VIROLOGY 213, 395–404 (1995)

HERV-H Endogenous Retroviruses: Presence in the New World Branch but Amplification in the Old World Primate Lineage

DIXIE L. MAGER[1] and J. DOUGLAS FREEMAN

Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, British Columbia V5Z 1L3; and Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada

Received July 5, 1995; accepted September 5, 1995

The evolutionary origin and age of the HERV-H family of human endogenous retrovirus-like sequences was investigated in this study. HERV-H elements exist in approximately 900 partially deleted copies and 50–100 more intact forms in humans and Old World monkeys. However, their possible presence in more divergent species is unknown. We have isolated a 1.6-kb genomic DNA segment from the New World monkey marmoset that had been PCR amplified using human HERV-H primers. DNA and protein comparisons and database searches indicate that this marmoset clone is more closely related to human HERV-H elements than to any other sequence, indicating that HERV-H-related sequences do exist in New World monkeys. In contrast to the high copy numbers of deleted elements in Old World primates, Southern blot analysis shows that such elements are present in less than 50 copies in two different species of New World monkey. To estimate evolutionary ages of the common deleted form of the element, a selected DNA segment from the *pol* region was compared from multiple human HERV-H elements. This comparison suggests that many HERV-H elements of the abundant deleted subfamily integrated approximately 30–35 million years ago. Very similar percentage divergence values between 5′ and 3′ long terminal repeats of individual elements of the deleted subfamily also suggest that these elements are close in age. These results indicate that HERV-H elements first appeared in the germline prior to the New World/Old World divergence over 40 million years ago. Interestingly, they remained in low numbers in the New World branch while a subfamily underwent a major amplification in Old World primates before the time of divergence of hominoids from Old World monkeys.

INTRODUCTION

It has been recognized for over 20 years that human DNA contains sequences related to retroviruses (Benveniste and Todaro, 1974). Since the first such sequences were isolated in 1981 (Martin *et al.,* 1981), it has become apparent that the genome harbors a variety of distinct families of human endogenous retroviral elements (HERVs), ranging in number from 1 to 1000 dispersed copies (for reviews see Larsson *et al.,* 1989; Wilkinson *et al.,* 1994). The origin of these sequences, however, remains obscure. Evolutionary analyses of HERV families have consisted primarily of Southern hybridizations to determine in which primate species a particular family is present. In most cases, these studies have found evidence for the HERV family in Old World monkeys but not in the more divergent New World monkeys (Mariani-Constantini *et al.,* 1989; Kannan *et al.,* 1991; Leib-Mösch *et al.,* 1992; Goodchild *et al.,* 1993; Dangel *et al.,* 1994). Genomic PCR analyses performed to look for the presence of particular elements in New World monkeys have also been negative (Shih *et al.,* 1991; Goodchild *et al.,* 1993; Haltmeier *et al.,* 1995). These types of studies have led investigators to conclude that most HERV families are not present in the New World lineage and hence are approximately 30–40 million years old. However, in some cases such a conclusion may not be warranted since both genomic Southern hybridizations and PCR analysis could give negative results due to sequence divergence between the New and the Old World primate species. Noncoding unique DNA segments (e.g., nonfunctional pseudogenes) differ by 12–13% between humans and New World monkeys (Koop *et al.,* 1986) and this value could be higher for interspersed repetitive elements. For example, long interspersed elements (LINEs) have diverged faster between species than within a species, thereby demonstrating the phenomenon of concerted evolution (Hutchison *et al.,* 1989). Although it has not been specifically studied, HERV families could display similar patterns of divergence.

We are therefore interested in examining the evolutionary history of HERVs using methods not directly reliant on close sequence similarity between species. The HERV-H (formerly RTVL-H) family of endogenous retroviral elements is the most abundant HERV family known, being present in approximately 1000 copies per haploid genome with a similar number of solitary long terminal repeats (LTRs) in humans, apes, and Old World monkeys

[1] To whom correspondence and reprint requests should be addressed at Terry Fox Laboratory, B.C. Cancer Agency, 601 West 10th Ave., Vancouver, B.C. V5Z 1L3. Fax: (604) 877-0712. E-mail: dixie@unix-g.ubc.ca.

0042-6822/95 $12.00

(Mager and Henthorn, 1984; Goodchild *et al.,* 1993). Most HERV-H elements in humans appear to be derived from a partially deleted precursor element that amplified in the genome (Wilkinson *et al.,* 1990; Goodchild *et al.,* 1993). The relative uniformity in copy number in Old World monkeys and humans and genomic PCR measurements of the age of a few individual elements have led to the conclusion that the majority of elements are at least 25 million years old (Goodchild *et al.,* 1993). This evolutionary age, however, is only a minimum estimate for the reasons outlined above. As is the case with most other HERV families that have been examined, Southern analysis of New World monkey DNA with HERV-H probes detects very weak or no hybridizing bands (Goodchild *et al.,* 1993; Wilkinson *et al.,* 1993). However, genomic PCR results suggest that HERV-H-related elements may be present in New World primates (Wilkinson *et al.,* 1993). In this study, we demonstrate that some HERV-H elements are indeed present in New World monkeys but at much lower copy numbers than in the Old World lineages. We also use two different sequence comparison methods to estimate age of HERV-H elements of the major deleted subfamily. These results suggest that the major expansion of HERV-H elements occurred in a relatively narrow window of time in the Old World primate lineage after separation from the New World branch.

MATERIALS AND METHODS

DNA sources and Southern analysis

New World monkey DNA was isolated from two different cell lines obtained from the American Type Culture Collection (ATCC). The marmoset cell line was B95-8, Epstein–Barr virus-transformed leukocytes, and the owl monkey cell line was OMK, kidney fibroblasts. Cells were cultured according to ATCC recommendations and DNA was isolated by standard procedures (Sambrook *et al.,* 1989). Eight micrograms of each DNA was digested with *Eco*RI, divided in half, and run on duplicate lanes of a 0.8% agarose gel. DNA was transferred to a Zetaprobe membrane (Bio-Rad) and hybridized to the two probes discussed in the text. Probes were labeled by the random primer method (Feinberg and Vogelstein, 1983) and 3–4 $\times$ 10^6 cpm probe/ml was used in the hybridizations. Hybridizations were performed as described previously (Mager and Goodchild, 1989) at a temperature of 60°. The final washes were done at the same temperature in 1$\times$ SSC, 0.3% SDS (1$\times$ SSC is 0.15 *M* NaCl/0.015 *M* sodium citrate).

HERV-H clones

The HERV-H clones used for the *pol* sequence analysis have all been reported previously and were the following: RTVL-H1 (Mager and Henthorn, 1984), RTVL-H2 (Mager and Freeman, 1987), RTVL-H3 and RTVL-H4 (Mager and Freeman, 1987; Mager, 1989), and 86K-24 (Fraser *et al.,* 1988), all derived from genomic libraries and all of the common deleted type. Z11837 (Accession No.) is an unpublished genomic sequence deposited in Genbank and is also of the deleted type. cDNA clones of the deleted type, N10-1, 5, 8, 13, and 14, were isolated from an NTera2D1 teratocarcinoma cell cDNA library (Wilkinson *et al.,* 1990). RGH1 and RGH2 were isolated from a genomic library and are of the intact type (Hirose *et al.,* 1993). Two other genomic clones of the intact type were RTVL-Hp1 (Wilkinson *et al.,* 1993) and env1a (Wilkinson and Mager, 1993).

Clones for which both LTRs were available for the LTR comparisons were RTVL-H1–4, and RGH2, as well as cH-4 (Mager and Goodchild, 1989; Mager, 1989) and F22, an unpublished genomic clone from our laboratory of the more intact type.

Sequence analysis

The PCR-amplified fragment from marmoset was subcloned and the termini of six individual subclones were sequenced. One was then completely sequenced by constructing smaller subclones and exonuclease III-generated deletion clones. The *pol* sequences from RTVL-H2, RGH1, RGH2, and Z11837 were reported previously. Regions encompassing the *pol* segment of interest have been sequenced in our laboratory from the clones RTVL-H1, RTVL-Hp1, and env1a (unpublished data). For the cDNA clones and for RTVL-H3, H4, and 86K-24, the selected region was sequenced using an oligonucleotide primer oriented in the 3′ to 5′ direction with respect to HERV-H and having the sequence 5′TGGCTGAGCTTGGTGAGGTGTG3′. The 3′ end of this primer is located 492 nucleotides 3′ to the region C deletion shown in Fig. 1. In some cases, portions were also sequenced by using vector primers after subcloning small fragments. The LTR sequences of RTVL-H1–4 and RGH2 have been reported in the studies cited above. For cH-4 and F22, subclones containing the LTRs were sequenced.

Sequences were determined using the dideoxy chain termination method with double-stranded templates (Tabor and Richardson, 1987). Sequences were analyzed using software provided by the Genetics Computer Group (GCG) (Devereux *et al.,* 1984) and homology searches to the databases were conducted using the BLAST server at NCBI.

For the purposes of this study, primate evolutionary branchpoints were taken from Gingerich (1984) and Sibley and Ahlquist (1987) and are as follows: 40 million years ago (MYr) for the platyrrhine/catarrhine (New World/Old World) divergence, 25–30 MYr for the divergence of hominoids from Old World monkeys, and 17–20 MYr for the divergence of gibbons from the higher apes.

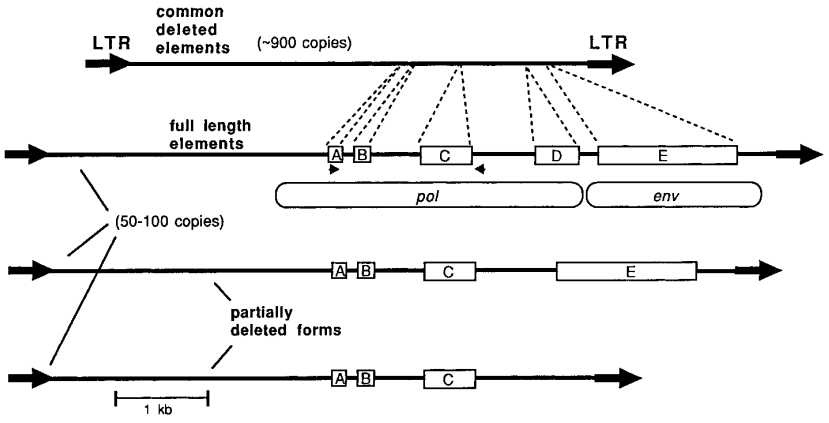

FIG. 1. Structures of various HERV-H elements that have been mapped or isolated. Boxes A–E represent segments that are deleted in the most abundant class. Large arrows at the termini are the LTRs and the small arrows shown under the fully intact element show locations of primers used for genomic PCR in Wilkinson *et al.* (1993) and discussed in the text. Extents of the *pol* and *env* regions are indicated.

RESULTS

HERV-H-related elements are present in New World monkeys

Figure 1 shows a diagram of HERV-H elements, illustrating the common deleted type of 5.8 kb, with the five deletions A–E, and other more intact forms that have been isolated, including a full-length element of 8.8 kb (Hirose *et al.,* 1993; Wilkinson *et al.,* 1993; Wilkinson and Mager, 1993; D. Mager, unpublished data). In a previous study (Wilkinson *et al.,* 1993), we used PCR on genomic DNA from various primates using a 5′ primer from within region A and a 3′ primer downstream of region C (see Fig. 1). Our original aim was to try to determine the spectrum of HERV-H forms present in different primates that contain region A. In that study, several PCR products were obtained in human and the Old World primates illustrating the complexity of partially deleted forms that exist in these species. Interestingly, a single product of 1.6 kb, the size expected for a structurally intact form (that is, containing regions B and C as well as region A) was obtained from a New World monkey DNA sample from marmoset. This result suggested that HERV-H elements are present in New World monkeys. However, since *pol* is the most highly conserved region among different retroviruses, the possibility exists that the HERV-H PCR primers used could have amplified a different type of retroviral element. Thus, to further investigate the possible existence of HERV-H elements in New World monkey, we have cloned the 1.6-kb fragment from marmoset and mapped and sequenced the termini of six individual subclones. This analysis revealed two classes of fragments, differing in DNA sequence by ~10%. A representative clone from one class was sequenced in its entirety and compared to the human HERV-H sequence RGH2 which is a full-length structurally intact element (Hirose *et al.,* 1993). A dot matrix comparison of the two sequences is shown in Fig. 2. As is evident, significant similarity extends throughout the sequence with the overall level of identity being 74%. The other marmoset sequence type is 73% identical to RGH2 (data not shown). Results of database searches indicated that the marmoset DNA sequence is more closely related to HERV-H than to other types of sequences. Other than HERV-H elements, the most closely related DNA sequences in the databases are also retroviral in origin but are only 55–63% identical in limited segments to the marmoset sequence. At the protein level, the marmoset sequence is also most closely related to human HERV-H. Figure 3 shows a comparison of the deduced amino acid sequences of the marmoset clone and a human HERV-H element. The overall level of amino acid identity is 67.5%. Comparisons of the marmoset protein sequence to the next most closely related retroviruses, Moloney murine leukemia virus or baboon endogenous virus, give identities of approximately 36% in both cases (data not shown). Thus, the marmoset clones we have

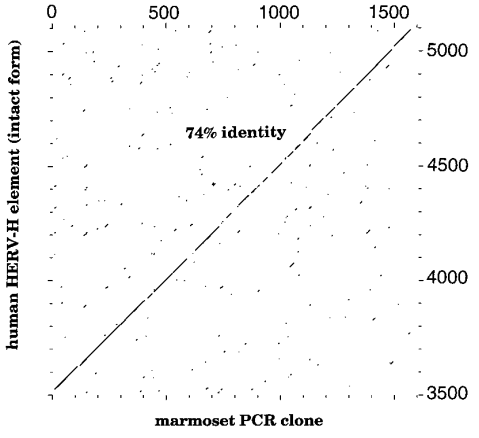

FIG. 2. Comparison of a marmoset PCR clone to a human HERV-H element. A dot matrix comparison plot run at a stringency of 14 matches in a window of 21 nucleotides is shown. The human element is RGH2 (Hirose *et al.,* 1993).

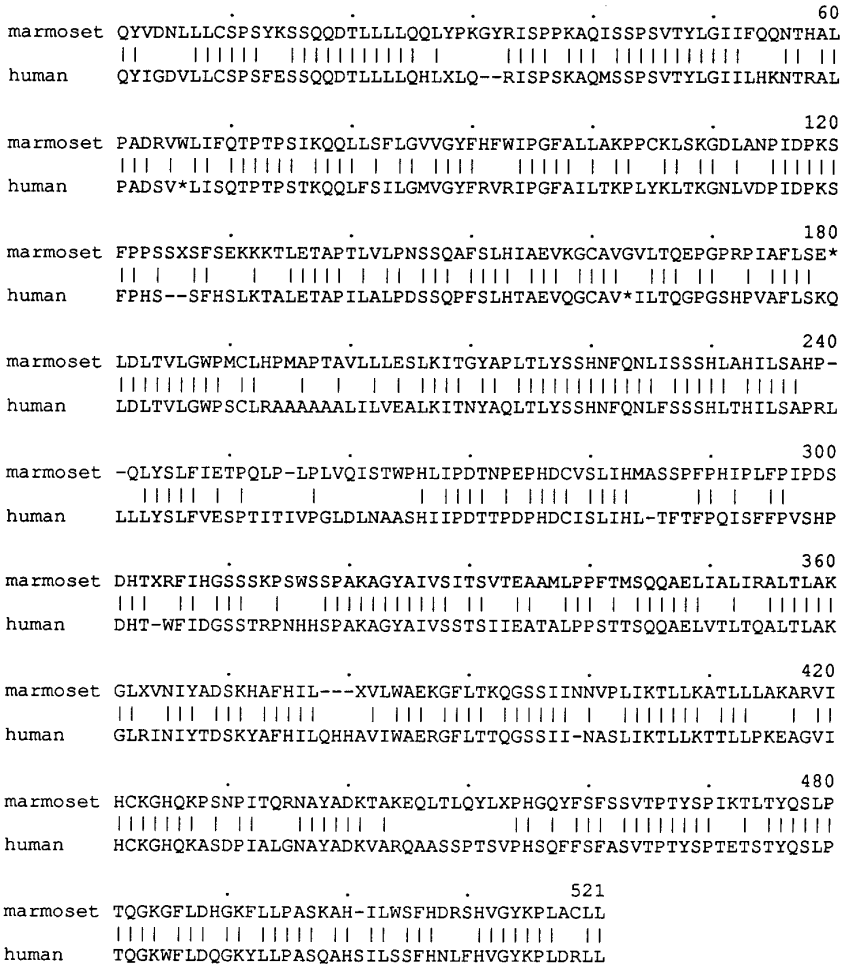

FIG. 3. Amino acid comparison of the marmoset sequence to a human HERV-H element. For this comparison, the terminal regions corresponding to the human PCR primers used to amplify the marmoset clone were not included. The human sequence is translated from a composite of RGH2 (positions 1–116) and RGH1 (positions 117–521). The partial clone RGH1 was used here since it has fewer termination codons and frame shifts compared to the full-length clone RGH2 (Hirose *et al.,* 1993). Positions shown with an X represent insertions of one or two spaces in the nucleotide sequences necessary to maintain the correct reading frame. Asterisks show termination codons.

isolated appear to be representatives of the HERV-H family of elements in New World monkeys.

HERV-H copy numbers in New World monkeys are low

In our previous study (Wilkinson *et al.,* 1993), the only segment that could be amplified by PCR from marmoset using the primers in Fig. 1 represented an intact form. This result may indicate that only intact elements and no deleted forms are present in this species. However, the primers used in that study will only amplify elements that contain region A and hence will not detect the common deleted form. To estimate the copy number of HERV-H elements in New World monkeys and to determine if deleted forms are present, two different probes were isolated from the marmoset clone for use in genomic Southern blot analysis. The location of these probes relative to the common HERV-H deletions, A, B, and C, is shown in Fig. 4C. Probe 1 is a 260-bp *Dra*II/*Eco*NI fragment contained within region C. This probe will detect only those elements that have region C intact. Probe 2 is a 350-bp *Nco*I/*Pvu*II fragment located between the B and C deletions. This probe should detect both intact and deleted elements. Genomic DNA from two New World monkey species, marmoset and owl monkey, was digested with *Eco*RI, Southern transferred, and hybridized to the two probes. Several observations can be made from the results shown in Figs. 4A and 4B. First, since probe 2 detects more bands than probe 1 in both species, there appear to be both intact and deleted HERV-H forms in New World monkeys. Second, there are greatly reduced numbers of HERV-H elements in the New World monkeys relative to the Old World species. Several different methods have previously estimated the copy number in humans and other Old World primates to be ~1000 elements with 50–100 of these being the more intact forms (Hirose *et al.,* 1993; Wilkinson *et al.,* 1993). In con-

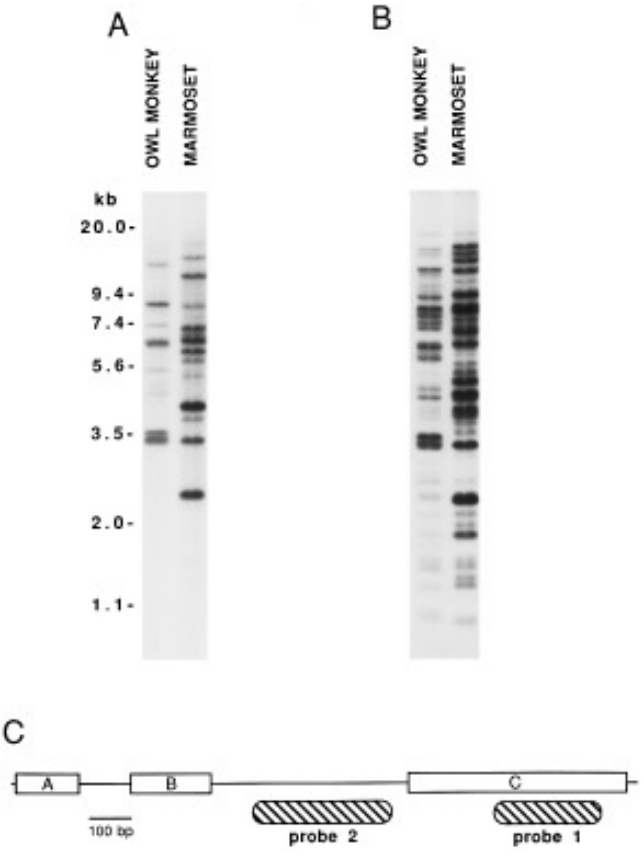

FIG. 4. Southern analysis of New World monkey DNA using marmoset probes. (A) Owl monkey and marmoset genomic DNA samples hybridized to probe 1. (B) Duplicate DNA lanes hybridized to probe 2. (C) Map showing the locations of the probes relative to the common HERV-H deletions.

trast, the banding patterns observed in Fig. 4B indicate that approximately 25–50 HERV-H-related sequences are present in the two New World species and Fig. 4A shows that only 10–15 of these are of the structurally more intact type (i.e., containing region C). It should be noted that both probes detect fewer strongly hybridizing bands in owl monkey compared to marmoset. This may indicate a lower number of elements in owl monkey or it could reflect sequence divergence since the probes were derived from marmoset. In either event, these data show that, while HERV-H family members exist in New World monkeys, they are present in much lower numbers.

HERV-H-deleted subfamily age based on *pol* sequence comparisons

The above results raise the question of when and why the HERV-H family greatly expanded in the Old World lineage while remaining in low numbers in New World species. We have previously shown that most HERV-H elements in humans appear to contain the five major deletions shown in Fig. 1 as well as other shared point mutations and frame shifts (Wilkinson *et al.,* 1990, 1993). This finding suggests that many elements were derived from a common structurally defective progenitor. To examine this further, we performed sequence comparisons of a segment of the *pol* gene from multiple HERV-H elements which do and do not contain the commonly deleted regions. The region chosen is a 305-bp segment located downstream of the C deletion. This region has been sequenced from 15 different elements, including 5 cDNA clones derived from an NTera2D1 teratocarcinoma cell library (Wilkinson *et al.,* 1990) and 10 genomic clones isolated by ourselves or others. Of these 15, 11 have been shown to be of the deleted type by mapping and/or sequence analysis and 4 have been identified as representing the more intact forms. The sequences were compared and Fig. 5 is a dendrogram based on sequence differences. It can be seen immediately that the sequences fall into two distinct groups. The group of 4 sequences at the bottom of the diagram are all derived from structurally intact elements, whereas the other 11 are of the deleted type. The fact that the sequences cluster in this way is additional evidence that the deleted elements were indeed derived from a common progenitor sequence and represent a distinct subfamily.

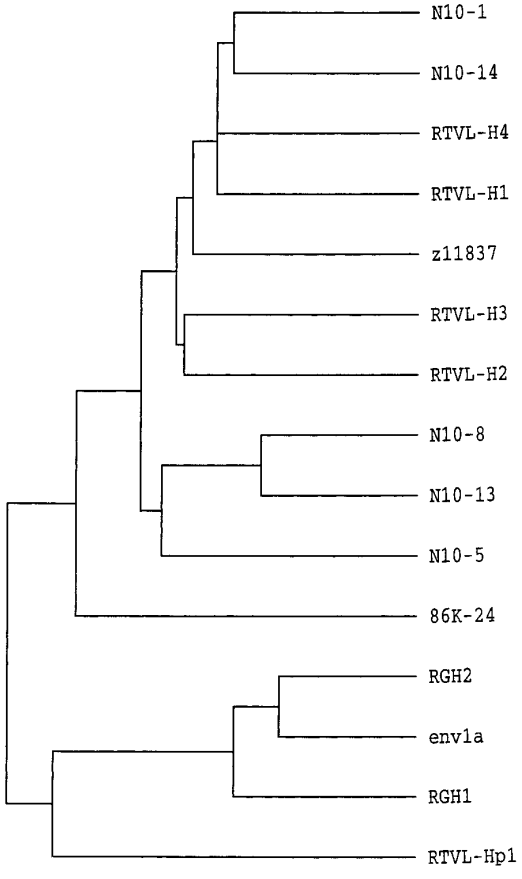

FIG. 5. Relatedness of HERV-H *pol* segments. DNA segments from 15 human HERV-H elements were compared using the GCG program Pileup which generates a dendrogram based on sequence differences. Note that the diagram is not a true phylogenetic tree.

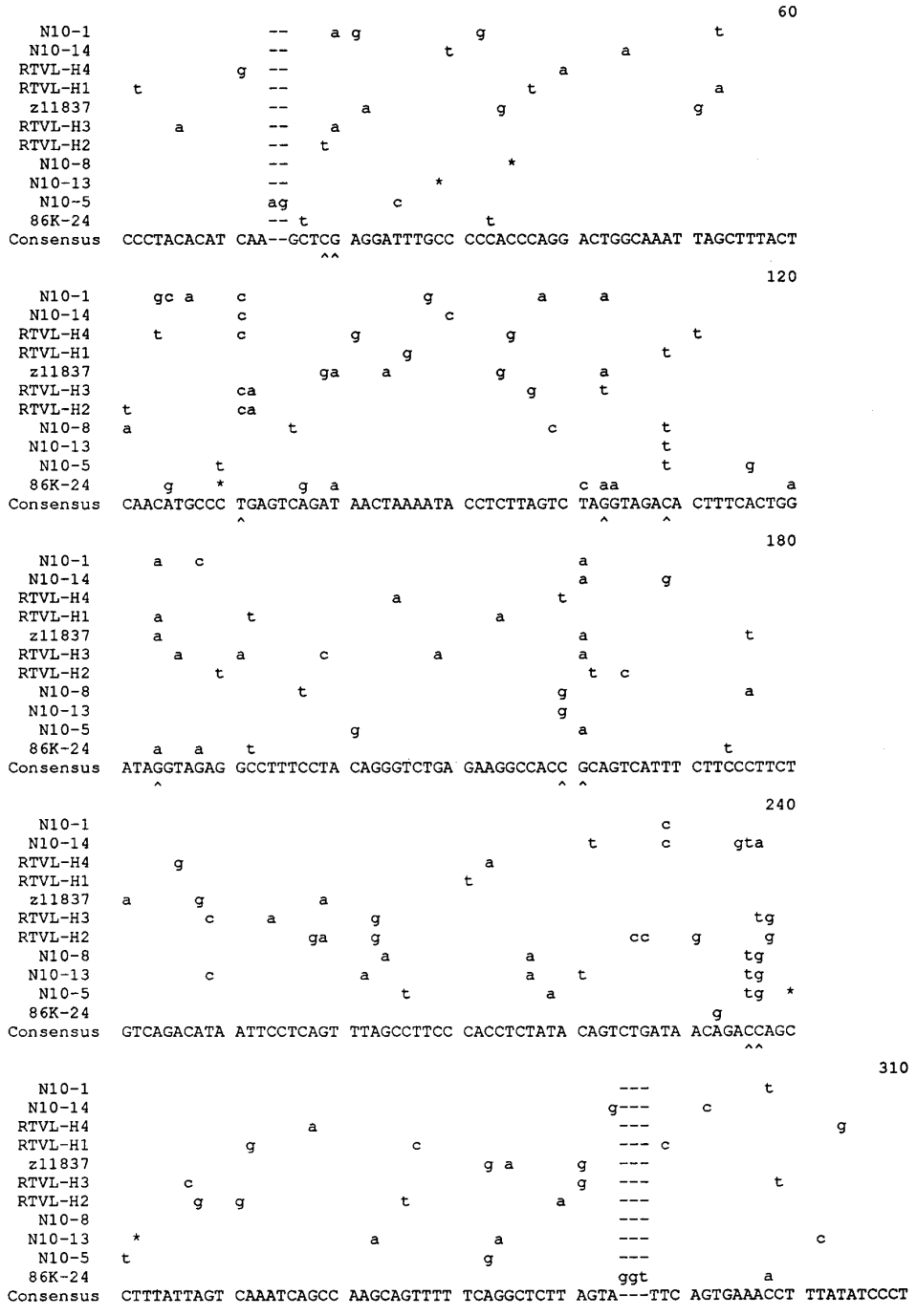

FIG. 6. Comparisons of HERV-H *pol* segments of the deleted subfamily. The consensus sequence is on the bottom line and bases shown in small letters are positions that differ from the consensus. Asterisks indicate single nucleotide deletions. Positions that were eliminated from the analysis are marked with a caret (^) under the consensus sequence.

Figure 6 is a comparison of the *pol* segments from the 11 deleted elements, showing where each element differs from the derived consensus sequence. The first observation to note is that there is little evidence for more restricted subfamilies within the 11 elements. That is, the nucleotide differences are distributed randomly and relatively few are shared among two or more sequences. This pattern suggests that each member was very close or identical to the progenitor element (represented by a consensus sequence) at the time of insertion and subsequently accumulated mutations at a neutral rate. The approximate age of this deleted subfamily can therefore be estimated by determining the average divergence from the consensus. To do this, the 10 positions marked with a caret (^) were first eliminated from the analysis either because they have no clear consensus (i.e., four

TABLE 1

HERV-H Ages Estimated by *pol* Sequence Divergence

Element	No. differences	Percentage divergence	Age (MYr)
N10-1	11	3.7	31
N10-14	9	3.1	26
RTVL-H4	11	3.7	31
RTVL-H1	10	3.4	28
z11837	14	4.7	39
RTVL-H3	14	4.7	39
RTVL-H2	16	5.4	45
N10-8	8	2.7	23
N10-13	9	3.1	26
N10-5	10	3.4	28
86K-24	15	5.1	43
Average	11.5 ± 2.7	3.9 ± 0.9	33 ± 8

or five sequences differed from the consensus) or because they could represent mutational "hotspots" due to the presence of CpG dinucleotides in the consensus. Such CpG positions mutate much faster than average and cannot be used to estimate evolutionary age (e.g., see Labuda and Striker, 1989). Of the remaining 295 positions, the differences from the consensus were counted with gaps and insertions each being treated as a single mismatch. The percentage divergence from the consensus was then determined for each element and is shown in Table 1. These values range from 2.7 to 5.4% with the average divergence being 3.9%. Using the neutral divergence rate of ~0.12% nucleotide differences/million years in humans (Li and Graur, 1991; Bailey *et al.,* 1991; Gibbons, 1995), the approximate ages of each element and the average age can be calculated and are also shown in Table 1. The average age for the subfamily determined by this method is 33 MYr, which by most estimates is before the time when hominoids diverged from Old World monkeys. Deviations of individual divergence values from the average could reflect different ages or could be partly due to differences in local mutation rates or to a slightly inaccurate consensus sequence caused by our limited sample size. Thus, the estimated ages of individual elements listed in Table 1 are most probably less accurate than the average value and should be treated with caution, particularly if there is no other measure of age available.

For two of the sequences, RTVL-H1 and RTVL-H3, we previously obtained an independent measure of age by performing PCR on genomic DNA from various primates using flanking primers (Goodchild *et al.,* 1993) (see Table 2). In that study, we found the RTVL-H1 element present in gibbon DNA and all higher primates, indicating that it has been in the germline for at least 17–20 million years which is the estimated time of divergence of gibbons from the higher apes. Its presence or absence in Old World monkeys could not be determined. The RTVL-H3 element was present in gibbons but absent in Old World monkeys, placing its age at 20–30 MYr. These direct measures of age agree fairly well with the average estimated age of this family. However, we cannot validate the ages of the other elements in a similar experimental manner since flanking sequences are either not available or are repetitive in those cases.

It is important to note that, except for the Z11837 sequence extracted from Genbank, the other 10 sequences used in the above analysis were isolated by hybridization with probes derived from RTVL-H1 or 2, suggesting potential ascertainment bias with this set of elements. It is possible that other more divergent HERV-H-related elements exist in the genome that would not be easily detected using RTVL-H1 or 2 probes and hence would be excluded from this analysis. However, when the same probes are used to estimate copy numbers in humans by genomic library screenings, values of close to 1000 copies are obtained (Mager and Henthorn 1984). Thus, we are confident that we have sampled from a large set of elements.

Relative ages of elements estimated by LTR–LTR divergence

Another way of estimating the age of an inserted retroviral element that does not rely on knowledge of the progenitor sequence is to compare the 5′ and 3′ LTRs. Assuming that the LTRs were identical at the time of insertion, the degree of divergence between them will reflect the relative age of the insertion. Table 2 shows the percentage differences between seven pairs of HERV-H LTRs and five additional LTR pairs from other human endogenous retroviral families for which independent data exist on evolutionary age. The first HERV-H pair listed, from the cH-4 element, are the most similar, having diverged by only 1.6%. These LTRs are of the type 1a subpopulation which is a small (~100 copy) HERV-H subfamily that has recently expanded in the hominoids (Goodchild *et al.,* 1993). Genomic PCR results have indicated that this element is present only in human, chimpanzee, and gorilla, making it the youngest HERV-H element that has been isolated (Goodchild *et al.,* 1993). The next four HERV-H elements listed, RTVL-H1, 2, 3, and 4, are all of the common deleted type as discussed above. It can be seen that the divergence values between these LTRs are very similar, being close to 4% in all four cases. This finding suggests that the ages of these four elements are much closer than indicated by the values in Table 1, which range from 28 to 45 MYr. This discrepancy suggests, as mentioned above, that the ages determined in Table 1 are most useful for calculating the average age of the subfamily, not for obtaining an accurate age of individual elements. Interestingly, the LTRs of the last two HERV-H elements listed in Table 2, RGH2 (Hirose *et al.,* 1993) and F22 (D. Mager, unpublished), show greater

TABLE 2

HERV LTR Divergence Compared to Evolutionary Age

Element	% LTR divergence	Independent measure of age		
		Age (MYr)	Evidence	Reference
cH-4	1.6	10–17	PCR	Goodchild *et al.,* 1993
RTVL-H1	3.9	>20	PCR	Goodchild et al., 1993
RTVL-H2	3.8	?		
RTVL-H3	4.3	20–30	PCR	Goodchild *et al.,* 1993
RTVL-H4	3.7	?		
RGH2	6.0	?		
F22	9.6	?		
RTVL-Ia	10.1	>30	Cloning	Erickson and Maeda, 1994
RTVL-Ib	14.3	>30	DNA analysis	Maeda and Kim, 1990
HERV-E (4.14)	5.8	>30	PCR	Shih *et al.,* 1991
HERV-R	8.8	>30	PCR	Shih *et al.,* 1991
HERV-K (C4)	9.1	>30	Cloning	Dangel *et al.,* 1995

sequence divergence. Both of these elements are of the more intact forms which, as a group, are older than the deleted elements.

Five LTR pairs from other HERV families are included in Table 2 for comparison. RTVL-Ia and b are members of the RTVL-I (or HERV-I) family found integrated into the primate haptoglobin gene cluster. Maeda and coworkers (Erickson and Maeda, 1994) have cloned RTVL-Ia from an Old World monkey and have used sequence comparisons of the locus to show that the RTVL-Ib element must be older (Maeda and Kim, 1990). Shih *et al.* (1991) used genomic PCR to demonstrate the presence of HERV-E (4.14 isolate) (Steele *et al.,* 1985), HERV-R (or ERV-3) (O'Connell and Cohen, 1984), and RTVL-Ia and b in Old World monkeys. Finally, the group of Yu (Dangel *et al.,* 1995) has used cloning and sequencing to show that the HERV-K (C4) element located in intron 9 of some complement component C4 genes is present in Old World monkeys. In all cases, the LTR divergence of these elements is more than HERV-H elements of the common deleted class, suggesting that they are older than the deleted HERV-H subfamily.

DISCUSSION

Figure 7 illustrates the current knowledge regarding the evolutionary history of HERV-H elements. Two primary expansions of HERV-H sequences have been identified to date. The largest expansion involved a structurally defective subfamily discussed here that appears to have amplified to several hundred copies in a relatively narrow window of time prior to the divergence of hominoids from Old World monkeys. In a previous study, we reported a second more recent expansion that was restricted to the hominoids and involved amplification to ~100 copies of a distinct subfamily of deleted elements associated with a subtype of LTR termed type 1a (Goodchild *et al.,* 1993). In that study, no evidence of HERV-H existence was found in New World monkeys by examining integration times of 11 individual elements. However, the findings reported here indicate that this family of sequences is indeed present in New World primates. The successive expansions of HERV-H in Old World primates are reminiscent of amplifications of different subfamilies of Alu and LINE-1 elements that have occurred throughout evolution (Deininger *et al.,* 1992; Smit *et al.,* 1995). As with these sequences, it is likely that analysis of a larger set of HERV-H elements from humans and other primates will uncover other subfamilies that may have been active at different times in evolution.

We have shown here that HERV-H elements are present in low numbers in New World monkeys, suggesting that they existed in the germline as recognizable entities for over 40 million years. However, since we have not yet shown that an element in New World monkeys occupies the same genomic site in an Old World primate, it is formally possible that separate germline infections of HERV-H as an exogenous virus occurred in the New and

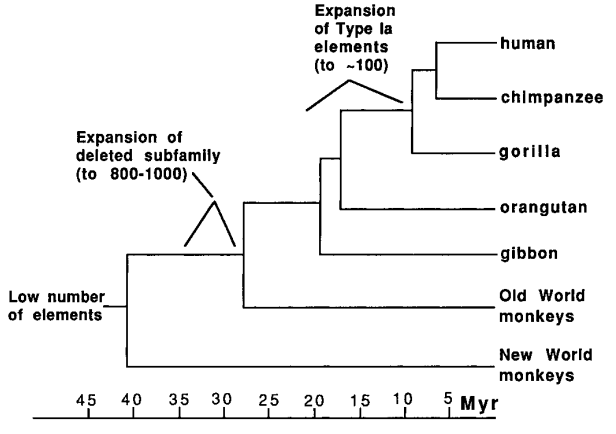

FIG. 7. Expansions of HERV-H elements during primate evolution. The branchpoint times shown should be considered approximate.

Old World branches shortly after their split. Such a chain of events is considerably more complex, however, and we consider it unlikely. The age of the HERV-H family determined here is a minimum estimate since its possible presence in more divergent species such as the prosimians has not yet been investigated. The age of the putative primordial HERV-H element therefore remains unknown as does its evolutionary origins. Structural similarities to exogenous retroviruses have led to the generally held belief that endogenous retroviruses are remnants of exogenous infections of the germline that became fixed in the species (e.g., see Coffin, 1982). This is most certainly true for many of the endogenous viruses of mice and chickens that are closely related to infectious retroviruses of those species (Coffin, 1982). If HERV-H elements originated in this way, it may be possible to identify the point during evolution at which the elements first appeared in the germline. However, it is noteworthy that the human endogenous retroviruses that have been isolated are generally more similar to rodent viruses than to the human infectious retroviruses known today (see Wilkinson *et al.,* 1994). Since retroviruses are related to retrotransposons, it is also possible that some HERV may have arisen from genomic retrotransposons. In this case, it would be impossible to determine the age of a particular family since it would simply have evolved along with the genome. A third possibility for the origin of some HERV families is that they are recombinational products of genomic retrotransposons and exogenous retroviruses. Indeed, some HERV families such as HERV-K (Callahan *et al.,* 1985) and HERV-H (Mager and Freeman, 1987) show regions of homology to different classes of retroviruses, suggesting that recombination has played a role in their history. Structural complexities such as this would make examinations of HERV molecular evolution even more difficult. Nonetheless, it would be of interest to attempt to trace the origins of HERV-H and other primate endogenous retroviral families further back in time. For instance, it would be interesting to determine their ages relative to the LINE-1 retrotransposon family, for which it has been estimated that over half of the human copies inserted into the genome before mammalian radiation (Smit *et al.,* 1995).

Although HERV-H family progenitors can be traced back to before the New World/Old World split, most copies in the Old World branch arose subsequent to that split from a partially deleted precursor element. We have previously obtained evidence for the RNA-mediated retrotransposition of HERV-H sequences through the detection of spliced integrated forms (Goodchild *et al.,* 1995). It is therefore possible that the major expansion of HERV-H elements involved spread as defective retrotransposons. However, the possibility of multiple germline infections by virions containing the defective HERV-H genome cannot be ruled out. One possible scenario for the expansion is that a structurally defective precursor element containing the five common deletions became the favored substrate for retrotransposition over the more intact forms. However, the reason for this is not clear. One explanation is that mutations may have occurred in the LTR which allowed the precursor "master" element to be transcribed at much higher levels compared to intact forms. It appears that such an event may have been responsible for the more recent hominoid-specific expansion of HERV-H elements with type 1a LTRs (Goodchild *et al.,* 1993). In the case of the deleted precursor element, such LTR mutations could have occurred after the New World/Old World divergence which would account for the lack of expansion in the New World branch. There are a number of nucleotide differences between the consensus LTR sequence of the deleted subfamily compared to the LTRs of the more intact elements RGH1 and 2 (unpublished observations) but it is unknown whether any are of functional relevance. To properly address this question, more LTR sequences of both the intact and the deleted types from Old World and New World primates need to be analyzed for structural and functional differences.

Regardless of what caused the amplification of the deleted translationally defective HERV-H subfamily, the retrotransposition of these elements must have been facilitated by proteins provided *in trans,* possibly encoded by rare intact HERV-H elements. Putative coding-competent HERV-H elements have not yet been isolated and their identification is hampered by the vast number of defective members in the human genome. The finding that New World monkeys contain relatively few deleted sequences suggests that these species are good candidates in which to conduct a search for protein-encoding elements. The existence of vastly different copy numbers of HERV-H elements in New and Old World primate species is also of interest as a distinguishing characteristic. Although difficult to approach experimentally, it has been speculated that transposable elements may be involved in speciation (e.g., see Li and Graur, 1991). Whether endogenous retroviral or other dispersed repetitive elements were a significant factor in the speciation of primates is an intriguing but unanswered question.

ACKNOWLEDGMENTS

We thank Christine Kelly for help with manuscript preparation. This work was supported by a grant from the Medical Research Council of Canada with core support provided by the British Columbia Cancer Agency.

REFERENCES

Bailey, W. J., Fitch, D. H. A., Tagle, D. A., Czelusniak, J., Slightom, J. L., and Goodman, M. (1991). Molecular evolution of the $\psi\eta$-globin gene locus: Gibbon phylogeny and the hominoid slowdown. *Mol. Biol. Evol.* 8, 155–184.

Benveniste, R., and Todaro, G. (1974). Evolution of type C viral genes.

I. Nucleic acids from baboon type C virus as a measure of divergence among primate species. *Proc. Natl. Acad. Sci. USA* 71, 4513–4518.

Callahan, R., Chiu, I-M., Wong, J. F. H., Tronick, S. R., Roe, B. A., Aaronson, S. A., and Scholm, J. (1985). A new class of endogenous human retroviral genomes. *Science* 228, 1208–1211.

Coffin, J. (1982). Endogenous viruses. *In* "Molecular Biology of Tumor Viruses: RNA Tumor Viruses" (R. A. Weiss *et al.,* Eds.), 2nd ed., pp. 1109–1203. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Dangel, A. W., Baker, B. J., Mendoza, A. R., and Yu, C. Y. (1995). Complement component C4 gene intron 9 as a phylogenetic marker for primates: Long terminal repeats of the endogenous retrovirus ERV-K (C4) are a molecular clock of evolution. *Immunogenetics* 42, 41–52.

Dangel, A. W., Mendoza, A. R., Baker, B. J., Daniel, C. M., Carroll, M. C., Wu, L. C., and Yu, C. Y. (1994). The dichotomous size variation of human complement C4 genes is mediated by a novel family of endogenous retroviruses, which also establishes species-specific genomic patterns among Old World primates. *Immunogenetics* 40, 425–436.

Deininger, P. L., Batzer, M. A., Hutchison, C. A., and Edgell, M. H. (1992). Master genes in mammalian repetitive DNA amplification. *Trends Genet.* 8, 307–311.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12, 387–395.

Erickson, L. M., and Maeda, N. (1994). Parallel evolutionary events in the haptoglobin gene clusters of rhesus monkey and human. *Genomics* 22, 579–589.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132, 6–13.

Fraser, C., Humphries, R., and Mager, D. (1988). Chromosomal distribution of the RTVL-H family of human endogenous retrovirus-like sequences. *Genomics* 2, 280–287.

Gibbons, A. (1995). When it comes to evolution, humans are in the slow class. *Science* 267, 1907–1908.

Gingerich, P. D. (1984). Primate evolution: Evidence from the fossil record, comparative morphology, and molecular biology. *Year Book Phys. Anthropol.* 27, 57–72.

Goodchild, N., Wilkinson, D., and Mager, D. (1993). Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements. *Virology* 196, 778–788.

Goodchild, N. L., Freeman, J. D., and Mager, D. L. (1995). Spliced HERV-H endogenous retroviral sequences in human genomic DNA: Evidence for amplification via retrotransposition. *Virology* 206, 164–173.

Haltmeier, M., Seifarth, W., Blusch, J., Erfle, V., Hehlmann, R., and Leib-Mösch, C. (1995). Identification of S71-related human endogenous retroviral sequences with full-length *pol* genes. *Virology* 209, 550–560.

Hirose, Y., Takamatsu, M., and Harada, F. (1993). Presence of *env* genes in members of the RTVL-H family of human endogenous retrovirus-like elements. *Virology* 192, 52–61.

Hutchison, C. A., Hardies, S. C., Loeb, D. D., Shehee, W. R., and Edgell, M. H. (1989). LINES and related retroposons: Long interspersed repeated sequences in the eucaryotic genome. *In* "Mobile DNA" (D. E. Berg and M. M. Howe, Eds). Am. Soc. Microbiol., Washington, DC.

Kannan, P., Buettner, R., Pratt, D., and Tainsley, M. (1991). Identification of a retinoic acid-inducible endogenous retroviral transcript in the human teratocarcinoma-derived cell line PA-1. *J. Virol.* 65, 6343–6348.

Koop, B. F., Goodman, M., Xu, P., Chan, K., and Slightom, J. (1986). Primate η-globin DNA sequences and man's place among the great apes. *Nature* 319, 234–238.

Labuda, D., and Striker, G. (1989). Sequence conservation in Alu evolution. *Nucleic Acids Res.* 17, 2477–2491.

Larsson, E., Kato, N., and Cohen, M. (1989). Human endogenous proviruses. *Curr. Top. Microbiol. Immunol.* 148, 115–132.

Leib-Mösch, C., Bachmann, M., Brack-Werner, R., Werner, T., Erfle, V., and Hehlmann, R. (1992). Expression and biological significance of human endogenous retroviral sequences. *Leukemia* 6, 72S–75S.

Li, W-H., and Graur, D. (1991). Evolution by transposition. *In* "Fundamentals of Molecular Evolution," pp. 172–203. Sinauer, Sunderland, MA.

Maeda, N., and Kim, H-S. (1990). Three independent insertions of retrovirus-like sequences in the haptoglobin gene cluster of primates. *Genomics* 8, 671–683.

Mager, D., and Goodchild, N. (1989). Homologous recombination between the LTRs of a human retrovirus-like element causes a 5-kb deletion in two siblings. *Am. J. Hum. Genet.* 45, 848–854.

Mager, D. L. (1989). Polyadenylation function and sequence variability of the long terminal repeats of the human endogenous retrovirus-like family RTVL-H. *Virology* 173, 591–599.

Mager, D. L., and Freeman, D. (1987). Human endogenous retroviruslike 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 Henthorn, P. S. (1984). Identification of a retrovirus-like repetitive element in human DNA. *Proc. Natl. Acad. Sci. USA* 81, 7510–7514.

Mariani-Costantini, R., Horn, T., and Callahan, R. (1989). Ancestry of a human endogenous retrovirus family. *J. Virol.* 63, 4982–4985.

Martin, M., Bryan, T., Rasheed, S., and Khan, A. (1981). Identification and cloning of endogenous retroviral sequences present in human DNA. *Proc. Natl. Acad. Sci. USA.* 78, 4892–4896.

O'Connell, C. D., and Cohen, M. (1984). The long terminal repeats of a novel human endogenous retrovirus. *Science* 226, 1204–1206.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shih, A., Coutavas, E. E., and Rush, M. G. (1991). Evolutionary implications of primate endogenous retroviruses. *Virology* 182, 495–502.

Sibley, C. G., and Ahlquist, J. E. (1987). DNA hybridization evidence of hominoid phylogeny: Results from an expanded data set. *J. Mol. Evol.* 26, 99–121.

Smit, A. F. A., Toth, G., Riggs, A. D., and Jurka, J. (1995). Ancestral mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.* 246, 401–417.

Steele, P. E., Rabson, A. B., Bryan, T., and Martin, M. A. (1985). Distinctive termini characterize two families of human endogenous retroviral sequences. *Science* 225, 943–947.

Tabor, S., and Richardson, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 polymerase. *Proc. Natl. Acad. Sci. USA* 84, 4767–4771.

Wilkinson, D. A., and Mager, D. L. (1993). RTVL-H human endogenous retrovirus-like elements with an intact *pol* and an *env*-related region. *J. Cancer Res. Clin. Oncol.* 199(Suppl. 1), S8.

Wilkinson, D. A., Freeman, J. D., Goodchild, N. L., Kelleher, C. A., and Mager, D. L. (1990). Autonomous expression of RTVL-H endogenous retroviruslike elements in human cells. *J. Virol.* 64, 2157–2167.

Wilkinson, D. A., Goodchild, N. L., Saxton, T. M., Wood, S., and Mager, D. L. (1993). Evidence for a functional subclass of the RTVL-H family of human endogenous retrovirus-like sequences. *J. Virol.* 67, 2981–2989.

Wilkinson, D. A., Mager, D. L., and Leong, J. C. (1994). Endogenous human retroviruses. *In* "The Retroviridae, Vol. 3" (J. Levy, Ed.), pp. 465–535. Plenum, NY.

Interestingly, they remained in low numbers in the New World branch while a subfamily underwent a major amplification in Old World primates before the time of divergence of hominoids from Old World monkeys.Â @article{Mager1995HERVHER, title={HERV-H endogenous retroviruses: presence in the New World branch but amplification in the Old World primate lineage.}, author={Dixie L. Mager and J. Douglas Freeman}, journal={Virology}, year={1995}, volume={213 2}, pages={. 395-404 } }. Dixie L. Mager, J. Douglas Freeman. Published in Virology 1995. Biology, Medicine. The evolutionary origin and age of the HERV-H family of human endogenous retrovirus-like sequences was investigated in this study. stated in. Europe PubMed Central. PubMed ID.Â HERV-H endogenous retroviruses: presence in the New World branch but amplification in the Old World primate lineage. (English). 1 reference. stated in. Endogenous retroviruses (ERVs) are endogenous viral elements in the genome that closely resemble and can be derived from retroviruses. They are abundant in the genomes of jawed vertebrates, and they comprise up to 5â€“8% of the human genome (lower estimates of ~1%). ERVs are a subclass of a type of gene called a transposon, which can be packaged and moved within the genome to serve a vital role in gene expression and in regulation. They are distinguished as retrotransposons, which are Class I elements