JOURNAL OF VIROLOGY, Sept. 1982, p. 914–924
0022-538X/82/090914-11$02.00/0

Vol. 43, No. 3

Molecular Cloning of a Family of Retroviral Sequences Found in Chimpanzee but Not Human DNA

TOM I. BONNER,[1]* EDWARD H. BIRKENMEIER,[1,2] MATTHEW A. GONDA,[3] GEORGE E. MARK,[1] GEORGE H. SEARFOSS,[3] AND GEORGE J. TODARO[1]

Laboratory of Viral Carcinogenesis, National Cancer Institute,[1] and National Cancer Institute-Frederick Cancer Research Facility,[3] Frederick, Maryland 21701, and The Jackson Laboratory, Bar Harbor, Maine[2]

Received 1 March 1982/Accepted 27 May 1982

A number of retrovirus-like sequences have been cloned from chimpanzee DNA which constitute the chimpanzee homologs of the endogenous colobus type C virus CPC-1. One of the clones contains a nearly complete viral genome, but others have sustained deletions of 1 to 2 kilobases in the polymerase gene. The pattern of related sequences detected in other primate species is consistent with the genetic transmission of these sequences for millions of years. However, the appropriately related sequences have not been detected in human, gibbon, or orangutan DNAs. These results suggest either that this family of sequences has been deleted from humans, gibbons, and orangutans, or that the genes were recently acquired in the chimpanzee and gorilla lineages.

Many retroviruses are considered to be endogenous because closely related sequences are found in the chromosomal DNA of all members of the species from which they were isolated (11). In addition, related sequences are found in the DNA of closely related species, suggesting that they have been genetically transmitted for millions of years. The endogenous retroviruses are thus distinguished from the exogenous viruses such as gibbon ape leukemia virus or Mason-Pfizer monkey virus, which have integrated DNA copies in only some members of a species and not in closely related species. Virus-related sequences occur in so many vertebrate genomes that they almost seem to be universal. However, until recently there has been no clear evidence for the presence of such sequences in humans (14; T. I. Bonner, C. O'Connell, and M. Cohen, Proc. Natl. Acad. Sci. U.S.A., in press). As a result it is difficult to assess the role of endogenous retroviral genomes in human carcinogenesis. Even though there is no direct evidence relating to humans, the high incidence of spontaneous leukemia in the AKR mouse is associated with the expression of integrated endogenous retroviruses (7). Furthermore, an avian leukosis virus, which is closely related to the endogenous retroviruses of chicken, has been shown to induce tumors at low frequency by activating certain cellular genes (15). Although it appears that similar mechanisms could be involved in human carcinogenesis, the testing of this hypothesis requires a means of detecting endogenous human retroviral sequences. For a number of years attention has therefore focused on viruses isolated from close relatives of humans.

The closest relatives of humans from which endogenous retroviruses have been isolated are the Old World monkeys. There are three distinct families of endogenous retroviruses of Old World monkeys: (i) the type D virus, LAD-1 isolated from langur (4); (ii) the numerous type C virus isolates from baboon (22); and (iii) the type C viruses MAC-1, MMC-1, and CPC-1, isolated from two species of macaque (stumptailed and rhesus) and from colobus, respectively (17, 19, 21). By using cDNA probes from any of these families, related sequences can be detected in most Old World primates, with sequence divergence which is proportional to the divergence of these primates from the species from which the virus was isolated (3, 4, 6, 19, 21). This result suggests that the sequences have been genetically transmitted for tens of millions of years. However, for the two type C virus families there are exceptions where the divergence is greater than expected. For example, there is easily detected homology in chimpanzee and gorilla DNA but not in human, gibbon, or orangutan DNAs, even though all of the apes are equidistant from the species from which these viruses have been isolated. It has been suggested that the rate of divergence is greater in some species and that the failure to detect related human sequences may be because the divergence is too great (3, 6). However, similar results in avian species have been interpreted as evidence of recent infections of some species and not of others (10).

If viral sequences have been in the primate genome for tens of millions of years, then the viral sequences from chimpanzees or gorillas,

the closest living relatives to humans, should provide the optimal probe for detection of human endogenous retroviral sequences. We have tested this hypothesis by cloning from chimpanzee cellular DNA sequences which are related to the colobus type C virus CPC-1. We find that essentially complete retroviral genomes are present in chimpanzee DNA even though no retrovirus has been isolated from chimpanzee. When these viral sequences were used as a probe, related sequences were found in several primate species consistent with the long-term genetic transmission of these sequences, and yet no closely related sequences were detected in human, orangutan, or gibbon DNAs.

MATERIALS AND METHODS

Preparation of DNA. High-molecular-weight cellular DNAs were extracted from primate tissues by common procedures. Two human DNAs were a gift from Maurice Cohen (Frederick Cancer Research Facility), the gorilla DNA was obtained from Mike Tainsky (Frederick Cancer Research Facility), and the orangutan and gibbon DNAs came from Raoul Benveniste (National Cancer Institute). Molecularly cloned CPC-1 viral DNA has been described (5) and was subcloned in the plasmid pBR322. Plasmid DNA was purified with a cleared lysate method followed by two steps of banding on ethidium bromide-cesium chloride density gradients. The ethidium bromide was removed from the banded DNA by isopropanol extraction, and residual proteins were removed by proteinase K digestion followed by phenol-chloroform extraction.

Nucleic acid hybridization. Probes were made by nick translation of plasmid DNAs (18). The standard low-stringency hybridization conditions for nitrocellulose filters from Southern blots (20) or plaque lifts (2) were 2 h of pretreatment of the filters in 4.5× SSC (1× SSC = 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% Ficoll–0.1% bovine serum albumin–0.1% polyvinylpyrrolidone at 60°C, followed by hybridization with denatured nick-translated probe in 4.5× SSC–0.02% Ficoll–0.02% bovine serum albumin–0.02% polyvinylpyrrolidone at 60°C. After hybridization for 16 to 60 h, the filters were extensively washed in 4.5× SSC at 60°C and given a final wash in 3× SSC–0.1% sodium dodecyl sulfate at room temperature. Where indicated, more stringent hybridizations were performed with 1× SSC instead of 4.5× SSC. Labeled markers of *Hin*dIII-digested λ DNA and *Taq*I-digested φX DNA were included on all blots as previously described (5). Sequence divergence between the viral portion of the clone 5CH3 and CPC-1 was determined by hybridization of nick-translated 5CH3 DNA to cloned CPC-1 and 5CH3 DNAs which had been fragmented by limited DNase treatment. After hybridization the hybrids were melted from hydroxyapatite columns, and the difference in melting temperatures was determined as described previously (6).

Molecular cloning. *Eco*RI or *Hin*dIII digests of high-molecular-weight chimpanzee DNA were size fractionated on neutral sucrose gradients. Portions of fractions from the gradients were electrophoresed on 0.7% agarose gels, blotted, and hybridized to CPC-1 probe to locate the fractions containing the optimal amounts of the desired sequences. These fractions were cloned in either λgtWES·λB or Charon 21A vectors as previously described (5). Preliminary experiments using probes to specific regions of the CPC-1 genome identified two major *Eco*RI fragments in chimpanzee DNA, at 2.4 and 2.0 kilobases (kb), as being related to the 3′ end of CPC-1. Three independent clones of the 2.4-kb fragment were obtained by cloning the 2 to 3-kb region of an *Eco*RI digest. After subcloning in pBR322, this fragment was used as a probe to isolate clones from the 5- to 7-kb region of a *Hin*dIII digest or the 7- to 10-kb region from a partial *Eco*RI digest. Fragments were isolated from these clones and subcloned into the plasmids pBR322 or pBR325 as previously described.

Electron microscopy. Heteroduplexes of cloned DNA were prepared according to methods previously described (23). The heteroduplexes were spread from formamide by the basic protein film method with cytochrome *c* and isodenaturing conditions (9).

RESULTS

Identification of an endogenous chimpanzee retroviral genome related to CPC-1. One of the clones of chimpanzee DNA, 5CH3, obtained from the partial *Eco*RI digest contains an essentially complete retroviral genome. Since the melting temperature of hybrids of 5CH3 and CPC-1 DNA is reduced by 27°C relative to the melting temperature of reassociated 5CH3 DNA, we estimate that the overall sequence divergence is approximately 27%. In spite of this large divergence we can align the sequence of this clone with the CPC-1 viral genome as summarized in Fig. 1. By electron microscopy of heteroduplexes under low-stringency spreading conditions, we found a 4.5-kb region of homology between 5CH3 and CPC-1 without any observable interruption due to insertion or deletion (Fig. 2 and 3). Heteroduplexes were prepared between an 8.25-kb portion of 5CH3, which is composed of the 5.9- and 2.35-kb *Eco*RI fragments after subcloning in the plasmid pBR325, and the 1.3- and 3.4-kb *Eco*RI fragments of CPC-1, each separately subcloned in pBR322. All of the viral sequences were subcloned in the same orientation so that duplex could form between both the viral and plasmid regions. Since pBR325 was derived by inserting extra sequences into the *Eco*RI site of pBR322, this construction of the heteroduplexes guaranteed the presence of single-stranded loops precisely at the ends of the CPC-1 sequences. Furthermore, the presence of the plasmid sequences helped to stabilize the heteroduplex and provided an internal size standard. Beyond the ends of regions A and B of Fig. 1, additional homology could be detected by low-stringency hybridization of labeled 5CH3 DNA to restriction digests of CPC-1 DNA. These experiments (not shown) demonstrated that detectable homology terminates between 0.4 and 0.7 kb from the 5′ end of

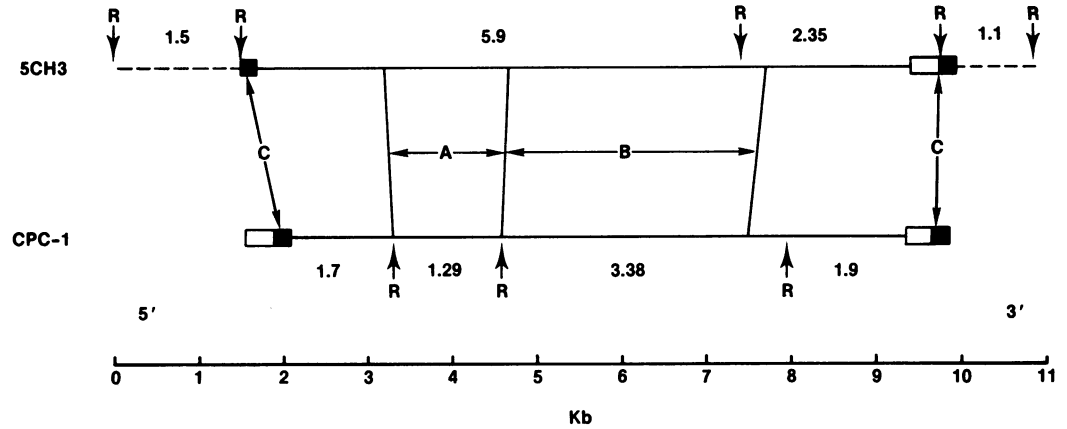

FIG. 1. Alignment of the chimpanzee DNA clone 5CH3 with the genome of the colobus type C virus CPC-1. The top line shows the four *Eco*RI fragments of the clone together with their sizes in kilobases. The *Eco*RI sites are denoted by arrows labeled R. The solid line indicates the viral sequences, and the dashed lines indicate nonviral sequences. The LTR sequences are indicated by boxes. The second line presents the map of CPC-1, again indicating the *Eco*RI fragments. The regions of alignment demonstrated in Fig. 2 through 4 are indicated by the letters A, B, and C. A 4.5-kb region of homology as determined by heteroduplex analysis is indicated by regions A and B. Region C at the 3′ end of the LTR sequences is determined by nucleotide sequencing. The bottom line defines a coordinate scale measured in kilobases from the 5′ end of the clone.

CPC-1 and between 0.4 and 0.8 kb from the 3′ end of CPC-1. Thus the region of homology detectable by hybridization contains between 6.8 and 7.5 kb of the CPC-1 genome but does not include the long terminal repeat (LTR) sequences.

To demonstrate the existence of an LTR sequence in 5CH3 we have sequenced the 5′ ends of the 5.9- and 1.1-kb *Eco*RI fragments as well as the 3′ ends of the 1.5- and 2.35-kb fragments. The sequences at the 5′ ends of the 5.9- and 1.1-kb *Eco*RI fragments are identical for at least 148 nucleotides (Fig. 4). Furthermore, comparison with the strong stop sequences (13) of CPC-1 and MAC-1, the 5′ ends of which correspond to the 5′ end of the viral RNAs and the 3′ ends of which define the 3′ end of their LTR sequences, shows extensive homology. Between nucleotides 88 and 187 there is 73% homology of 5CH3 with CPC-1 and 77% homology with MAC-1. Additional homology to CPC-1 occurs from nucleotides 15 to 37 and includes a Goldberg-Hogness sequence at nucleotides 31 to 37. Thus we have a directly repeated sequence which is clearly related to part of the CPC-1 LTR sequence. The alignment of the 5CH3 sequence (which is shown as region C in Fig. 1) with the CPC-1 sequence allows us to define the 3′ end of the LTR as occurring 187 nucleotides to the 3′ side of the *Eco*RI sites. In addition, transcription of a viral RNA would be predicted to begin about 63 nucleotides to the 3′ side of the *Eco*RI site, which occurs at 1.5 kb from the 5′ end of 5CH3.

Since *Eco*RI apparently cuts within the LTR, the 3′ ends of the 1.5- and 2.4-kb *Eco*RI fragments would be expected to contain the 5′ portion of the LTR and thus be homologous. However, we were unable to detect any hybridization between the two fragments. This result suggests either that the bulk of the LTR sequence has been deleted from the 1.5-kb *Eco*RI fragment or that the presence of the 1.5-kb fragment in the clone is due to its accidental ligation to the remainder of 5CH3 during cloning. Except in the improbable case of a deletion extending exactly to the *Eco*RI site, there should be at least some sequence homology between the two fragments if the presence of the 1.5-kb fragment is not fortuitous. Comparison of the nucleotides adjacent to the *Eco*RI site yields an ambiguous result. Ten of the first 18 nucleotides adjacent to the *Eco*RI site (nucleotides −2 to −19) are identical, but beyond this point in the 5′ direction there is essentially no sequence homology. Thus if we were to propose a deletion model it should occur at about nucleotide −20. The probability of a chance occurrence of 10 homologous nucleotides in 18 is 4×10^{-3}, assuming that all nucleotides are equally probable. Thus, by itself the observed homology does not provide compelling evidence for either alternative. However, since there appear to be no differences between the two LTR sequences for nucleotides 1 to 147, it seems unlikely that the first 18 nucleotides on the other side of the *Eco*RI site would accumulate eight changes. Thus it appears that the presence of the 1.5-kb fragment is fortuitous. However, we cannot exclude the possibility that there was a deletion

of all but 20 nucleotides of the LTR sequence which should have occurred in the 1.5-kb fragment. Such a deletion would remove most of the putative promoter sequences and might allow the cell to inactivate the viral genome. Regardless of the nature of the 1.5-kb fragment and whether there is a complete promoter sequence at the 5′ end of the viral sequence, it appears that the remainder of 5CH3 contains all the sequences needed to encode a retroviral RNA. Furthermore, the combined 5.9- and 2.35-kb *Eco*RI fragments can be used to make a probe representative of the whole genome without any nonviral flanking sequence.

Identification of related sequences in chimpanzee DNA. Having identified 5CH3 as a retrovirus-like sequence related to CPC-1, we then determined that this clone is representative of the majority of the chimpanzee sequences detected with CPC-1 probe. The most direct evidence is provided by the results of Fig. 5A and B, in which we see that the patterns of chimpan-

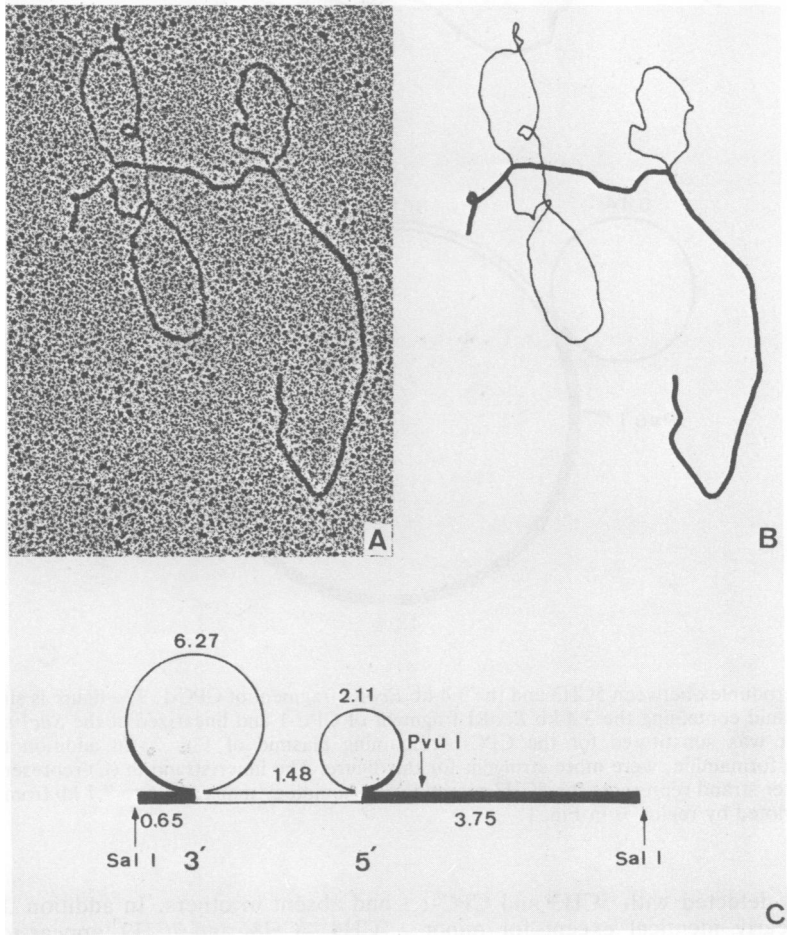

FIG. 2. Heteroduplex between 5CH3 and the 1.3-kb *Eco*RI fragment of CPC-1. Plasmids containing the 5.9- and 2.35-kb *Eco*RI fragments of 5CH3 and the 1.3-kb *Eco*RI fragment of CPC-1 were linearized by cutting with *Sal*I. Heteroduplexes were formed and spread with 30% formamide. Although not evident in the heteroduplex, the 5CH3-containing plasmid was also cut at *Pvu*I to eliminate the formation of hairpins in the pBR325 sequences. (A) Typical heteroduplex molecule. (B) Interpretive drawing in which the double-stranded regions are shown with a heavier line than the single-stranded regions. (C) Measurements of the single- and double-stranded regions in kilobases. The typical standard deviation is 0.2 kb. The vector regions are shown with heavier lines. The strand containing 5CH3 is shown on the top and includes 1.2- and 0.4-kb stretches of pBR325 specific sequences in the large and small single-stranded loops, respectively. The 5′ to 3′ orientation of the CPC-1 sequences from mapping of the plasmid is shown below. The region of homology maps at 3.2 to 4.7 kb from the 5′ end of 5CH3 and is denoted by region A in Fig. 1.

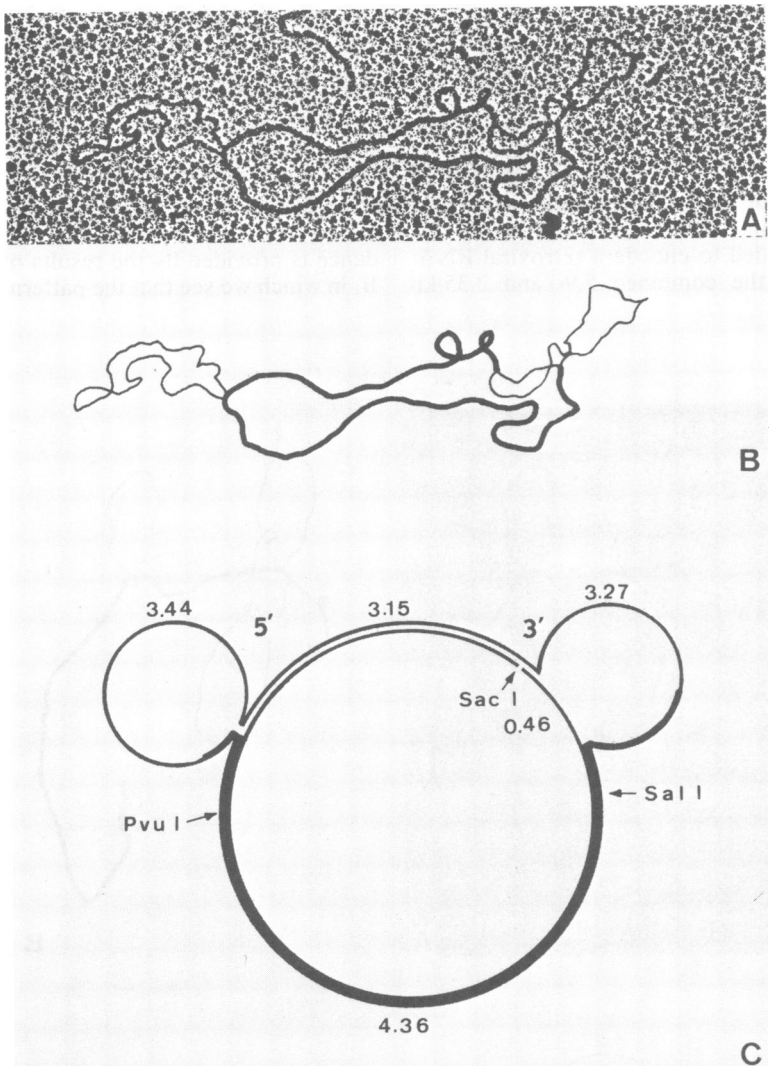

FIG. 3. Heteroduplex between 5CH3 and the 3.4-kb *Eco*RI fragment of CPC-1. The figure is similar to Fig. 2 except that plasmid containing the 3.4-kb *Eco*RI fragment of CPC-1 and linearized at the *Sac*I site within the CPC-1 fragment was substituted for the CPC-1–containing plasmid of Fig. 2. In addition the spreading conditions, 40% formamide, were more stringent for this figure. The inner strand in (C) represents the CPC-1 plasmid; the outer strand represents the 5CH3 plasmid. The homology maps at 4.6 to 7.7 kb from the 5′ end of 5CH3 and is denoted by region B in Fig. 1.

zee fragments detected with 5CH3 and CPC-1 probes are nearly identical except for minor differences in relative intensities.

Additional information can be derived from the restriction map of 5CH3 and several other clones from the 3′ end of the genome (Fig. 6). All of the clones have similar but not identical maps within the viral sequences and different maps in the 3′ flanking sequences. There are a number of restriction sites such as the *Eco*RI site at 5.4 kb, the *Sac*I sites at 9.2 and 9.65 kb, and the *Bam*HI site at 8.55 kb, which are present in some clones and absent in others. In addition three clones, 7CH4, 7CH8, and 7CH9, appear to have deletions of 1.7, 1.4, and 1.2 kb, respectively. The deletion in 7CH4 has been mapped by heteroduplexing to 5CH3. Heteroduplexes of 7CH4 with 7CH9 confirm that 7CH9 also has a deletion which maps entirely within the deletion region of 7CH4. To determine whether these deletions occur within the chimpanzee genome or during the cloning process, we examined *Eco*RI and *Hin*dIII digests of chimpanzee DNA by using probes defined by the *Eco*RI and *Hin*dIII sites at

3.7, 5.2, and 7.4 kb on the map. Figure 5D reveals two major *Eco*RI-*Hin*dIII bands of 2.5 and 2.2 kb, as well as a major *Eco*RI band of 2.0 kb. The absence of the 2.0-kb *Eco*RI band in Fig. 5C implies that it is the 2.0-kb *Eco*RI fragment of 7CH8. The presence of the 2.5-kb band in Fig. 5C implies that it spans the *Hin*dIII site at 5.2 kb, and thus it appears to correspond to the 1.2-kb deletion of 7CH9. The 2.2-kb *Eco*RI-*Hin*dIII band almost disappears in Fig. 5C, which suggests that the major contribution to the band in Fig. 5D is the fragment of 5CH3 which spans the 5.2- to 7.4-kb region of the map. However, the presence of a faint 2.2-kb band in Fig. 5C suggests that some genomic copies have a deletion of approximately 1.5 kb, comparable to 7CH8. Judging from the intensities of the bands in Fig. 5D, we estimate that 30% of the genomic copies are equivalent to 5CH3, 30% have the additional *Eco*RI site of 5.4 kb, 30% have a 1.2-kb deletion which spans the *Hin*dIII site at 5.2 kb, and the remaining 10% have other deletions mapping between 4 and 6 kb on the map. Comparison of the intensities of *Eco*RI bands with those of known amounts of cloned DNA also allows us to estimate that there are 10 to 20 copies of the 2.0-kb band and roughly four times as many copies of the 2.4-kb band which appears to be common to all of the genomic copies.

Using the information from the restriction maps as well as the existence of viral copies with

```
                -48          -40          -30          -20          -10
5CH3 2.4 kb      TGAGGGAC     AGCCGTTCAA   AAGTTTTACT   GCAAGAGCGG   GAACCAAAAG
5CH3 1.5 kb      .T.ATA.G     TAAATAAATG   GCAGACA.AG   A....TATCT   ..TG....T.
CPC-1            CC.ACTCG     CCTT......   .T...-C.GC   CA.TCATACA   TT.GATTTGT

                      10           20           30           40           50
5CH3 5.9 kb      AATTCCTTTG   TTCCCCTGTA   ACTTTCAGGC   TATAAAAAAG   CAAACACTCG
5CH3 1.1 kb          ....X    ..........   ..........   ..........   ..........
CPC-1            .TAATAAC.C   .GTA.....T   TACCCT..A.   ....T..TGT   TGTCTTGAGC

                      60           70                        80
5CH3 5.9 kb      CATTGTTCGG   GGCCXTCTTG   ---TA-TGCG   GTGGAAT---   ---GGAGGG-
5CH3 1.1 kb      ..........   ....C.....   ..........   ..........   ..........
CPC-1            .CACTACT..   A..TC..CC.   CTT..C.C..   C.CCT.GAG-   GCGT...A.T
MAC-1                           GC..C.G   C--------..   T.TC.CGAGC   AAGT...A.C

                90           100          110          120          130
5CH3 5.9 kb      ACCAGGTTCG   AACTTGTAGT   AAAGATCCTT   GCCGCTT-GG   CTTGACTCTG
5CH3 1.1 kb      ..........   ..X.......   ..N.......   X...X.....   ..X.......
CPC-1            T.........   ...C....A.   ..........   TG.A...GAC   A.....C.CT
MAC-1            T.........   ...C....A.   ..........   ..T....A.C   T.........

                140          150          160          170          180   187
5CH3 5.9 kb      XACTCTGGTG   GGGTCTTCTT   TGGGGAACTA   ACGGGTCTGA   GCATAACAC
5CH3 1.1 kb      G..NX....X
CPC-1            G.........   A.C.T....-   -.....-...   GAC..A..TG   ...C....A
MAC-1            G.........   .TC.TC.TC-   -.....-A..   .AC......G   ........T
```

FIG. 4. Comparison of sequences in the LTR region of 5CH3 with LTR sequences of CPC-1 and MAC-1. The sequence of 5CH3 was determined in the vicinity of the *Eco*RI sites at 1.5 and 9.75 kb and aligned with the sequences of CPC-1 and MAC-1. The sequences of 5CH3 are labeled with the size of the *Eco*RI fragment from which they originated. The positions of the 5CH3 sequence are numbered relative to the two *Eco*RI sites with negative numbers to the 5′ side and positive numbers to the 3′ side. Deletions introduced to align the sequences are indicated with dashes. X denotes an ambiguous position in one of the two 5CH3 sequences in which the sequence could be read as either of two nucleotides, one of which is the same as the corresponding nucleotide of the other 5CH3 sequence. The dots indicate that the nucleotide is the same as the corresponding nucleotide of the sequence of the 2.4- or 5.9-kb 5CH3 sequences. N denotes an indeterminant nucleotide. Although not actually sequenced, nucleotides 1 to 5 of the 1.1-kb 5CH3 fragment must be the same as those in the 5.9-kb fragment since they are part of the *Eco*RI site. The strong stop sequence of CPC-1 begins at nucleotide 63, whereas the strong stop sequence of MAC-1 starts at nucleotide 64. Both strong stop sequences terminate at nucleotide 187.

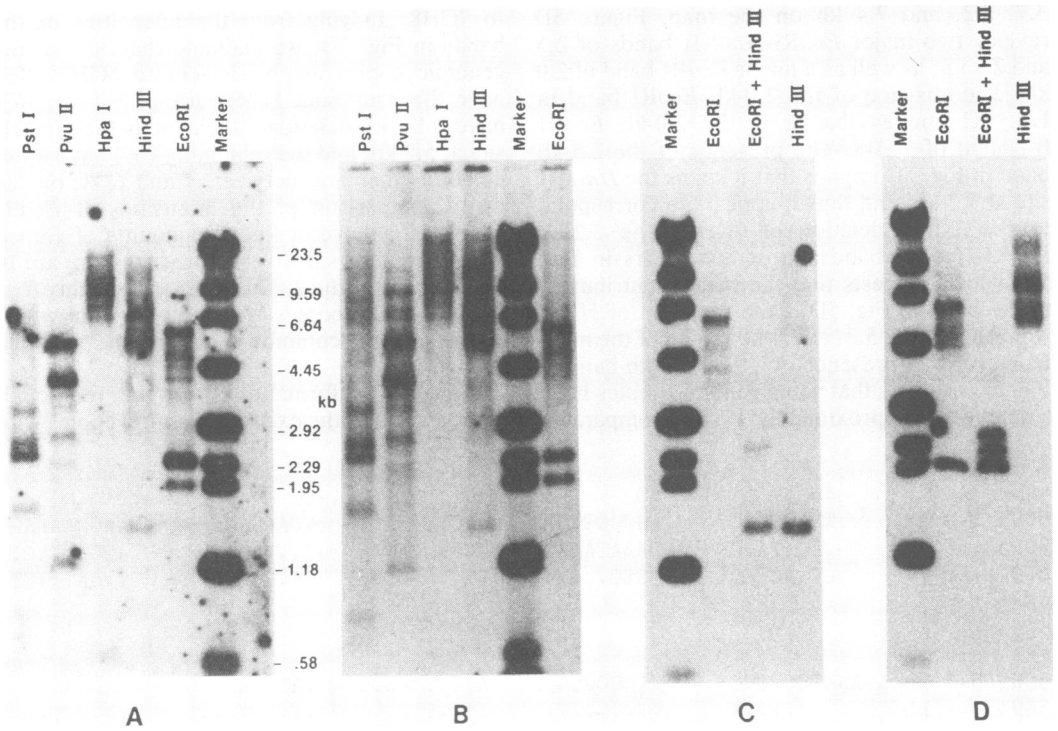

FIG. 5. Hybridization of chimpanzee DNA with CPC-1 and 5CH3 probes. Identical Southern blots of chimpanzee DNA digested with the indicated restriction enzymes were hybridized with nick-translated probes of either the 5.9- and 2.4-kb *Eco*RI fragments of 5CH3 (A) or the entire CPC-1 genome (B). The hybridization in (A) was done under more stringent conditions to minimize the contribution from distantly related sequences. The probes in (C) and (D) are individually subcloned *Hin*dIII or *Hin*dIII-*Eco*RI fragments of 5CH3, whose location is indicated in Fig. 6.

deletions, we can account for most of the bands observed in Fig. 5. Thus the 2.35- and 5.9-kb *Eco*RI bands of 5CH3 and the 2.0-kb *Eco*RI band of 7CH8 account for the major *Eco*RI bands. The *Eco*RI band of 4.0 kb is presumably the result of cutting the 5.9-kb fragment into 3.9- and 2.0-kb fragments, and the series of bands from 4.5 to 5.0 kb represent deletions from the 5.9-kb fragment. In addition, the *Hin*dIII 1.5-, *Hpa*I 0.55-, *Pvu*II 5.3- and 1.2-, and *Pst*I 2.35-, 1.68-, 0.78-, and 0.65-kb fragments of 5CH3 are all seen as relatively common bands. Some of the larger *Pvu*II and *Pst*I bands are not easily accounted for, but they presumably represent a combination of restriction site differences and deletions relative to 5CH3. The observation that the chimpanzee bands detected by 5CH3 and CPC-1 are essentially identical, as well as the fact that most of the bands can be accounted for by the restriction map of 5CH3 and closely related clones, implies that these clones are representative of a single family of sequences which constitute the chimpanzee equivalent of the CPC-1 virus.

Related sequences in other species. Since the clone 5CH3 provides a probe representative of the CPC-1 family of sequences from chimpanzee, we can now determine whether this family of sequences is present in the other apes. Hybridization of this probe to *Eco*RI digests of several primate DNAs (Fig. 7) reveals related sequences in all the primates except humans, gibbons, and orangutans. There is even hybridization to mouse DNA, which presumably reflects the distant homology that exists between primate and murine retroviruses (6). It is difficult to define precisely an upper limit on the amount of related sequence in the human, gibbon, and orangutan genomes since the signal to be detected depends both on the number of copies of viral sequence and on their extent of homology with 5CH3. We have calibrated our sensitivity by reconstruction experiments in which picogram amounts of plasmid DNAs containing *Eco*RI fragments of CPC-1 were added to microgram quantities of human DNA. These mixed DNAs were included in blots along with the unmixed DNAs. Under the conditions of Fig. 7, we detected bands containing about 0.5 pg of completely homologous DNA, 1 pg of the 3.4-kb

fragment of CPC-1, 2 pg of the 1.3-kb fragment, or 8 pg of the 3.1-kb fragment which contains the 5′ and 3′ ends of the CPC-1 genome. Since 1 pg of DNA is equivalent to 3,000 nucleotides of single copy sequence in 1 μg of cellular DNA and we have at least 3 μg of human 1 and human 3 DNAs, we should have detected a band containing 1,000 to 2,000 nucleotides of a single copy sequence if it was no more diverged than CPC-1. To improve our sensitivity we have hybridized several blots with probes of 10-fold higher specific activity. In most cases our sensitivity increases about 10-fold so that we should have detected 200 nucleotides of single copy sequences in 3 μg of cellular DNA. Under these conditions we typically saw two to five bands in the human, gibbon, and orangutan digests, with intensities corresponding to a few hundred nucleotides of single copy sequence as diverged as CPC-1. These bands have not been very reproducible, and we suspect that, if they are real, they represent sequences that are so diverged that the hybrids can barely form at our stringency of hybridization. We therefore conclude that, if the viral sequences in these apes are not more diverged than CPC-1, then the cellular genomes contain at most a single copy or probably not even a single copy. However, if the human sequences were more diverged than CPC-1, then it is improbable that they were present in the chromosomal DNA of the common ancestor of humans and chimpanzees. This hypothesis would require the viral sequences to evolve after the divergence at a rate of at least 10 times the rate of the single copy sequences. This in turn requires the rate to be at least 10 times greater than the estimated rate for fixation of mutations of neutral selective value (12). This rate argu-

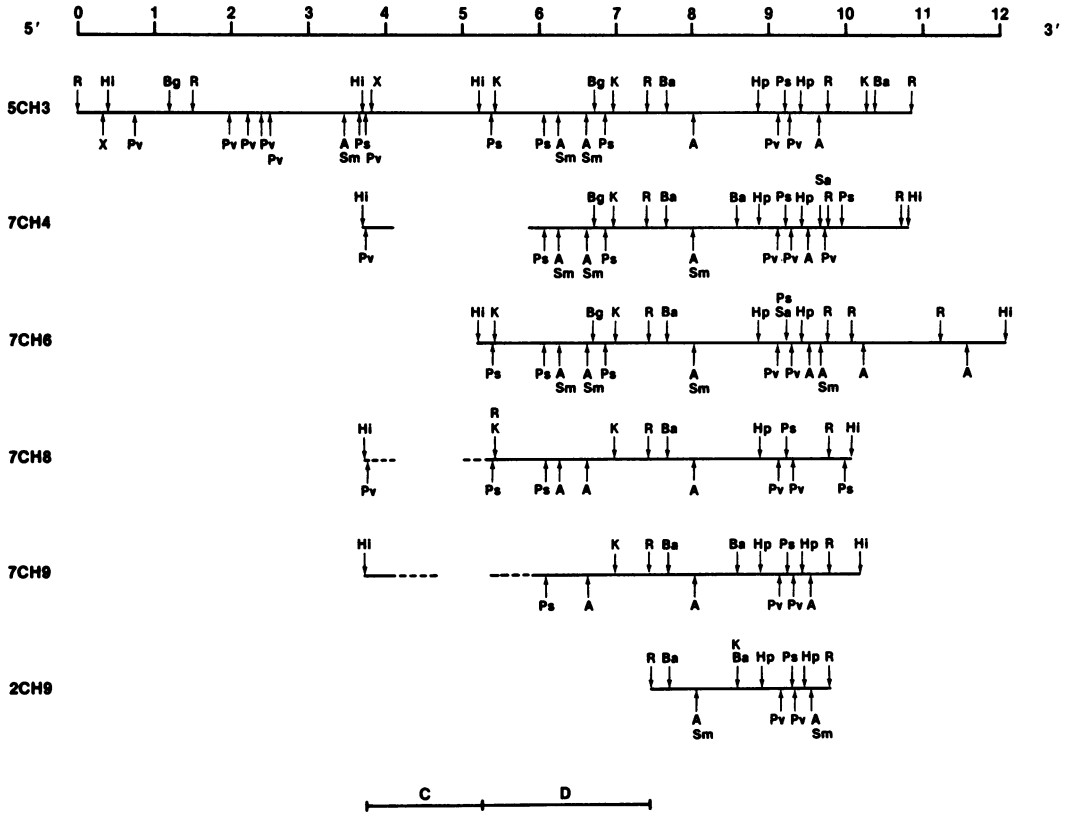

FIG. 6. Restriction map of 5CH3 and related clones. The top line provides a scale in kilobases originating at the 5′ end of the 5CH3 clone. Other clones are aligned with 5CH3, and deletions are indicated by breaks in the solid lines. Where the exact endpoint of the deletion is undetermined, dashed lines are used to represent sequence of which half is absent. The restriction enzymes are abbreviated as follows: A, *Aval*; Ba, *Bam*HI; Bg, *Bgl*II; Hi, *Hin*dIII; Hp, *Hpa*I; K, *Kpn*I; Ps, *Pst*I; Pv, *Pvu*II; R, *Eco*RI: Sa, *Sac*I; Sm, *Sma*I; and X, *Xba*I. The enzymes *Xho*I and *Sal*I do not cut 5CH3. The enzymes *Bgl*II, *Sac*I, *Sma*I, and *Xba*I were not mapped on the clones 7CH7-7CH9. The lines labeled C and D at the bottom of the figure represent the regions used as probes in Fig. 5C and D.

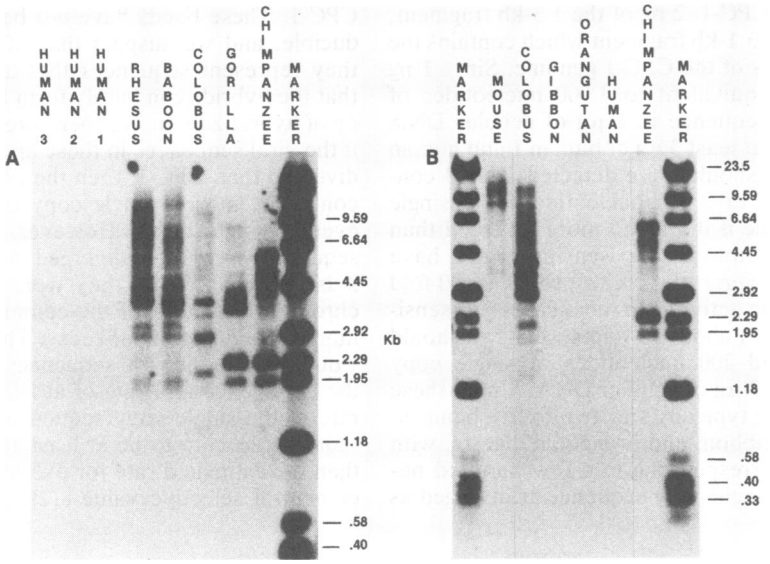

FIG. 7. Hybridization of 5CH3 probe to various DNAs. Southern blots of *Eco*RI-digested DNAs were hybridized to the combined 5.9- and 2.4-kb *Eco*RI fragments of 5CH3 labeled to a specific activity of 2×10^7 cpm/μg. All the digests contained about 1 μg of DNA except for human 1 and human 3 in (A), which contained 3 to 4 μg; the chimpanzee digest in (B), which contained 0.4 μg; and the human digest in (B), which contains 2 μg. The human DNA in (B) is the human 1 DNA of (A).

ment is subject to the assumption that moderately repeated sequences cannot evolve substantially more rapidly than single-copy sequences. We know of no data to suggest the contrary except for the data on endogenous retroviral sequences, whose evolutionary history is unclear.

Although these results weaken the hypothesis that these viral sequences have been transmitted in the germ line of primates for tens of millions of years, the results of Fig. 7 also contain information supportive of the hypothesis. If we look at the closely related species, baboon and rhesus as one pair and chimpanzee and gorilla as the other, we see that the restriction patterns within pairs are very similar, whereas between pairs they are very different. The close similarity of the chimpanzee and gorilla restriction patterns is supported by experiments (not shown) using the 2.35-kb *Eco*RI fragment as a probe, which demonstrate that the 2.35-kb *Eco*RI band of gorilla is homologous to the 2.35-kb *Eco*RI band of chimpanzee and contains *Bam*HI and *Hpa*I sites identical to the chimpanzee sites. Whereas the many faint bands larger than 5 kb in the rhesus and baboon digests may represent sites in the adjacent cellular sequences, the more intense bands below 5 kb presumably represent restriction sites contained exclusively within multiple copies of viral sequence. In addition, *Bam*HI, *Sac*I, *Bgl*II, and *Hpa*I digests (not shown) reveal eight major viral fragments of which seven are identical in baboon and rhesus DNAs. Thus the viral sequences are closely related in closely related species but quite different in more distantly related species, as would be expected for long-term germ-line transmission.

DISCUSSION

We have cloned several members of a family of retrovirus-like sequences which constitute the chimpanzee equivalent of the CPC-1 virus. These clones have the characteristic structure of an integrated retrovirus. This was demonstrated for the clone 5CH3, which has a pair of LTR sequences separated by approximately 8 kb of sequence having extensive, although distant, homology to the CPC-1 virus. Both from the Southern blots (Fig. 5) and from the restriction maps of the clones it is clear that there are multiple copies of this sequence which are closely related but not identical. Approximately one-third of the cellular copies have sustained deletions of 1 to 2 kb within the region between 4 and 6 kb on the restriction map. By comparison with the map of Moloney leukemia virus, these deletions are within the polymerase gene. Excluding restriction sites within the deletion region, the restriction maps of Fig. 6 indicate that approximately 74% of the restriction sites are the same among different copies. Using the approach of Nei and Li (16), we deduce that the average divergence between copies is approximately 5%.

The 3' copy of the LTR (assumed to begin at 9.3 kb and extend to 9.9 kb in Fig. 6) contains a large fraction of the restriction site heterogeneity. In spite of this divergence between LTR sequences of different copies of the viral sequence, the limited sequence of 5CH3 suggests that its two copies of the LTR are identical. This result suggests that either 5CH3 is a recently integrated sequence, so that the two copies of LTR which are identical upon integration have not had enough time to diverge, or else there is a cellular mechanism which maintains the homogeneity of the two LTR sequences.

When we examine other primate species we are confronted with an enigmatic result. The restriction patterns indicate that closely related species have closely related viral sequences, whereas distantly related species have distantly related viral sequences. This evolutionary pattern is supported by the strong stop sequence comparison which shows 5CH3 to be about equally diverged from the CPC-1 and MAC-1 viruses of colobus and macaque, as well as by hybridization data with CPC-1 and MAC-1 cDNAs (19, 21; Bonner, unpublished data). Although the similarity of restriction patterns provides only qualitative data, the data taken together would define a phylogenetic tree for the viral sequences which must be nearly congruent to the tree defined by the single-copy DNA of the primates. On the other hand, the apparent absence of closely related sequences in humans, gibbons, and orangutans suggests that they either were not present in the common ancestor of the apes or were deleted from these three lineages. The former alternative is difficult to reconcile with the evolutionary pattern of the related viral sequences in the other primates including chimpanzees and gorillas. The attractiveness of the deletion hypothesis depends on the details of the hypothesis. According to most phylogenies, the gibbon and orangutan lineages diverged substantially before humans, chimpanzees, and gorillas diverged from each other. In addition the gibbons appear to have branched off before the orangutan lineage. Thus it is likely that independent deletion events would have had to occur in each of the three lineages. This might have occurred easily if the common ancestor of the apes contained a single viral copy or even multiple copies in a single cluster. However, it is difficult to see how it would have occurred if the common ancestor had multiple viral copies dispersed throughout the genome unless the viral sequences were selected against in some species but not others. Even in this case, one might expect to find inactive remnants of the viral sequences rather than their complete absence.

Its apparent absence from three genera of apes suggests that the CPC-1 famly of retroviral sequences is not essential for the propagation of these apes. Similar results have been previously reported for different retroviral families in individual chickens (1) and mice (8). Since the CPC-1 family is absent in three genera which have existed for millions of years, our results suggest that in addition these viral sequences have been of no selective advantage to these apes. This is not to say, however, that no retroviral sequences are present in human DNA. Such sequences have been cloned from human DNA (14; Bonner et al., Proc. Natl. Acad. Sci. U.S.A., in press). Indeed, if the stringency of hybridization is reduced by another 7°C, 5CH3 clearly detects human sequences. However, these sequences should not be considered to belong to the CPC-1 or 5CH3 family of sequences since the same restriction fragments are detected under more stringent conditions by murine retroviral probes (Bonner, unpublished data). Thus the human sequences detected with 5CH3 are more closely related to murine retroviruses than they are to the CPC-1 family of primate retroviruses.

ACKNOWLEDGMENTS

We thank Kay Reynolds for technical assistance in the early stages of this project and Kunio Nagashima for assistance with the heteroduplex analysis.

LITERATURE CITED

1. **Astrin, S. M., E. G. Buss, and W. S. Hayward.** 1979. Endogenous viral genes are non-essential in the chicken. Nature (London) **282**:339–341.
2. **Benton, W. D., and R. W. Davis.** 1977. Screening λ gt recombinant clones by hybridization to single plaques in situ. Science **196**:180–182.
3. **Benveniste, R. E., and G. J. Todaro.** 1976. Evolution of type C viral genes: evidence for the Asian origin of man. Nature (London) **261**:101–108.
4. **Benveniste, R. E., and G. J. Todaro.** 1977. Evolution of primate oncornaviruses: an endogenous virus from langurs (*Presbytis* spp.) with related virogene sequences in other Old World monkeys. Proc. Natl. Acad. Sci. U.S.A. **74**:4557–4561.
5. **Birkenmeier, E. H., T. I. Bonner, K. Reynolds, G. H. Searfoss, and G. J. Todaro.** 1982. Colobus type C virus: molecular cloning of unintegrated viral DNA and characterization of the endogenous viral genomes of colobus. J. Virol. **41**:842–854.
6. **Bonner, T. I., and G. J. Todaro.** 1980. The evolution of baboon endogenous type C virus: related sequences in the DNA of distant species. Virology **103**:217–227.
7. **Chattopadhyay, S. K., M. W. Cloyd, D. L. Linemeyer, M. R. Lander, E. Rands, and D. R. Lowy.** 1982. Cellular origin and role of mink cell focus forming viruses in murine thymic lymphomas. Nature (London) **295**:25–31.
8. **Cohen, J. C., and H. E. Varmus.** 1979. Endogenous mammary tumour virus DNA varies among wild mice and segregates during inbreeding. Nature (London) **278**:418–423.
9. **Davis, R. W., M. Simon, and N. Davidson.** 1971. Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. Methods Enzymol. **21**:413–428.

10. **Frisby, D. P., R. A. Weiss, M. Roussel, and D. Stehelin.** 1979. The distribution of endogenous chicken retrovirus sequences in the DNA of Galliform birds does not coincide with avian phylogenetic relationships. Cell **17**:623–634.
11. **Gardner, M. B.** 1980. Historical background, p. 1-46. *In* J. R. Stephenson (ed.), Molecular biology of RNA tumor viruses. Academic Press, Inc., New York.
12. **Kimura, M.** 1968. Evolutionary rate at the molecular level. Nature (London) **217**:624–626.
13. **Lovinger, G. G., G. E. Mark, G. J. Todaro, and G. Schochetman.** 1981. 5'-Terminal nucleotide noncoding sequences of retroviruses: relatedness of two Old World primate type C viruses and avian spleen neurosis virus. J. Virol. **39**:238–247.
14. **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. U.S.A. **78**:4892–4896.
15. **Neel, B. G., W. S. Hayward, H. L. Robinson, J. Fang, and S. M. Astrin.** 1981. Avian leukosis virus-induced tumors have common proviral integration sites and synthesize discrete new RNAs: oncogenesis by promoter insertion. Cell **23**:323–334.
16. **Nei, M., and W.-H. Li.** 1979. Mathematical model for studying genetic variation in terms of restriction nucleases. Proc. Natl. Acad. Sci. U.S.A. **76**:5269–5273.
17. **Rabin, H., C. V. Benton, M. A. Tainsky, N. R. Rice, and R. V. Gilden.** 1979. Isolation and characterization of an endogenous virus from Rhesus monkeys. Science **204**:841–842.
18. **Rigby, P. W., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
19. **Sherwin, S. A., and G. J. Todaro.** 1979. A new endogenous primate type C virus isolated from the Old World monkey *Colobus polykomos*. Proc. Natl. Acad. Sci. U.S.A. **76**:5041–5045.
20. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
21. **Todaro, G. J., R. E. Benveniste, S. A. Sherwin, and C. J. Sherr.** 1978. MAC-1, a new genetically transmitted type C virus of primates: "low frequence" activation from stumptail monkey cell cultures. Cell **13**:775–782.
22. **Todaro, G. J., C. J. Sherr, R. E. Benveniste, M. M. Lieber, and J. C. Melnick.** 1974. Type C viruses of baboon: isolation from normal cell cultures. Cell **2**:55–61.
23. **Young, H. A., M. A. Gonda, D. DeFeo, R. W. Ellis, K. Nagashima, and E. M. Scolnick.** 1980. Heteroduplex analysis of cloned rat endogenous replication-defective (30S) retrovirus and Harvey murine sarcoma virus. Virology **107**:89–99.