Identification of a novel family of human endogenous retroviruses and characterization of one family member, *HERV-K(C4)*, located in the complement *C4* gene cluster

Mayada Tassabehji¹, Tom Strachan^{1,2}, Michael Anderson^{3,+}, R.Duncan Campbell³, Simon Collier^{1,§} and Majlinda Lako^{2,*}

¹Department of Medical Genetics, University of Manchester, St Mary's Hospital, Manchester M13 0JH, ²Department of Human Genetics, University of Newcastle upon Tyne, Ridley Building, Claremont Place, Newcastle upon Tyne NE1 7RU and ³MRC Immunochemistry Unit, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Received September 27, 1994; Revised and Accepted November 7, 1994

EMBL accession no. X80240

ABSTRACT

We have identified a novel family of about 10–50 human endogenous retrovirus elements (*HERVs*) and have characterized one family member (*HERV-KC4*). This retrovirus element is integrated within intron 9 of the complement *C4A* genes and also in some *C4B* genes, and is a principal contributor to interlocus and interallelic length heterogeneity of *C4* genes. The *HERV-K(C4)* sequence has a typical retrovirus structure with elements of *gag*, *pol* and *env* domains, flanked by two long terminal repeats (LTRs) and is similar to type A, B and D retroviruses. Multiple termination codons preclude the existence of long open reading frames, suggesting that the *HERV-K(C4)* sequence is no longer functional. Zoo blot hybridization reveals that New World monkeys appear to lack sequences similar to *HERV-K(C4)*, suggesting that integration has occurred after the divergence of Old and New World monkeys. Retrotransposition of prototype viruses is presumed to have led to the amplification and integration of the members of the family in different loci, which in humans, appear to be dispersed over several chromosomes. The absence of the *HERV-K(C4)* element in some *C4B* genes in both humans and orang-utans indicate that the retrovirus inserted into the *C4A* gene after the duplication of the cluster. Subsequent spread of the *HERV-K(C4)* sequence to *C4B* genes presumably occurred by interlocus sequence exchange mechanisms, such as unequal crossover and gene conversion-like mechanisms.

INTRODUCTION

The retroviral superfamily is composed of retroviruses and retrovirus-like elements (*RTVL*) and is marked by the capacity for transposition through an RNA intermediate via the action of reverse transcriptase. Retroviruses typically contain flanking long terminal repeats (LTRs) and a central element which contains various genes, notably *gag* (which encodes capsid proteins), *env* (which encodes envelope proteins) and *pol*, a gene which encodes a multifunctional protein with various enzyme activities, including reverse transcriptase (RT) activity and endonuclease and integrase (transposase) functions. While retroviruses are able to produce infectious viruses, retrotransposons (which lack the *env* gene) and retroposons (which lack both the *env* gene and also the LTRs of retrotransposons) do not pass through the infectious form and their life cycle is limited within the cell.

Transcription of the intermediate RNA starts at one LTR and ends at the other one. Reverse transcription requires a primer to initiate strand synthesis, which in retroviruses is usually a tRNA carrying lysine, phenylalanine, histidine, isoleucine or proline anticodons, having the last bases complementary to the retrovirus sequence (1,2). The integration of newly synthesized double-stranded DNA at new loci may cause gene disruption, gene activation or translocation. Furthermore, the incorporation of a cellular gene may occur with subsequent transfection of other cells (3). These findings indicate that retroviruses are major contributors to insertional mutagenesis and tumourigenesis.

Over the last two decades, at least 7 families of *HERVs* (human endogenous retrovirus) and *RTVL* elements have been identified in humans (1,4–11). According to their sequence homology to known viruses, they are usually divided into two groups: *HERV-I*

*To whom correspondence should be addressed

Present addresses: ⁺University of Manchester, Oxford Road, Manchester M13 9PL and [§]University of Cambridge, Department of Genetics, Downing Street, Cambridge CB2 3EH, UK

(which includes *HERV*s similar to type C mammalian viruses) and *HERV-II* (which includes *HERV*s similar to type A, B and D mammalian viruses and to type C avian viruses) (12). Some of the cloned members of these families present full length proviral DNA, some are truncated either in the 5' or 3' end and some exist as solitary LTRs (4,6,8). In most of them, multiple termination codons interrupt reading frames, precluding the production of functional proteins necessary for retrotransposition (6–8). However, the dispersion and integration at new loci of defective retroviruses through *trans*-complementation has already been demonstrated (13–15). Members of the same family are located on the same chromosome or dispersed randomly, revealing that the integration process depends not only on the characteristics of retroviruses, but also on the chromatin structure of host chromosomes (16). There is already evidence, such as differences in copy number and the tRNA primers used for the initiation of (–) strand synthesis, which indicates that different families of *HERV*s derive from different exogenous viruses.

In the present report, we present evidence for a new family of human endogenous retroviruses, which can be classified within the *HERV-II* group; this family is dispersed over several human chromosomes, and appears to be conserved in Old World monkeys, but not in New World monkeys. We also report the complete nucleotide sequence and sequence analysis of a 6,357 bp family member which is located in intron 9 of the complement *C4A6* allele and which has LTRs and sequences similar to *gag*, *env* and *pol* genes. Although this family member is presumed to have integrated initially in the *C4A* gene, interlocus sequence exchanges between the duplicated *C4A* and *C4B* genes have resulted in its transfer to a considerable proportion of *C4B* genes, contributing to the well-characterized interlocus and interallelic length variation of *C4* genes.

MATERIALS AND METHODS

DNA sequencing

Two neighbouring 4.8 kb *Bam*HI fragments from the cosmid clone cos210 (17) containing the complement *C4A6* allele were subcloned into pAT-X to generate B5' 4.8-3 and B3' 4.8-3 (see Figure 1). The 4.8 kb inserts were individually digested with one or two restriction endonucleases in a series of independent reactions to generate a series of overlapping restriction fragments which were subcloned into pTZ18/19R (Pharmacia) and

sequenced in both directions by using standard dideoxynucleotide sequencing methods and vector-specific or insert-specific primers.

Southern blotting

High molecular weight primate genomic DNA samples (Scotlab Ltd) were digested with restriction enzymes according to the manufacturer's instructions. 5 µg of each sample was subjected to electrophoresis on 0.8% agarose gels and fractionated, restriction fragments were transferred to Hybond N+ membrane (Amersham International plc) in 20×SSC overnight and immobilized using 0.4 M NaOH. The primate blot and a somatic hybrid panel blot (Scotlab Ltd) were hybridized with an [α - 32 P]-labelled subclone, env-1.1, a 1.1 kb *Eco*RI–*Bam*HI restriction fragment derived from the *env* gene of *HERV-K(C4)* (see text). Hybridization washes were concluded in 20×SSC and 0.1% SDS or 1×SSC and 0.1% SDS.

Computer analysis

Most of the computer analysis was done by using the GCG (Wisconsin computer package) program (gap weight used in gap comparisons is 3.00 and gap length 0.10, unless otherwise indicated). FASTA and BLAST searches were conducted against available sequences in different databases (July 1994) (18,19).

RESULTS

The additional 6.5 kb element found in *C4* genes is a defective retrovirus

Short- and long-range restriction mapping have previously revealed the variable presence of a c. 6.5 kb element in complement *C4* genes in the class III region of the HLA complex. This region contains a tandemly duplicated segment that is c. 30 kb long, resulting in duplication not only of the *C4* genes, but also of steroid 21-hydroxylase genes (*CYP21*) and also of genes of unknown function (20,21). Those complement *C4* genes that possess the variable 6.5 kb element have traditionally been described as long *C4* genes and include *C4A* genes and ~2/3 of *C4B* genes (22–27). The remaining *C4B* alleles which lack the 6.5 kb element are traditionally referred to as short *C4* genes.

In order to study this 6.5 kb element, we subcloned and sequenced appropriate restriction fragments from the *C4A6* allele (see Materials and Methods; Figure 1). Computer analysis revealed that the additional element has a retrovirus-like structure

Figure 1. Partial restriction map of the *C4A6* allele. Restriction sites are as follows: B, *Bam*HI; Bg, *Bgl*II; C, *Cla*I; H, *Hind*III; R1, *Eco*RI; K, *Kpn*I; S, *Sal*I; St, *Sst*I; X, *Xho*I.

with sequences similar to *gag*, *pol* and *env* genes, and is flanked by two LTRs. Including the two LTRs, the full length element is 6,357 bp long. Comparison of the sequence with the reverse complement of the *C4A* gene sequence enabled identification of a common 6 bp sequence (TGTCTG) which marked the target integration site. An 18 bp sequence (TGGCGCCCAA-CGTGGGGC) with 12 bases complementary to the last 18 bases of the 3' end of tRNA^{Lys}, but only 8 bases complementary to tRNA^{Trp} and tRNA^{Pro}, is classified as the primer binding site. Accordingly, tRNA^{Lys}, carrying a CUU anticodon, is considered to be the primer for (–) strand synthesis. We therefore propose that the retrovirus which integrated in *C4* genes should be named *HERV-K(C4)*, where K stands for lysine, in accordance with the terminology used for other *HERVs* (1,2,4,9). *HERV-K(C4)* is located within the distal half of the intron and the transcriptional orientation of *HERV-K(C4)* is opposite to that of the *C4* gene (Figure 2). The presence of multiple stop codons in all the reading frames suggests that *HERV-K(C4)* is no longer transcriptionally active.

The structure of *HERV-K(C4)*

The analysis of the *HERV-K(C4)* sequence reveals that the two long terminal repeats at the 5' and 3' end, which are 546 bp and 548 bp respectively, differ by a few deletions and nucleotide substitutions, contributing to an overall divergence level of 7.8%. At the ends of the LTRs, there is a 7 bp inverted repeat TGTGGG/CCCAACA which encompasses at the distal ends an inverted dinucleotide repeat TG/CA that is also found in *HERV-K10* and other proviruses (2,28,29). As in other retrovirus LTRs, transcription activation sequence elements are found in *HERV-K(C4)*, such as a TATA box sequence (TATAAA; 29 bp downstream of the starting TG sequence) and an 8 bp sequence (GTGGAAAG) corresponding to the consensus sequence for the SV40-type enhancer. Usually the SV40 enhancer core is followed by two tandem repeats, 72 and 73 bp long successively, or short repeats, which are GC-rich and preceded by a glucocorticoid response element (9,28). However, none of these elements are found in either LTRs of *HERV-K(C4)*. The CAT box which is usually located 40–50 bp upstream of the TATA box is also missing.

A typical polyadenylation signal AATAAA is found in the 5' LTR, and a similar sequence is also found in the 3' LTR. However, comparison with sequences existing in different databases reveal that the LTRs of *HERV-K(C4)* do not show extensive homology throughout their length to LTRs of other retroviruses, although some homology is observed around the end of LTRs, where polyadenylation signals are located. In spite

Figure 2. The location and orientation of *HERV-K(C4)* in the *C4A6* allele.

of the fact that *HERV-K(C4)* would be expected to be defective in expression, the LTRs have all the elements necessary for initiation, polyadenylation and termination of the transcription process.

Sequence comparisons have revealed that the closest homologue to *HERV-K(C4)* is *HERV-K10* [*HERV-K10(+)*] (9): the overall similarity calculated using a gap weight of 5 and a gap length weight of 0.3 (see Materials and Methods) is 40.17%. The latter virus, however, has LTRs which are 968 bp in length, considerably longer than those of *HERV-K(C4)*, and is also quite divergent in sequence (43% similarity). Dot matrix and gap comparison analysis between the nucleotide sequence of the two retroviruses allowed the definition in the *HERV-K(C4)* sequence of regions corresponding to *gag*, *pol* and *env* genes (Figure 3). The highest homology is observed in the *pol* (overall homology 64.5%) and *env* (overall homology 60.3%) regions. As in most other retroviruses, the *gag* region of *HERV-K(C4)* starts with a sequence specifying the dipeptide Met–Gly (9). However, multiple stop codons, present throughout the length of *HERV-K(C4)*, preclude the production of functional proteins. Comparison of amino acid sequences of the three reading frames in the *gag*, *pol* and *env* regions revealed short stretches of amino acids which are well conserved between the two retroviruses and probably represent functional domains of the proteins. The RT

Figure 3. Dot matrix analysis between *HERV-K10(+)* and *HERV-K(C4)*. The window size is 21 and stringency is 14.

Figure 4. Partial restriction map of *HERV-K(C4)*. Restriction sites are as follows: B, *Bam*HI; Bg, *Bgl*II; K, *Kpn*I; R1, *Eco*RI; T, *Taq*I. The probe region (*env*-1.1), the *Eco*RI, *Bam*HI and the *Taq*I fragments overlapping with the probe are shown. The interrupted lines indicates parts of restriction fragments which extend out from the *HERV-K(C4)* sequence.

Figure 5. Southern blotting of human genomic DNA samples with *env*-1.1 probe. Lanes are as follows: 1, *Taq*I digest; 2, *Eco*RI digest; 3, *Bam*HI digest. The main hybridizing bands are indicated.

shows up to 70% nucleotide sequence homology to the RT sequences of *HERV-K10*, *HML-3* (human mouse mammary tumour virus-like sequences), *HML-5* and *HML-6* (8; data not shown). The amino acid comparison revealed also that the 7 domains of the RT protein which are reported to be very well conserved between RNA polymerase II of different species and the RT of retroviruses are dispersed in three reading frames of the RT region (30).

FASTA searches for individual *gag*, *pol* and *env* regions showed that they are similar to the reciprocal regions of type A, B and D mammalian viruses in particular, consistent with *HERV-K(C4)* belonging to the same group *HERV-II* (data not shown). The same searches also revealed sequences similar to separate regions as LTRs, *gag*, *pol*, *env* or LTRs and part of *gag* and LTRs and part of *env*, implying the presence of similar truncated elements located elsewhere in the genome. Very often these sequences are located in the vicinity of Alu repeats, implying a preference for insertion in early replicating chromosome domains.

Figure 6. Southern blotting of somatic cell hybrid panel with *env*-1.1 probe. The restriction enzyme used for digestion is *Eco*RI. The main hybridizing bands, 3.8 kb, are indicated on the left hand side of the figure.

HERV-K(C4) belongs to a novel family of *HERVs*

To identify sequences related to *HERV-K(C4)*, Southern blot hybridizations were conducted against genomic DNA samples. *env*-1.1, a 1.1 kb *Eco*RI-*Bam*HI fragment (see Figures 1 and 4), which spans most of the *env* region of *HERV-K(C4)* detected a variety of bands when hybridized to *Eco*RI-, *Bam*HI- and *Taq*I-digested human genomic DNA samples (Figure 5). As expected from the known restriction sites (Figure 5), the main hybridizing bands were as follows: *Eco*RI, 3.8 kb; *Bam*HI, 4.8 kb; *Taq*I, 2.1 kb and 6.0 kb plus 7.0 kb (when both *C4A* and *C4B* genes are long; the 6.0 kb band is *C4B*-specific, while the 7.0 kb band is *C4A*-specific). In addition to the main hybridizing bands, several fainter bands were detected when less stringent conditions were applied (Figure 5).

The uniformity of the pattern in different human individuals shows that almost no polymorphism exists, except that which is due to the presence or absence of the *HERV-K(C4)* element (as seen in the *Taq*I bands of *C4B*). We estimate that there are about 10–50 sequences related to *HERV-K(C4)*. However, interpretation of the significance of the fainter bands is not

	010	016•	212	324	423	683	734	750	756	803	811	852	867	909	937	940	1006	1049	1079	1099
1						*							+		+					+
2												+								
3					+														+	
4										*							+			
5			+D			+D	+	+	+D	+			+	+D	+	+	+	+	+	+D
6									+					+						
7									+								+			
8										+	+			+						
9							+													
10	+																			
11																		+		
12						+			*											
13								+	+				+				+			+
14						+		+	+				+	+	+					
15								+							*		+			
16		+																		
17											+				*					
18				+							+		+							
19						+		+	+				+				+			+
20									+							+				
21						+			+						+		+			+
22						+				+										+
X										+				+						
Y			+						+								*			

Figure 7. Chromosomal distribution in the somatic cell hybrid panel blot. ● Human/mouse somatic cell hybrid, all others are human/hamster hybrids; D, multiple deletions; + >30% of the cells contain the given chromosome; * 5–30% of the cells contain the given chromosome.

straightforward. First, the related sequences could be expected to show different degrees of sequence homology to *HERV-K(C4)*. Secondly, the number of loci represented by each hybridization band could be different because of co-migration of non-allelic restriction fragments. Moreover, a single locus can contribute two hybridizing bands, if the cross-hybridizing sequence contains an internal restriction site for the enzyme used to prepare the relevant genomic blot. As the *TaqI* digest revealed different intensities of 7.0 and 6.4 kb bands characteristic of *C4A* and recombinant *C4A/C4B* locus genes, respectively, in comparison with the 2.1 kb *TaqI* band, specific to *HERV-K(C4)*, one would expect the existence of very closely related copies to *HERV-K(C4)* in the family, conserving *TaqI* restriction sites.

To investigate the chromosomal location of *HERV-K(C4)*-related sequences in the human genome, the env-1.1 kb probe was hybridized to a somatic cell hybrid panel containing *EcoRI*-digested genomic DNA (Figure 6 and Figure 7). The results clearly show that a prominent 3.8 kb *EcoRI* hybridizing fragment is found in lanes containing human chromosome 6, as expected, but also in a variety of other human chromosomes.

To test the evolutionary conservation of *HERV-K(C4)*, the env-1.1 kb probe was hybridized to a primate genomic blot (Figure 8). An almost uniform pattern is observed within the chimpanzee cluster and within the orang-utang cluster, revealing no polymorphism. Also, unique hybridizing patterns are seen in gorilla, macaque and human. Besides the main hybridizing bands, several fainter bands are observed in each of them. The main *EcoRI* hybridization band in the orang-utang cluster and human is 3.8 kb, which corresponds exactly to that predicted from the restriction map of *HERV-K(C4)* and indicates its likely presence within the *C4* genes of these species. In contrast, the chimpanzee appears to show other cross-hybridizing bands, suggesting the presence of closely related members in this species (Figure 8). The chimpanzee, however, appears only to have short *C4* genes

Figure 8. Hybridization of the env-1.1 probe to *EcoRI*-digested genomic DNA samples from primates. Lanes represent individual genomic DNA samples as follows: 1–5, chimpanzee (*Pan troglodytes*); 6–8, orang-utang (*Pongo pygmaeus*); 9, gorilla (*Gorilla gorilla*); 10, macaque (*Macaca fascicularis*); 11, marmoset (*Leontopithecus saguinus*); 12, human; 13, marmoset (*Cebuella pygmaea*).

(33; M.Lako, unpublished observations), which indicates the absence of any *HERV-K(C4)* sequence within its *C4* gene. As a result, sequences closely related to *HERV-K(C4)* appear to be integrated in the chimpanzee genome at locations outside the *C4* gene.

No hybridizing bands were detected in marmoset DNA samples (*Leontopithecus saguinus* and *Cebuella pygmaea*) (Figure 8), a

finding which suggests that the integration of these elements may have occurred after the New World monkeys diverged from the Old World monkeys. The multiple faint bands in the primate blot may reflect the existence of 10–50 copies of less related or truncated sequences in primate genomes, a number comparable to that expected in the human genome.

DISCUSSION

The human complement *C4A* and *C4B* genes are known to reside on duplicated segments of DNA about 30 kb long, containing *CYP21* genes and also other genes of uncertain function (20,21). Length variation for individual *C4* genes has been recorded: *C4A* genes are 'long' (22.5 kb), whereas *C4B* genes may be 22.5 kb ('long' *C4*) or 16 kb ('short' *C4*) (22–25). The size difference between long and short *C4* genes has been demonstrated to be due to an additional 6.5 kb found in *C4* long genes (26). A subsequent study has located this additional sequence, 2–2.5 kb downstream of the transcription point of the *C4A* gene (27). We have shown in the present study that this 6.5 kb sequence is a member of a novel family of human endogenous retroviruses, which integrated in intron 9 of the *C4A* gene.

Studies on integration site preferences for elements that undergo retrotransposition have revealed that host genomes contain a limited number of integration sites that are used as targets at very high frequency, as well as a number of sites where integration occurs spontaneously (31). Alu repeats have been found inside the LTRs of *RTVL* elements or in the vicinity of *HERVs*, indicating some preferential ability for integration in early-replicating chromatin domains (4). However, although the chromatin structure of early-replicating domains favours the integration process, it does not appear to override the retrovirus's capacity for transcription. Localization of *RTVL* sequences in the vicinity of each other or in the same chromosomes also implies that some spatial factors must also play a role in integration site selection (16). In addition to *HERV-K(C4)*, other *RTVL* elements integrated in the HLA complex include the *SINE-R*. *C2* in intron 3 of the *C2* gene (11) and LTRs in the *HLA-DQ* locus (4). Additionally, a deleted provirus which is 6 kb long and is similar to type C viruses is known to be inserted 2 kb upstream of the transcription point of a murine complement C4 homologue, *Slp* (sex limited protein) (32). The provirus has conferred androgen response on the adjacent *Slp* gene and therefore has been termed an impositon. In contrast to the impositon inserted in the *Slp* gene, however, *HERV-K(C4)* does not affect gene expression, as the same allotypes are experienced by long or short *C4B* genes (24).

The apparent abundance of such *RTLV* elements in the MHC may indicate that the MHC is a 'hot spot' for retrovirus integration. The human MHC is 4 Mb long and is thought to contain at least 170–200 genes. This high density of genes and the high transcriptional activity may reflect a favourable DNA matrix for retrovirus integration. However, some retroviruses are known to have integrated in late replicating domains with low transcriptional activity (16), and a combination of different factors most likely influences the integration site preference.

HERV-K(C4) shows the typical structure of a proviral DNA in spite of multiple terminations caused by stop codons in the *gag*, *pol* and *env* regions which preclude the production of functional proteins. As with other members of the *HERV-II* families, *HERV-K(C4)* is related to type A, B and D mammalian retroviruses, with the strongest homologies encompassing the RT and the rest of the *pol* region, indicating the conservation of the

most important domains of the polypeptide. The identification of a sequence complementary to tRNA^{Lys} located immediately after the LTRs suggests that *HERV-K(C4)* should be included in the group of *HERV-K* elements. However, differences found between the LTRs, *gag* and *pol* regions of *HERV-K(C4)* and those of other *HERV-K* sequences (2,9) indicate that it is distinct from them.

Hybridization techniques have revealed the presence of several sequences related to *HERV-K(C4)*, dispersed over several human chromosomes. Currently, however, the identity of these sequences is unknown. We have grouped these elements in a family of *HERVs*, which is novel when the sequence, integration and dispersion specificity are taken into account. Homologous sequences exist also in the genomes of other primates, indicating the evolutionary conservation of some members of this family. The unique and uniform pattern of hybridization within the genome of primates shows no polymorphism and stability of the proviral DNA.

Some inferences can be drawn on the timing of the integration of *HERV-K(C4)* into the *C4* gene cluster, and its relationship with the timing of the *C4* duplication event. The lack of cross-hybridizing bands in the marmoset suggests that integration most likely occurred after the divergence of New World monkeys from Old World monkeys, as previously demonstrated for *HML-2* sequences (12). The *C4* genes are known to be duplicated in humans and also in Old World monkeys (33), suggesting that the gene duplication event has predated the divergence of Hominoids from Old World monkeys (~30 million years ago). There are cross-hybridizing bands in chimpanzee, indicating the integration of closely related members of the same family. However, this integration did not apparently extend to the *C4* genes, as indicated by the presence of only 'short' *C4A* and *C4B* genes in the chimpanzee (33; M.Lako, unpublished observations). In humans and orang-utangs, the integration most likely occurred in a *C4A* gene following the duplication event, because all extant *C4A* alleles but only about 65% of the *C4B* alleles appear to possess the *HERV-K(C4)* element. After the initial integration event in the *C4A* gene, subsequent interlocus exchange mechanisms, such as unequal crossover or gene conversion-like events, resulted in a proportion of *C4B* genes acquiring this element. In favour of this idea, a unidirectional exchange of information from *C4A* to *C4B* genes has been reported (34). The alternative explanation that a *HERV* element integrated in a *C4* gene prior to duplication seems unlikely: one would have to postulate that this element was subsequently lost from both *C4* genes in chimpanzee, while in humans and orang-utangs it was lost from only some *C4B* genes. Therefore, the gene duplication and the integration of the retrovirus most likely represent independent events, as previously demonstrated for the haptoglobin gene cluster of primates (5).

For the moment, we know nothing about the nature of the original provirus which gave rise to the defective copies, one of which integrated in *C4* genes. Whether the members of this novel family are the result of retrotransposition of a few copies or of only one original provirus remains unknown. Furthermore, we do not know whether the primary pattern of amplification was of a defective or of a functional element. If the original provirus were able to code for the necessary proteins for retrotransposition, it is likely that its amplification led to dispersion of a few more *RTVL* elements into the genome. This amplification and dispersion process might also be the cause of mutation accumulation in defective elements. Otherwise, amplification and dispersion of defective elements, which might

have happened in the case of *RTVL-H* (13), implies the existence of a source of proteins necessary for reverse transcription and integration in new loci. It is likely that functional elements similar to *HERV-K(C4)* existing in the human genome provide this necessary pool of proteins, making possible the amplification and dispersion of some defective family members.

ACKNOWLEDGEMENTS

We thanks Dr M.Jackson for a useful discussion during the preparation of the manuscript and Mr A.Curtis for technical assistance. This work has been supported by Wellcome Trust grants (030289/Z/89 and 038716/Z/93).

REFERENCES

1. Kroger, B. and Horak, I. (1987) *J. Virol.*, **61**, 2071–2075.
2. Ono, M. (1986) *J. Virol.*, **58**, 937–944.
3. Varmus, H.E. (1982) *Science*, **216**, 812–820.
4. Kambhu, S., Falldorf, P. and Lee, J.S. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4927–4931.
5. Maeda, N. and Kim, H.S. (1990) *Genomics*, **8**, 671–683.
6. Mager, D.L. and Henthorn, P.S. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7510–7514.
7. Martin, M.A., Bryan, T., Rasheed, S. and Khan, A.S. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 4892–4896.
8. Medstrand, P. and Blomberg, J. (1993) *J. Virol.*, **67**, 6778–6787.
9. Ono, M., Yasunaga, T., Miyata, T. and Ushikubo, H. (1986) *J. Virol.*, **60**, 589–598.
10. Repaske, R., Steele, P.E., O'Neill, R., Rabson, A.B. and Martin, M.A. (1985) *J. Virol.*, **54**, 764–772.
11. Zhu, Z.B., Jian, B. and Volanakis, J.E. (1994) *Hum. Genet.*, **93**, 545–551.
12. Constantini-Mariani, R., Horn, T. and Callahan, R. (1989) *J. Virol.*, **63**, 4982–4985.
13. Godchild, N.L., Wilkinson, D.A. and Mager, D.L. (1993) *J. Virol.*, **196**, 778–788.
14. Wilkinson, D.A., Goodchild, N.L., Saxton, T.M., Wood, S. and Mager, D.L. (1993) *J. Virol.*, **67**, 2981–2989.
15. Steele, P.E., Martin, M.A., Rabson, A.B., Bryan, T. and O'Brien, S.J. (1986) *J. Virol.*, **59**, 545–550.
16. Taruscio, D. and Manuelidis, L. (1991) *Chromosoma*, **101**, 141–156.
17. Anderson, M.J., Milner, C.M., Cotton, R.G.H. and Campbell, R.D. (1992) *J. Immunol.*, **148**, 2795–2802.
18. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
19. Pearson, W.R. and Lipman, D.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.
20. Gitelman, S.E., Bristow, J. and Miller, W.L. (1992) *Mol. Cell. Biol.*, **12**, 2121–2134.
21. Sargent, C.A., Anderson, M.J., Hsieh, S., Kendall, E., Gomez-Escobar, N. and Campbell, R.D. (1994) *Hum. Mol. Genet.*, **3**, 481–488.
22. Carroll, M.C., Belt, K.T., Palsdottir, A. and Porter, R.R. (1984) *Phil. Trans. R. Soc. Lond.*, **306**, 379–388.
23. Carroll, M.C., Campbell, R.D., Bentley, D.R. and Porter, R.R. (1984) *Nature*, **307**, 237–241.
24. Palsdottir, A., Fossdal, R., Arnason, A., Edwards, H.J. and Jensson, O. (1987) *Immunogenet.*, **25**, 299–304.
25. Rittner, C., Hauptman, G., Grosse-Wilde, H., Grosshans, E., Tongio, M.M. and Mayer, S. (1975) In F. Kissmeyer-Nielsen (ed.), *Histocompatibility testing*. Munksgaard, Copenhagen, pp 945–953.
26. Schendel, D.J., O'Neill, G.J. and Wank, R. (1984) *Immunogenet.*, **20**, 23–31.
27. Yu, C.Y. (1991) *J. Immunol.*, **146**, 1057–1066.
28. Khoury, G. and Gruss, P. (1983) *Cell*, **33**, 313–314.
29. Temin, H.M. (1981) *Cell*, **27**, 1–3.
30. Xiong, Y. and Eickbush, T.H. (1990) *EMBO J.*, **9**, 3353–3362.
31. Shih, C.C., Stoye, J.P. and Coffin, J.M. (1988) *Cell*, **53**, 531–537.
32. Stavenhagen, J.B. and Robins, D.M. (1988) *Cell*, **55**, 247–254.
33. Bontrop, R.E., Broos, L.A.M., Otting, N. and Jonker, M.J. (1991) *Tissue Antigens*, **7**, 145–151.
34. Partanen, J. and Campbell, R.D. (1989) *Immunogenet.*, **30**, 520–523.