTLV-I high-stringency-primed fragment containing both monkey and HTLV-I-like characteristics; (ii) B.3, an HTLV-II high-stringency-primed fragment containing some monkey characteristics but quite different from any reported reverse transcriptase; (iii) C.1, an MOP moderate-stringency-primed

TABLE 1. Predicted amino acid sequences encoded by cloned PCR-amplified DNAs^a

Template DNA	Annealing temp (°C)	No. of independent clones	Primer	Internal amino acid sequence	
A.1 CEM	37	1	MOP	o ooooo oo ooo oo o MLNSPTICQTYVVGKVIKPVREQF*KCYSIH	A3 = A4
A.2 CEM	55	1	HTLV-I	MLNSPTLCQYFVGRVLQPVVDQFPRCYIVH	A9 = A12
A.3 CEM	55	1	HTLV-I	MLNSPTLCQYFVGRVLQPVVDQFPRCYIVY	A5 = A16
A.4 HeLa	55	3	HTLV-I	MLNSPTLCQYFVGRVLQPVVDQFPRCYIVY	
A.5 HeLa	55	1	HTLV-I	MLNSTICQYYVGTILKPVVDQFPQCYVIH	
A.6 CEM	55	2	HTLV-I	MLNSIISACCRIGIKG--ASEYVSTAYIRH	
A.7 CEM	60	1	HTLV-I	MLNSPTIYQYFVGHGLQPVVDQFPRCYIVH	
A.8 CEM	60	1	HTLV-I	MLNSIISACCRIGIKG--ASEYLSAYIRH	
A.9 HeLa	60	1	HTLV-I-HTLV-II	MLNSPTICQYFVGRVLQPVVDQFPRCYIVH	
A.10 HeLa	60	1	HTLV-I	MLNSPTICQYFVGRVLQPVRIISFPRCYIVH	
A.11 HeLa	60	1	HTLV-I	MLNSLTICQTYVRKAIK*VRE*FKKCYIIH	
A.12 TC	60	1	HTLV-I	MLNSPTICQYFVGRVLQPVVDQFPRCYIVH	
A.13 PTL	60	1	HTLV-I	MLNSPTICQYFVGRVLQPVVDQFPRCYIVY	
A.14 PTL	60	1	HTLV-I	MLNSPTICQYFVGHVLQPVVDQFPRCYIVH	
A.15 PTL	60	1	HTLV-I	RLNSIISACCRIGIKG--ASDMFPTAYIRH	
A.16 PTL	60	1	HTLV-I	MLNSTICQYYVGTILKPVVDQFPQCYVIH	
A.17 PTL	60	1	HTLV-I	MLNSPTICQYFVGRVLQPCQGHVFRCYILH	
				FKNSPTLFEMQLAHILQPIRQAFPOCTILQ FKNSPTLFEQQLAAVLNPMRKMFPSTTIYQ MANSPTLCQKYVAAAIEPVRKSWAQMYIIH MANSPTLCQKYVATAIHKVRHAWKQMYIIH MKNSPTLCQKFVDKAILTVRDKYQDSYIVH MLNSPTICQTFVGRALQPVREKFSDCYIIH	HTLV-I Homolog HTLV-II Homolog SRV-II Homolog MPMV Homolog MMTV Homolog HuRRS-K Homolog
B.1 CEM	37	1	MOP	o oo oo oo oooo o oo ooo FKNSPTIFGEALARDLQKFPTRDLGCVLLK ←	B3 = B6
B.2 CEM	37	1	MOP	FRDSSHLPFGEALTRALSQFSYLD--TLVLW	
B.3 CEM	55	2	HTLV-II	FRDSPHYFGQALQLDLSQLHLQ-PS-ILLQ	
B.4 CEM	60	1	HTLV-I-HTLV-II	FRDSPHYFGQALQLDLSQLHLQ-PS-IIILQ	
B.5 HeLa	60	1	HTLV-I-HTLV-II	FRDSPHYFGQALQLDLSQIHIQ-PS-ILLQ	
B.6 HeLa	60	1	HTLV-II	FRDSPHYFGQALQLDLSQLHLQ-PS-ILLQ	
B.7 TC	60	1	HTLV-I	FKDSPLYLASISPR-LEPILI--PDTFVLQ	
B.8 PTL	60	1	HTLV-I	FKDSPLYLASISPR-LEPILI--PDTFVLQ	
				WKGSPAIFQSSMTKILEPFKKQNPDIYIYQ FKNSPTLFDALHRDLTDFRTQHPVETLLQ FKNSPTLFDALHRDLADFRIQHDPDLILLQ FKNSPTIFGEALARDLQKFPTRDLGCVLLQ	HIV Homolog BaEV Homolog MoMLV Homolog HuRRS-E Homolog
C.1 CEM	37	3	MOP	ooo o o RGLSCL-FII*LCCGKTCFFHAQAICYADTLGNWSFGSGK VGLSP--FLLAQFTSAICSVVRRAPPHCLAFS CPLSPLLFNIVLEVLARAIRHELEIKKIQLGKEEVKLSL	HBV Homolog L1 Homolog
D.1 MU-S	37	3	MOP	o ooooo oo oo o ooo o MSNSPTMCQLYMQEALLPVREQFPSLILL	
D.2 MU-S	37	1	MOP	MSNSPTMCQLYVQEALLPVREQFPSLILL ←	
D.3 MU-S	37	1	MOP	MKNSPTLCQKFVDKAILTVRDKYQDSYIVH ←	
D.4 MU-S	37	1	MOP	MANSPTMCQLYVGKAVEPIRKEYPKLRCVH	
				MKNSPTLCQKFVDKAILTVRDKYQDSYIVH MSNSPTMCQLYVQEALLPVREQFPSLILL	MMTV Homolog MU-IAP Homolog
E.1 CEM	37	3	MOP	o ooooo oo oo oo ooo oooo MSNSPTICQTYVVGQAIIEPTRKKFSQCY-IH	
E.2 TC	60	2	HTLV-I	MLNSPTVCQIYVRKAILPVREQFKKCYIIH	
				MANSPTLCQKYVAAAIEPVRKSWAQMYIIH	SRV-II Homolog
F.1 CEM	37	1	MOP	RSRG*RTVNHAFLYLCPLAAGGLWCSLEVDF*WGS	
F.2 CEM	37	1	MOP	KKDAPSPSAAIPLYCKQHDSANQPQIV	
F.3 CEM	37	1	MOP	FSSPR*VYR*VLALR*	
F.4 CEM	37	1	MOP	SQGEAGEVGRSHSY	
F.5 CEM	60	1	HTLV-I	RLNSLFLVKLMGKLLSQLENSKNGILS	
F.6 HeLa	60	1	HTLV-I-HTLV-II	FINSLLCVIILFGENLIAPHIRQIHWS	
F.7 HeLa	60	1	HTLV-II	YQCSSLHFHIVQN*RILIAFPQKMPHWS	

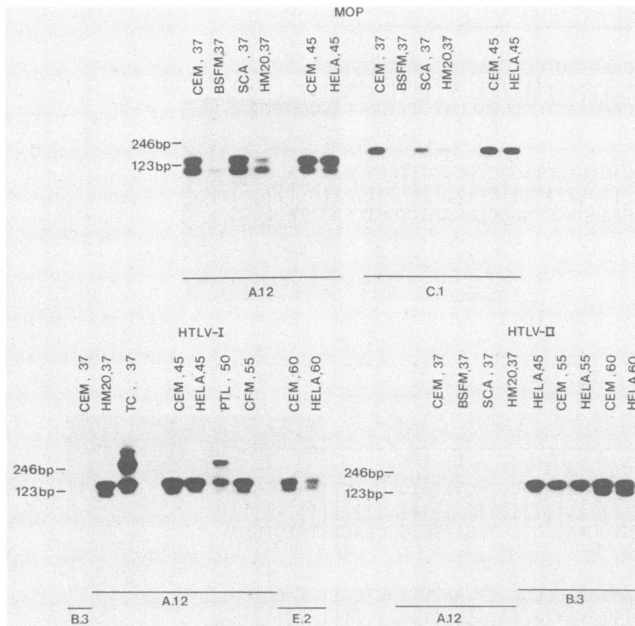


FIG. 6. Southern blot analysis of PCR-amplified DNAs. The gels photographed for Fig. 4 were blot transferred to nylon filters (Biotrans; ICN Pharmaceuticals Inc.), and cut sections of these filters were hybridized separately to one of four cloned PCR-amplified, reverse transcriptase-related DNA fragments. Lane designations and groupings shown on the top of each autoradiograph are as in Fig. 4; labeled lines within each autoradiograph span the regions hybridized to probe A.12, B.3, C.1, or E.2 (Table 1, see the text). The autoradiographs correspond to the following hybridizations. MOP was probed with A.12; MOP (after stripping of the filter probed with A.12) probed with C.1 (note that the material hybridizing to C.1 migrated more slowly than did that hybridizing to A.12, as expected for the larger size of group C clones). HTLV-I lane CEM,37 was probed with B.3, lanes HM20,37 to CEM,55 was probed with A.12, and lanes CEM,60 and HELA,60 were probed with E.2. HTLV-II lanes CEM,37 to HM20,37 were probed with A.12, and lanes HELA,45 to HELA,60 (right) were probed with B.3. Autoradiographs were exposed for 1 h.

MOP-primed reaction mixtures but not in HTLV-I-primed mixtures (data not shown), whereas the HTLV-I-like sequence (A.12) was present in both. Also, the monkeylike sequence (E.2) was present in HTLV-I- but not HTLV-II-primed reaction mixtures (data not shown), whereas the rather novel sequence (B.3) was present in HTLV-II- but not HTLV-I-primed mixtures. (The nature of the lower band of the doublet in the 123-base-pair range, observed by both blotting and UV absorption in many PCR products, is probably due to the amplification of slightly shorter templates as reflected by products such as A.6 or B.7, shown in Table 1.) Due to the efficiency of amplification after the first successful priming reaction and the reasonable selectivity of mixed oligonucleotide priming at 37°C (MOP clones), it is probable that additional blotting studies will reveal isolated cases of unexpected priming. In any event, the data shown in Table 1 and Fig. 6 indicate that high-stringency-primed amplifications generally yield specific and distinct families of related sequences.

Distribution of some reverse transcriptase-related cloned DNA fragments in genomic DNA. To investigate the distribution of some specific amplified PCR fragments in human DNA, the four reverse transcriptase-related clones described above were labeled and used to probe blots of gels

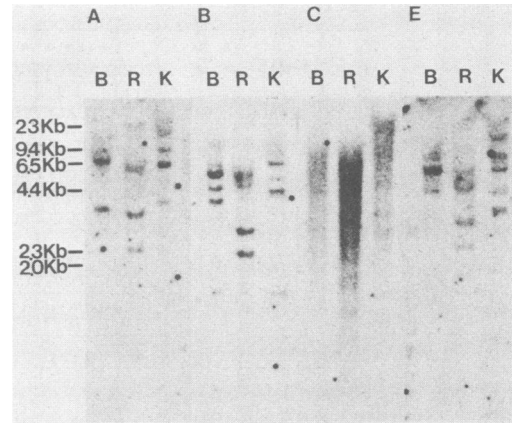


FIG. 7. Southern blot analysis of human CEM DNA probed with four cloned PCR amplified, reverse transcriptase-related, DNA fragments (see legend to Fig. 6 and the text). Restriction endonuclease-digested DNA samples (10 μ g per lane) were electrophoresed for 11 h in 1% agarose at 2 V/cm, transferred to nitrocellulose, and hybridized as previously described (see Materials and Methods). Each panel shows the autoradiograph of a filter containing the same DNA (CEM), digested with the same enzymes (B, *Bam*HI; R, *Eco*RI; K, *Kpn*I) but hybridized to different probes as follows: A, A.12; B, B.3; C, C.1; E, E.2. Autoradiographs were exposed for 15 h.

containing electrophoretically resolved restriction endonuclease-digested DNAs. All of the reverse transcriptase-related cloned probes detected repetitive sequences in chromosomal DNA (Fig. 7). In the case of the monkey retroviruslike probes (A.12, B.3, and E.2), bands with signal intensities corresponding to 10 to 100 copies per cell were evident, and there were definite differences in the patterns obtained with different probes. In the case of the L-1-like probe (C.1), bands and smears with combined signal intensities of more than 100 copies per cell were evident, and the pattern is suggestive of a long interspersed nucleotide sequence family. Blotted human scalp or HeLa DNAs yielded similar if not identical patterns when hybridized to each of these probes (data not shown). Interestingly, C.1 exhibited no obvious homology to any region of an L-1 compilation sequence (53) or to any published primate L-1 sequences. The results of the genomic Southern blotting experiments presented here suggest that many of the reverse transcriptase-related sequences amplified by PCR are members of multigene subfamilies. In addition, since most of the bands in Fig. 7 are significantly larger than the reverse transcriptase coding region, it is likely that each reverse transcriptase is part of a larger repetitive element.

DISCUSSION

The human genome apparently contains a wide spectrum of sequences that are closely related to the reverse transcriptase genes of retroviruses, and it is assumed that most of these sequences are part of defective proviral DNAs. In this communication we show that synthetic oligonucleotides homologous to highly conserved regions of reverse transcriptase genes are particularly effective in detecting such sequences when used as primers in PCR amplification. These primers may be either unique (HTLV-I and HTLV-II; Fig. 1), or mixed (MOP; Fig. 1), and the nature of the amplified sequences appears to depend on both the primer sequence and annealing temperature (although the relative

importance of each has not been investigated systematically).

Surprisingly, MOP amplification of genomic DNA or cDNA (21) is certainly a more productive approach to identifying reverse transcriptase sequences than is conventional mixed-oligonucleotide screening of genomic or cDNA libraries. Amplification of a true sequence is less dependent on the extent of primer-target homology than is the detection of a true positive signal in oligonucleotide probing. Published data on MOP-amplified cDNA (21) and the results shown here (Fig. 5) on MOP-amplified genomic DNA suggest that primer-target homologies of less than 70% can amplify expected sequences, at least with 37°C annealing and either 37 or 72°C extension. It is assumed that this fidelity of PCR is due primarily to the requirement for two independent primer-target interactions in specific orientation within a defined distance, and that the use of increased annealing temperatures and/or unique primers would limit the composition of the amplified products. Indeed, unique oligonucleotide primers of 20 residues and annealing temperatures of 55°C supposedly require perfect or nearly perfect primer-target matching for successful amplification (48). (The annealing limit for our HTLV-I and HTLV-II primers was between 65 and 70°C; the limit of our mixed primers was not tested.)

The nature of the human reverse transcriptase-related amino acid sequences detected by PCR in both genomic and cDNAs was quite varied (Table 1). Neglecting putative L-1-like sequences (group C) and those sequences in no clear context (group F), one might distribute the remaining clones between two major mouse-related classes, one broadly homologous to MMTV (including groups A and E) and the other broadly homologous to Moloney MLV (group B). This classification is reasonably consistent with previous studies of human endogenous retroviruses, which suggested that MMTV- and Moloney MLV-related species account for thousands of human proviral DNAs. However, it does not take into account other striking amino acid sequence relationships, nor does it consider the more likely origin of many or most human endogenous retroviruses or retrovirus-related genes, which would be primate rather than murine.

Many of the amino acid sequences shown in Table 1 corresponding to HTLV-I-primed, high-stringency (55 and 60°C) PCR amplifications revealed striking similarities to reverse transcriptases of primate retroviruses. Specifically, these are the FPQC-like or identical blocks noted previously for most genomic and cDNA members of group A (related to the HTLV-I homolog), 18 out of 30 amino acid identities noted for group E (related to the simian retrovirus type II [SRV-II] homolog), and possibly the LEP block noted for the two cDNA-derived (TC, PTL) members of group B (related to the HIV homolog). These homologies are admittedly limited, as is the entire region here under study, but when considered in conjunction with their amplification by very high-stringency priming with primate-specific oligonucleotides and present views regarding the evolution of retroviruses (28), many or all of these reverse transcriptase-like regions may be primate related. Evolutionarily, MMTV appears to be closely related to SRV-I, SRV-II, and Mason-Pfizer monkey virus, whereas Moloney MLV, itself quite distantly related to MMTV, is very closely related to baboon endogenous virus. At least in terms of the amino acid sequences presented in Table 1, groups A and B can just as easily be related to their monkey homologs (SRV-II/Mason-Pfizer monkey virus and baboon endogenous virus, respectively) as to their murine homologs; as noted above, the

inclusion of HTLV-I and HIV homologs often strengthens the primate relationship. As might be expected from genetic drift and recombination (including gene conversion) among presumed predominantly defective human proviruses, this relationship is not straightforward and may be quite complex. The peculiar Mason-Pfizer monkey virus/HTLV-I hybridlike character of many group A sequences (such as A.16) could reflect this complexity, as could the MMTV/SRV hybridlike character of HuRRS-K noted by Ono et al. (37). Eventually, additional PCR amplifications with different primers and stringencies, characterizations of genomic library clones, and cross-probed Southern genomic blots of many primate DNAs should reveal the evolutionary history of human endogenous retroviruses and/or related sequences. If human and monkey retroviruses diverged and evolved relatively independently, it is possible that many human endogenous retrovirus-related sequences are derived from human viruses.

Irrespective of the origin of human retroviruses, their presence leads to both practical and theoretical concerns. Presently, the major practical concern is that the effective use of PCR as a screening procedure for HTLV-I, HTLV-II, and HIV infections must always include appropriate controls to ensure that no endogenous sequences contribute to positive signals. As previously noted, HIV unique primers corresponding to the highly conserved reverse transcriptase region shown in Fig. 1 function well in the PCR amplification of HeLa DNA even at annealing temperatures of 60°C. Although it is unlikely that the internal nucleotide sequence(s) would be detected after stringent hybridization to an HIV-specific probe, other sites in HIV are probably more appropriate for its unambiguous detection by PCR.

Another practical concern is that the use of PCR for determining the possible retroviral etiology of a variety of human diseases may be complicated by endogenous retroviruses. Even if cDNAs are used for PCR templates, the transcriptional activities of endogenous sequences must be considered.

The major theoretical concern related to the presence of endogenous human retroviruses is their potential expression. In analogy with murine systems, the following might be expected: (i) that such sequences in humans will be classed as low-, high-, or, most often, non-virus-producing and that this classification will depend on the nature of both the provirus and the host; (ii) that some human endogenous proviral DNAs that are defective nonvirus producers might yield infectious and/or oncogenic recombinants with exogenous viruses or with other defective or nondefective proviruses; (iii) that some human sequences might be partially transcribed and translated into functional gene products; and (iv) that some endogenous elements might be responsible for known or new mutations as well as for the insertional activation of oncogenes, as are some human L-1 elements (16, 32).

In addition, it is also possible that the human genome contains reverse transcriptases that are not associated with endogenous viruses. The use of MOP amplification of cDNA [either from cloned cDNAs or from reverse transcriptase-copied poly(A)⁺ RNAs] coupled with cDNA library characterizations, should provide a reasonably rapid approach to investigating reverse transcriptase gene expression. Investigations with our human teratocarcinoma and T-cell cDNA PCR amplification products are in progress.

In conclusion, the importance of reverse transcription in regard to the development and evolution of mammalian genes and its present importance as a mechanism of DNA

transposition suggest that a variety of reverse transcriptase genes, including those in unexpected contexts, are present in humans and mice. Moreover, the homology relationships of human putative reverse transcriptases identified by PCR amplification suggest that many or most of them are related to primate retroviruses. Such a relationship, if meaningful, implies an ancient ($>10^5$ years) association between retroviruses and humans. It is possible that the proposed transfer of an HTLV-I-like monkey virus to humans in the 16th century and of an HIV-like monkey virus to humans in the 20th century merely represent the most recent detectable examples of monkey-human cross-infections. It is also possible that both human and monkey exogenous retroviruses have been evolving relatively independently (8), and that many of the human endogenous sequences were derived from human retroviruses.

ACKNOWLEDGMENTS

We thank Horace Lozina and Heinz Annus for excellent technical help, Roger Karess, Peter D'Eustachio, Hannah Klein, and Robert Schneider for helpful discussions and support, Steve Sloane for assistance with the polymerase chain reaction, Lara Schulman and Ross Smith for help with computer analysis, and Aileen Siegel for typing the manuscript.

This work was supported by National Science Foundation grant PCM 8203006 and by grant BRSG S07RR5399-25 of the Division of Research Resources, National Institutes of Health. A.S. and R.M. are predoctoral trainees supported by Public Health Service training grants 1GM07827 and 5T32GM07238208, respectively, from the National Institutes of Health.

LITERATURE CITED

- Callahan, R., I. M. Chiu, J. H. F. Wong, S. Tronick, B. A. Roe, S. A. Aaronson, and J. Schlom. 1985. A new class of endogenous human retroviral genomes. *Science* **228**:1208-1211.
- 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.
- Chiu, I.-M., A. Yaniv, J. E. Dahlberg, A. Gazit, S. F. Skuntz, S. R. Tronick, and S. A. Aaronson. 1985. Nucleotide sequence evidence for relationship of AIDS retrovirus to lentivirus. *Nature (London)* **317**:366-368.
- Finnegan, D. J. 1985. Transposable elements in eukaryotes. *Int. Rev. Cytol.* **93**:281-326.
- Foley, G. E., H. Lazarus, S. Farber, B. G. Uzman, B. A. Boone, and R. E. McCarthy. 1965. Continuous culture of human lymphoblasts from peripheral blood of a child with acute leukemia. *Cancer* **18**:522-529.
- Franck, A., H. Guillely, G. Jonaro, K. Richards, and L. Hirth. 1980. Nucleotide sequence of cauliflower mosaic virus DNA. *Cell* **21**:285-294.
- Franklin, G. C., S. Chretien, I. M. Hanson, H. Rochefort, F. E. B. May, and B. R. Westley. 1988. Expression of human sequences related to those of mouse mammary tumor virus. *J. Virol.* **62**:1203-1210.
- Fukasawa, M., T. Miura, A. Hasegawa, S. Morikawa, H. Tsujimoto, K. Miki, T. Kitamura, and M. Hayami. 1988. Sequence of simian immunodeficiency virus from African green monkey, a new member of the HIV/SIV group. *Nature (London)* **333**:457-461.
- Gattoni-Celli, S., K. Kirsch, S. Kalled, and K. J. Isselbacher. 1986. Expression of type C-related endogenous retroviral sequences in human colon tumors and colon cancer cell lines. *Proc. Natl. Acad. Sci. USA* **83**:6127-6131.
- Gonzalez, I. L., J. L. Gurski, T. J. Campen, D. J. Dorney, J. M. Erickson, J. E. Sylvester, and R. D. Schmickel. 1985. Variation among human 28S ribosomal RNA genes. *Proc. Natl. Acad. Sci. USA* **82**:7666-7670.
- Graham, F. L., J. Smiley, W. C. Russell, and R. Nairn. 1977. Characteristics of a human cell line transformed by DNA from human adenovirus types. *J. Gen. Virol.* **36**:59-72.
- 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.
- Hattori, M., S. Kuhara, O. Takenaka, and Y. Sakaki. 1986. L1 family of repetitive sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. *Nature (London)* **321**:625-628.
- Kato, N., S. Pfeifer-Ohlsson, M. Kato, E. Larsson, J. Rydnert, R. Ohlsson, and M. Cohen. 1987. Tissue-specific expression of human provirus ERV3 mRNA in human placenta: two of the three ERV3 mRNA contain human cellular sequences. *J. Virol.* **61**:2182-2191.
- 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.
- Kazanian, H. H., Jr., C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips, and S. E. Antonarakis. 1988. Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature (London)* **332**:164-166.
- Kobayashi, M., and K. Koike. 1984. Complete nucleotide sequence of HBV DNA of subtype ADR and its conserved gene organization. *Gene* **30**:227-234.
- Kroger, B., and I. Horak. 1987. Isolation of novel human retrovirus-related sequences by hybridization to synthetic oligonucleotides complementary to the tRNA^{PRO} primer-binding site. *J. Virol.* **61**:2071-2075.
- Kwok, S., D. H. Mack, K. B. Mullis, B. Poiesz, G. Ehrlich, D. Blair, A. Friedman-Kien, and J. J. Sninsky. 1987. Identification of human immunodeficiency virus sequences by using in vitro enzymatic amplification and oligomer cleavage detection. *J. Virol.* **61**:1690-1694.
- Lathe, R. 1985. Synthetic oligonucleotide probes deduced from amino acid sequence data: theoretical and practical considerations. *J. Mol. Biol.* **183**:1-12.
- Lee, C. C., X. Wu, R. A. Gibbs, R. G. Cook, D. M. Muzny, and C. T. Caskey. 1988. Generation of cDNA probes directed by amino acid sequence: cloning of urate oxidase. *Science* **239**:1288-1291.
- Littman, D. R., Y. Thomas, P. J. Maddon, L. Chess, and R. Axel. 1985. The isolation and sequence of gene encoding T8: a molecule defining functional classes of T lymphocytes. *Cell* **40**:237-246.
- Loeb, D. D., R. W. Padgett, S. C. Hardies, W. R. Shehee, M. B. Comer, M. H. Edgell, and C. A. Hutchison. 1986. The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. *Mol. Cell. Biol.* **6**:168-182.
- Mager, D. L., and J. D. Freeman. 1987. Human endogenous retrovirus-like genome with type C *pol* sequences and *gag* sequences related to human T-cell lymphotropic viruses. *J. Virol.* **61**:4060-4066.
- 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.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 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.
- 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.
- Mietz, J., 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.

30. Misra, R., A. Shih, M. Rush, E. Wong, and C. W. Schmid. 1987. Cloned extrachromosomal circular DNA copies of the human transposable element THE-1 are related predominantly to a single type of family member. *J. Mol. Biol.* **196**:233-243.
31. 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.
32. Morse, B., P. G. Rotherg, V. J. South, J. M. Spandorfer, and S. M. Astrin. 1988. Insertional mutagenesis of the *myc* locus by a LINE-1 sequence in a human breast carcinoma. *Nature (London)* **333**:87-90.
33. O'Connell, C. D., and M. Cohen. 1984. The long terminal repeat sequences of a novel human endogenous retrovirus. *Science* **226**:1204-1206.
34. O'Connell, C. D., S. O'Brien, W. G. Nash, and M. Cohen. 1984. ERV3, a full-length human endogenous provirus: chromosomal localization and evolutionary relationships. *Virology* **138**:225-235.
35. Ono, M. 1986. Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to type A and B retrovirus genes. *J. Virol.* **58**:937-944.
36. Ono, M., M. Kawakami, and H. Ushikubo. 1987. Stimulation of expression of the human endogenous retrovirus genome by female steroid hormones in human breast cancer cell line T47D. *J. Virol.* **61**:2059-2062.
37. Ono, M., T. Yasunaga, T. Miyata, and H. Ushikubo. 1986. Nucleotide sequence of human endogenous retrovirus genome related to mouse mammary tumor virus genome. *J. Virol.* **60**:589-598.
38. Patarca, R., and W. A. Haseltine. 1984. Matters arising: alignment of the similar regions in polymerase gene products among different viruses. *Nature (London)* **309**:359-361.
39. Paulson, K. E., N. Deka, C. W. Schmid, R. Misra, C. W. Schindler, M. G. Rush, L. Kadyk, and L. Leinwand. 1985. A transposon-like element in human DNA. *Nature (London)* **316**:359-361.
40. Pustell, J., and F. C. Kafatos. 1982. A convenient and adaptable package of DNA sequence analysis programs for microcomputers. *Nucleic Acids Res.* **10**:51-59.
41. Pustell, J., and F. C. Kafatos. 1982. A high speed, high capacity homology matrix: zooming through SV40 and polyoma. *Nucleic Acids Res.* **10**:4765-4782.
42. Rabson, A. B., P. E. Steele, C. F. Garon, and M. A. Martin. 1983. mRNA transcripts related to full-length endogenous retroviral DNA in human cells. *Nature (London)* **306**:604-607.
43. Ratner, L., W. Haseltine, R. Patarca, K. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalsk, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway Jr., M. L. Pearson, J. A. Lautenberger, T. S. Papas, J. Ghayeb, N. T. Chang, R. C. Gallo, and F. Wong-Staal. 1985. Complete nucleotide sequence of the AIDS virus HTLV-III. *Nature (London)* **313**:277-284.
44. Repaske, R., R. R. O'Neill, P. E. Steele, and M. A. Martin. 1983. Characterization and partial nucleotide sequence of endogenous type C retrovirus segments in human chromosomal DNA. *Proc. Natl. Acad. Sci. USA* **80**:678-682.
45. Repaske, R., P. E. Steele, R. R. O'Neill, A. B. Rabson, and M. A. Martin. 1985. Nucleotide sequence of a full-length human endogenous retroviral segment. *J. Virol.* **54**:764-772.
46. Rice, N. R., R. M. Stephens, A. Burny, and R. V. Gilden. 1985. The *gag* and *pol* gene of bovine leukemia virus: nucleotide sequence and analysis. *Virology* **142**:357-377.
47. Rogers, J. H. 1985. The structure and evolution of retroposons. *Int. Rev. Cytol.* **93**:231-279.
48. Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**:487-491.
49. Schwartz, D. E., R. Tizard, and W. Gilbert. 1983. Nucleotide sequence of Rous sarcoma virus. *Cell* **32**:853-869.
50. 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.
51. Shimotohno, K., Y. Takahashi, N. Shimizu, T. Gojobori, D. W. Golde, I. S. Y. Chen, M. Miwa, and T. Sugimura. 1985. Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus Type II: an open reading frame for the protease gene. *Proc. Natl. Acad. Sci. USA* **82**:3101-3105.
52. Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukemia virus. *Nature (London)* **293**:543-548.
53. 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.
54. Skowronski, J., T. G. Fanning, and M. F. Singer. 1988. Unit-length LINE-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* **8**:1385-1397.
55. Skowronski, J., and M. F. Singer. 1985. Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. *Proc. Natl. Acad. Sci. USA* **82**:6050-6054.
56. Sonigo, P., M. Alizon, K. Staskus, D. Klatzmann, S. Cole, O. Danos, E. Retzel, P. Tiollais, A. Haase, and S. Wain-Hobson. 1985. Nucleotide sequence of visna lentivirus: relationship to the AIDS virus. *Cell* **42**:369-382.
57. Sonigo, P., C. Baker, E. Hunter, and S. Wain-Hobson. 1986. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. *Cell* **45**:375-385.
58. Staden, R. 1984. Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Res.* **12**:505-519.
59. Staden, R. 1984. Graphic methods to determine the function of nucleic acid sequences. *Nucleic Acids Res.* **12**:521-538.
60. Stoye, J., and J. Coffin. 1985. Endogenous viruses, p. 357-404. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), *RNA tumor viruses*, vol. 2, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
61. Thayer, R. M., M. D. Power, M. L. Bryant, M. B. Gardner, P. J. Barr, and P. A. Luciw. 1987. Sequence relationships of type D retrovirus which cause simian acquired immunodeficiency syndrome. *Virology* **157**:317-329.
62. Toh, H., H. Hayashida, and T. Miyata. 1983. Sequence homology between retroviral reverse transcriptase and putative polymerase of hepatitis B virus and cauliflower mosaic virus. *Nature (London)* **305**:827-829.
63. Toh, H., R. Kikuno, H. Hayashida, T. Miyata, W. Kugimiya, S. Inouye, S. Yuki, and K. Saigo. 1985. Close structural resemblance between putative polymerase of a *Drosophila* transposable genetic element 17.6 and *pol* gene product of Moloney murine leukemia virus. *EMBO. J.* **4**:1267-1272.
64. Varmus, H. 1988. Retroviruses. *Science* **240**:1427-1435.
65. Weiner, A. M., P. L. Deininger, and A. Efstratiadis. 1986. Nonviral retroposons. *Annu. Rev. Biochem.* **55**:631-661.
66. Wood, W. J., J. Gitschian, L. A. Lasky, and R. M. Lawn. 1985. Base composition-independent hybridization in tetramethyl ammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. *Proc. Natl. Acad. Sci. USA* **82**:1585-1588.