Journal of General Virology (1999), **80**, 835–839. Printed in Great Britain SHORT COMMUNICATION

A human endogenous retrovirus-like (HERV) LTR formed more than 10 million years ago due to an insertion of HERV-H LTR into the 5′ LTR of HERV-K is situated on human chromosomes 10, 19 and Y

Anna V. Lapuk, Pavel P. Khil, Irina V. Lavrentieva, Yuri B. Lebedev and Eugene D. Sverdlov

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya 16/10, Moscow, GSP-7, 117871 Russia

A chimeric long terminal repeat (LTR) containing the whole LTR of a human endogenous retrovirus-like element of the H family (HERV-H) inserted downstream of the core enhancer region of the 5′ LTR of a HERV-K retroelement was detected and sequenced in the human 19p12 locus, known to be enriched with genes encoding zinc finger proteins. Similar chimeras were also detected in human chromosomes 10 and Y in human–hamster hybrid cells containing individual human chromosomes. This finding was interpreted as evidence of transpositions of the chimera in the genome. PCR analyses detected the chimera in the genomes of chimpanzee and gorilla, but not in that of orang-utan. These data demonstrate that the chimera appeared in the primate germ cells more than 10 million years ago, before divergence of the human/chimpanzee and the gorilla lineages. The combination of the two LTRs forms a new regulatory system that can be involved in nearby gene expression.

Endogenous retrovirus-like (ERV) sequences inherited as stable Mendelian genes have been found in the genomes of most vertebrate species. Human ERVs (HERVs) exogenous progenitors are thought to infect the germ line in the course of the evolution of primates and then spread throughout genomes by retrotranspositions (for review see Lower *et al.*, 1996). HERVs represented by several distinct families were estimated to occupy up to 1% of the human genome (Leib-Mosch *et al.*, 1993; Lower *et al.*, 1996). A multitude of HERVs lost their viral genes, most probably due to homologous recombination between 5′ and 3′ long terminal repeats (LTRs), thus forming so-called solitary LTRs. The plethora of HERV elements containing promoters, enhancers and other regulatory regions can affect the genome regulation and evolution by means of interactions with adjacent cellular genes. There is well-documented evidence of HERVs functioning as components of cellular gene promoters (Di Cristofano *et al.*, 1995; Schulte *et al.*, 1996) or signals of transcript polyadenylation (Kazmierczak *et al.*, 1996), as well as evidence of their involvement in tissue-specific expression of human genes (Schulte & Wellstein, 1998).

To gain a deeper insight into the mechanisms of interplay between ERV elements and other components of the human genome, we are now trying to implement a genome-wide scan of coincidences of LTRs and human genes using the resources provided by the Human Genome Program. Previously, we identified all of the HERV-K LTRs on human chromosome 19 (Vinogradova *et al.*, 1997; Lavrentieva *et al.*, 1998). Moreover, we found about 40 of the 72 mapped full-size LTRs located close to known genes, some of which code for transcriptional factors and signal pathway proteins (Vinogradova *et al.*, 1997). The studied LTRs relate to the HERV-K10 sequence belonging to the HML-2 group of elements and represent one of the most abundant HERV families (Leib-Mosch *et al.*, 1993; Zsíros *et al.*, 1998). The nucleotide sequences of HERV-K LTRs found on human chromosome 19 and retrieved from databases fall into two major subfamilies comprising 16 distinct groups with estimated evolutionary ages of 7–45 million years (Lavrentieva *et al.*, 1998).

One of these LTRs gave a product of about 1400 bp in length after PCR amplification from a chromosome-19-specific cosmid with primers targeted at conservative regions at the 3′ and 5′ ends of the LTR consensus sequence (Lavrentieva *et al.*, 1998); other cosmids produced LTR amplicons about 920 bp long. This 'long' LTR was mapped on a 7·02 kb *Eco*RI restriction fragment of cosmid R31578 in the 19p12.0 locus (http://www-bio.llnl.gov/genome/html/chrom map.html) and sequenced. A set of LTR-specific primers and the *fmol* DNA Cycle sequencing system (Promega) were used for

Author for correspondence: Yuri Lebedev.
Fax +7 095 330 6538. e-mail yuri@humgen.siobc.ras.ru

The GenBank accession number of the sequence reported in this paper is AF078838.

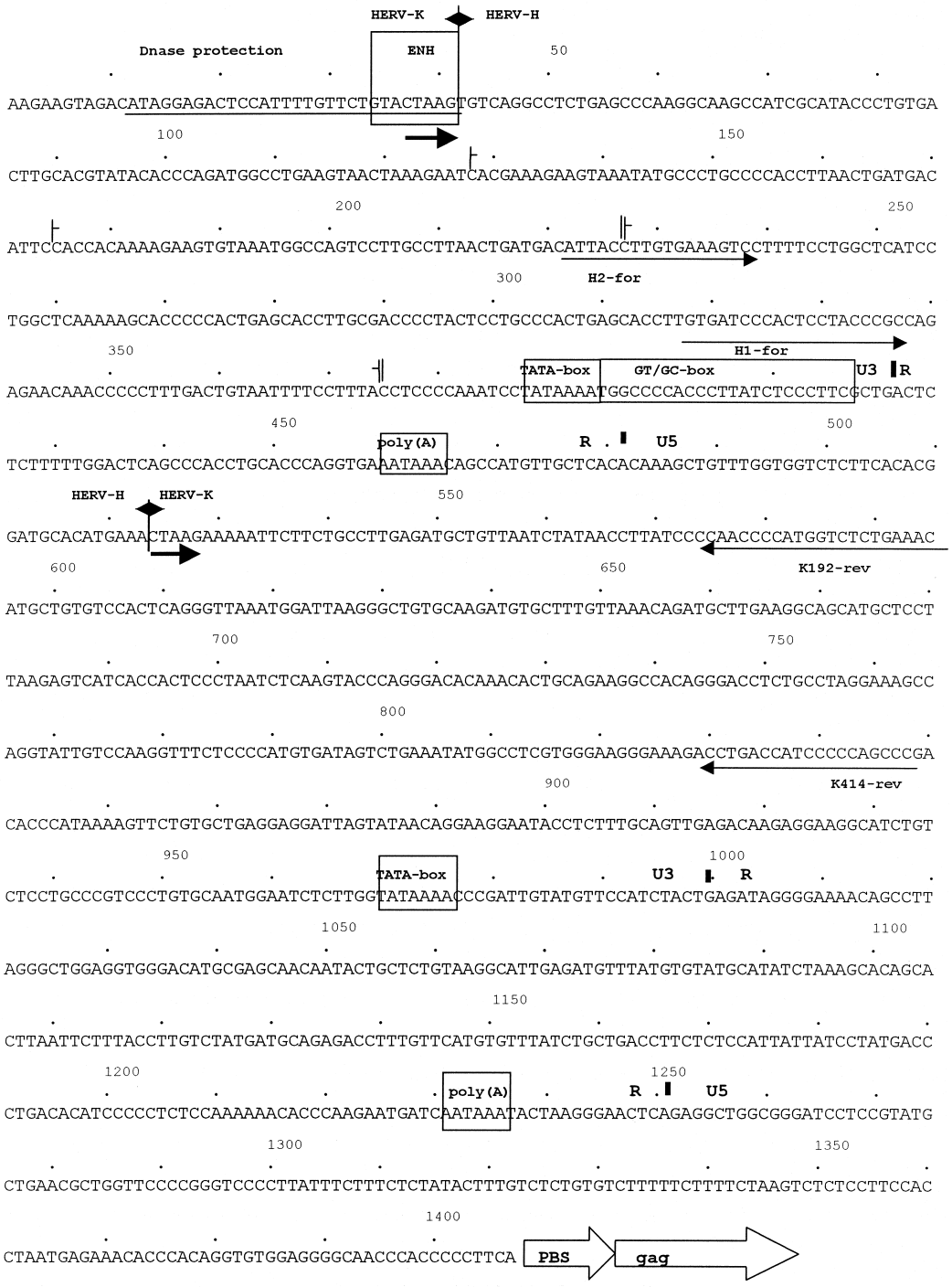

Fig. 1. The chimeric HERV-K–HERV-H LTR primary structure (GenBank accession no. AF078838). The double-headed bold arrows mark the borders between the HERV-K and HERV-H LTRs. The LTR regulatory regions are boxed and designated as TATA-box, ENH (core enhancer homology), poly(A) for the polyadenylation signal, and GT/GC-box. The borders of the U3, R and U5 functional regions of LTRs are marked by vertical lines. Symbols ⊢ and ⊩ mark the border between type I repeats and unique region I sequences of the HERV-H LTR. The bold horizontal arrows below the sequence mark the short direct repeat flanking the HERV-H LTR insert. The long arrows show the location of the oligonucleotide primers used in PCR analyses. The position of the area protected from DNase action by nuclear proteins is indicated by underlined letters. Empty arrows at the 3′ terminus of the chimeric LTR sequence indicate positions of the primer-binding site (PBS) and the 5′ region of the *gag* gene attached to the LTR.

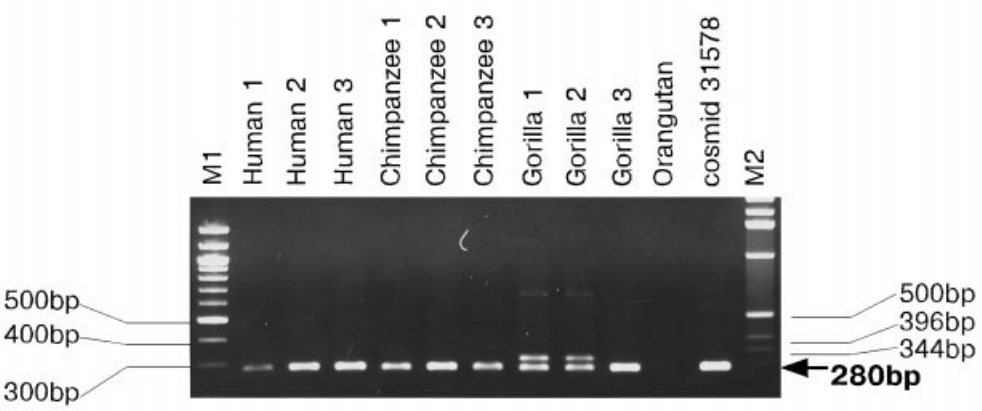

Fig. 2. Nested PCR of primate genomic DNA (10 ng samples). The first round was performed with forward primer H2-for (5′ ACATTACCTTGTGAAAGTCCT 3′) and reverse primer K414-rev (5′ GGGCTGGGGGATGGTCAGGT 3′), both shown in Fig. 1. Three human DNAs from unrelated individuals, three chimpanzee and three gorilla DNAs and one sample of orangutan DNA were used as indicated. The PCR products were diluted 250-fold and used as templates for the second round with forward primer H1-for (5′ TGTGATCCCCACTCCTACCCG 3′) and reverse primer K192-rev (5′ TGTTTCAGAGACCATGGGGTTGGG 3′) (see Fig. 1 for the primer positions). Cosmid, amplification of cosmid 31578 DNA from a chromosome-19-specific library. The bold arrow marks the position of the expected PCR product. The fragment lengths of 100 bp DNA (M1) and 1 kb DNA ladders (M2) (Gibco BRL) are indicated.

sequencing. A sequence downstream of the LTR was obtained using the primer-walking strategy. The resulting structure of about 1400 bp corresponds to the 5′ LTR of a HERV-K family retroelement with an insert (Fig. 1). This insert represents a solitary 482 bp HERV-H LTR (Anderssen *et al*., 1997) flanked by the short direct CTAAG repeats. This LTR mosaic was followed by a 300 bp DNA stretch containing a primer-binding site (PBS) complementary to the $tRNA^{Lys}$ and partial *gag* sequences characteristic of HERV-K.

Our finding that the HERV-H element was integrated into the HERV-K proves that this integration is more recent than the integration of HERV-K into the human genome. The HERV-K LTR was assigned to the II-V subgroup, which started its expansion in the genome about 25 million years ago (Lavrentieva *et al*., 1998). The sequences of the LTR and the II-V consensus are 2·4% diverged. Assuming no selection for or against mutation in the integrated sequence, the estimated age of the HERV-K LTR should be about 13–15 million years, provided that the mutation rate is 0·2% (Anderssen *et al*., 1997) or 0·26% (Lavrentieva *et al*., 1998) per million years. The integrated HERV-H LTR is structurally similar to the I-2 subtype of HERV-H sequences. The evolutionary age of the HERV-H type I elements was assessed to be 18–67 million years, and several expansion waves of the elements were proposed (Anderssen *et al*., 1997). The apparent discrepancy between the order of integration of the two LTRs and their estimated ages can be explained by the assumption that the original HERV-H master gene was still transpositionally active at the time of the insertion or that the age estimations are inaccurate due to some selective pressure having affected the mutation rate of both the LTRs.

The occurrence of the chimeric retroelement in the genomes of the apes was tested by PCR amplification of the HERV-K–HERV-H junction region. Two pairs of specific primers (Fig. 1) corresponding to the HERV-H and HERV-K parts, respectively, were used according to the nested primer strategy. PCR products of the expected 280 bp size were detected with human, chimpanzee and gorilla DNA, but not with orangutan DNA (Fig. 2). The negative result can be explained either by the absence of the chimera in the orangutan genome or by its higher divergence in regions corresponding to the PCR primers. Anyway, the data obtained demonstrate that the chimera appeared in the primate germ line prior to the divergence of the human/chimpanzee and gorilla lineages, i.e. over 10 million years ago, the figure being in good agreement with the chimera age deduced from the mutation rates.

The distribution of the chimeric retroelement in the human genome was studied by PCR analysis with the same chimera-specific primers and DNA from a panel of rodent–human somatic cell hybrids containing individual human chromosomes (Fig. 3). The expected PCR products were obtained from the DNA of human chromosomes 10, 19 and Y. Therefore, the chimeric retroelement was most probably propagated after the chimera appearance since repeated independent formation of identical structures seems rather unlikely. In turn, the retro-positions might indicate the recent master gene function of the chimeric structure. If a prerequisite for the master gene activity of a sequence is its ability to be reverse-transcribed, then the sequence could be expected to bear a PBS. Indeed, the HERV-K–HERV-H LTR present on chromosome 19 was found to be joined to a PBS for the $tRNA^{Lys}$. As mentioned above, there are also *gag*-like sequences in the neighbourhood, but it remains unknown whether other retroviral genes are presented in the retroelement.

Nested integration is characteristic of retroelements (Kidwell & Lisch, 1997). We have found other examples of

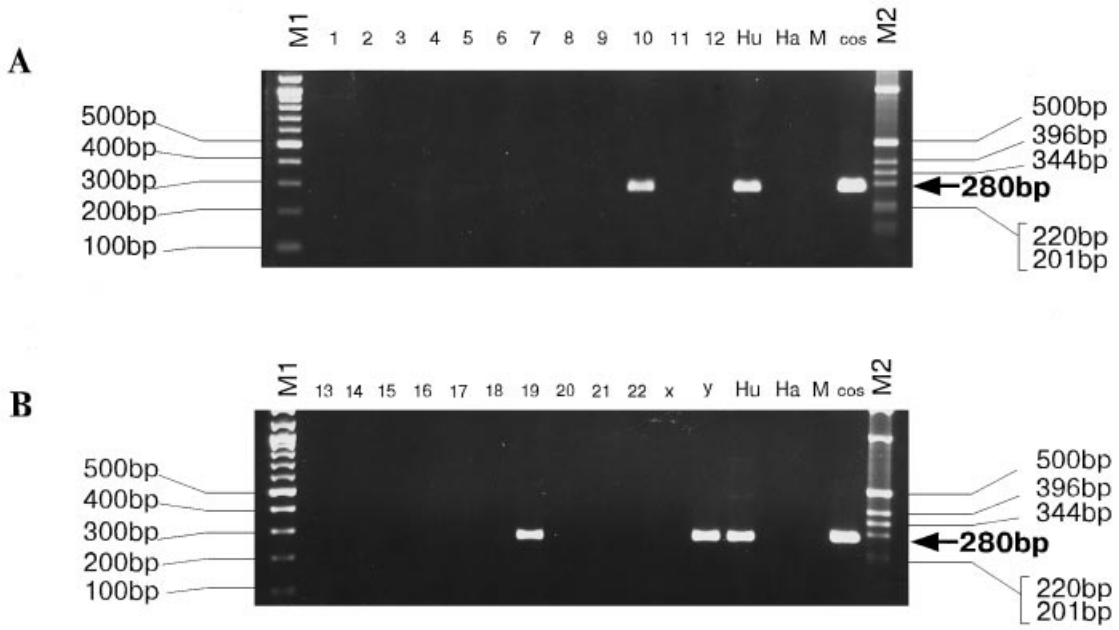

Fig. 3. PCR amplification on DNAs from human–rodent hybrid cell lines containing single human chromosomes (PCRable DNA MSC Hybrid panel, QUANTUM Biotechnologies) 1–12 (A) and 13–22, X and Y (B). The same strategy as in Fig. 1 was used except that the primers for the first PCR round were H1-for and K414-rev (for their structures see Fig. 1 and the legend to Fig. 2). The first round PCR product was diluted 125-fold and used as a template for the second round. Hu, Ha, M: human, hamster and mouse DNA control amplifications, respectively; cos: the amplification product of cosmid 31578 DNA. Bold arrows mark the position of a 280 bp long PCR product. Molecular mass markers M1 and M2 were as in Fig. 2.

similar LTR-into-LTR insertions in human genomic sequences deposited in databases. For instance, an ERV9 LTR inserted into a unique HuERs-P3 LTR is located in the Prader–Willi/Angelman syndrome region of human chromosome 15q11–q13 (accession no. AC004738). Also, a HERV-K LTR integrated into a unique HERV-H LTR was detected among sequences of human Y chromosome (accession no. AC002992).

Thus, integration/recombination events between different HERVs and/or solitary LTRs seem to be common (Goodchild *et al.*, 1993; Lindeskog *et al.*, 1998). This kind of event might enrich the process of genome evolution with new opportunities due to shuffling the modules that affect the expression of the neighbouring genes. In particular, the HERV-K LTR of the chimera includes a conservative DNA stretch capable of the specific binding of human nuclear protein factors (Akopov *et al.*, 1998). On the other hand, its HERV-H LTR insert contains a GC/GT-box necessary for transcriptional activity of this LTR subtype (Sjottem *et al.*, 1996; Nelson *et al.*, 1996). Therefore, a new combination of regulatory modules appeared that might play some role in the gene regulation. It is interesting to note that the chimera is located within a chromosome 19 area enriched with genes encoding zinc finger proteins known as transcription regulators (Ashworth *et al.*, 1995).

Note added in proof. Further research revealed that not simply was a solitary LTR integrated into the HERV-K 5′ LTR but rather the HERV-H provirus with two LTRs was integrated, at the same position. PCR amplifications of a hybrid retroelement with our primers reproducibly and efficiently produced a fragment containing a single LTR as described in the paper. The effect did not depend on the template contained in the hybrid. Most probably, the viral sequences were excised due to extended primer displacement from the 5′ HERV-H LTR to the homologous sequence in the 3′ HERV-H LTR. Such an excision is probably very typical in PCR amplification of endogenous retroviruses.

We are grateful to Professor G. Hunsmann for providing DNA samples from apes. This work was partially supported by a grant from the Human Genome State Project of Russia and an INTAS-96-1710 grant.

References

Akopov, S. B., Nikolaev, L. G., Khil, P. P., Lebedev, Y. B. & Sverdlov, E. D. (1998). Long terminal repeats of human endogenous retrovirus K family (HERV-K) specifically bind host cell nuclear proteins. *FEBS Letters* **421**, 229–233.

Anderssen, S., Sjottem, E., Svineng, G. & Johansen, T. (1997). Comparative analyses of LTRs of the ERV-H family of primate-specific retrovirus-like elements located from marmoset, African green monkey, and man. *Virology* **234**, 14–30.

Ashworth, L. K., Batzer, M. A., Brandriff, B., Branscomb, E., de Jong, P., Garcia, E., Garnes, J. A., Gordon, L. A., Lamerdin, J. E., Lennon, G., Mohrenweiser, H., Olsen, A., Slezak, T. & Carrano, A. V. (1995). An integrated metric physical map of human chromosome 19. *Nature Genetics* **11**, 422–427.

Di Cristofano, A., Strazzullo, M., Longo, L. & La Mantia, G. (1995). Characterization and genomic mapping of the ZNF80 locus: expression

of this zinc-finger gene is driven by a solitary LTR of ERV9 endogenous retroviral family. *Nucleic Acids Research* **23**, 2823–2830.

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

Kazmierczak, B., Pohnke, Y. & Bullerdiek, J. (1996). Fusion transcripts between the HMGIC gene and RTVL-H-related sequences in mesenchymal tumors without cytogenetic aberrations. *Genomics* **38**, 223–226.

Kidwell, M. G. & Lisch, D. (1997). Transposable elements as sources of variation in animals and plants. *Proceedings of the National Academy of Sciences, USA* **94**, 7704–7711.

Lavrentieva, I., Khil, P., Vinogradova, T., Akhmedov, A., Lapuk, A., Shakhova, O., Lebedev, Y., Monastyrskaya, G. & Sverdlov, E. D. (1998). Subfamilies and nearest-neighbour dendrogram for the LTRs of human endogenous retroviruses HERV-K mapped on human chromosome 19: physical neighbourhood does not correlate with identity level. *Human Genetics* **102**, 107–116.

Leib-Mosch, C., Haltmeier, M., Werner, T., Geigl, E. M., Brack-Werner, R., Francke, U., Erfle, V. & Hehlmann, R. (1993). Genomic distribution and transcription of solitary HERV-K LTRs. *Genomics* **18**, 261–269.

Lindeskog, M., Medstrand, P., Cunningham, A. A. & Blomberg, J. (1998). Coamplification and dispersion of adjacent human endogenous retroviral HERV-H and HERV-E elements: presence of spliced hybrid transcripts in normal leukocytes. *Virology* **244**, 219–229.

Lower, R., Lower, J. & Kurth, R. (1996). The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proceedings of the National Academy of Sciences, USA* **93**, 5177–5184.

Nelson, D. T., Goodchild, N. L. & Mager, D. L. (1996). Gain of Sp1 sites and loss of repressor sequences associated with a young, transcriptionally active subset of HERV-H endogenous long terminal repeats. *Virology* **220**, 213–218.

Schulte, A. M. & Wellstein, A. (1998). Structure and phylogenetic analysis of an endogenous retrovirus inserted into the human growth factor gene pleiotrophin. *Journal of Virology* **72**, 6065–6072.

Schulte, A. M., Lai, S., Kurtz, A., Czubayko, F., Riegel, A. T. & Wellstein, A. (1996). Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus. *Proceedings of the National Academy of Sciences, USA* **93**, 14759–14764.

Sjottem, E., Anderssen, S. & Johansen, T. (1996). The promoter activity of long terminal repeats of the HERV-H family of human retrovirus-like elements is critically dependent on Sp1 family proteins interacting with a GC/GT box located immediately 3′ to the TATA-box. *Journal of Virology* **70**, 188–198.

Vinogradova, T., Volik, S., Lebedev, Y., Shevchenko, Y., Lavrentyeva, I., Khil, P., Grzeschik, K.-H., Ashworth, L. K. & Sverdlov, E. D. (1997). Positioning of 72 potentially full size LTRs of human endogenous retroviruses HERV-K on the human chromosome 19 map. Occurrence of the LTRs in human gene sites. *Gene* **199**, 255–264.

Zsíros, J., Jebbink, M. F., Lukashov, V. V., Voûte, P. A. & Berkhout, B. (1998). Evolutionary relationships within a subgroup of HERV-K-related human endogenous retroviruses. *Journal of General Virology* **79**, 61–70.

Received 7 August 1998; Accepted 17 December 1998