



## UvA-DARE (Digital Academic Repository)

### Evolutionary relationships within a subgroup of HERV-K-related human endogenous retroviruses

Zsiros, J.; Jebbink, M.F.; Lukashov, V.V.; Voûte, P.A.; Berkhout, B.

#### Publication date

1998

#### Published in

Journal of general virology

[Link to publication](#)

#### Citation for published version (APA):

Zsiros, 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*, 61-70.

#### General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

## Evolutionary relationships within a subgroup of HERV-K-related human endogenous retroviruses

J. Zsíros,<sup>1,2</sup> M. F. Jebbink,<sup>2</sup> V. V. Lukashov,<sup>2</sup> P. A. Voûte<sup>1</sup> and B. Berkhout<sup>2</sup>

Departments of Pediatric Oncology<sup>1</sup> and Human Retrovirology<sup>2</sup>, Academic Medical Center, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

The prototype endogenous retrovirus HERV-K10 was identified in the human genome by its homology to the exogenous mouse mammary tumour virus. By analysis of a short 244 bp segment of the reverse transcriptase (RT) gene of other HERV-K10-like sequences, it has become clear that these elements represent an extended family consisting of multiple groups (the HML-1 to HML-6 subgroups). Some of these elements are transcriptionally active and contain an intact open reading frame for the RT protein, raising the possibility that this family is still expanding through retrotransposition. To better define the relationship of these endogenous retroviruses, we identified ten new members of the

HML-2 subgroup. PCR was used to amplify reverse-transcribed RNA of a 595 bp region of the RT gene in a variety of human cell samples, including normal and leukaemic bone marrow and peripheral blood, placenta cells and a transformed T cell line. We provide an extensive phylogenetic analysis of the relationships for this cluster of HERV-K-related endogenous retroviral elements. Nucleotide diversity values for nonsynonymous versus synonymous codon positions indicate that moderately strong selection is or was operating on these retroviral RT gene segments. The evolution of this class of endogenous retroelements is discussed.

### Introduction

The human genome contains a wide variety of endogenous retrovirus-like elements that may constitute as much as 0·1–0·6% of the total genome (reviewed in Wilkinson *et al.*, 1994). These endogenous proviruses presumably result from an ancient infection of the germ cells by exogenous viruses. Such elements became fixed within the population and currently appear to form relatively stable components of the human genome. Some of these elements are present at the same location in the genomes of all primates, indicating that the element inserted at the time when these species shared a common ancestor, which is estimated to be at least 25 million years ago.

The HERV-K family of endogenous retroelements is particularly interesting because there is some evidence that members of this family are still biologically active. First, although HERV-K sequences have been detected in the genomes of great apes and Old World monkeys, New World monkeys do not possess this family of sequences (Mariani-

Costantini *et al.*, 1989). Furthermore, there are indications of a unique insertion in the chimpanzee genome, which could be the result of either a retrotransposition event or reinfection during recent evolutionary history (Craig *et al.*, 1991). Second, most HERV-K members are transcriptionally active (Medstrand & Blomberg, 1993) and, unlike many of the other sequenced HERV elements, the prototype HERV-K10 provirus contains very few mutations that interrupt the viral open reading frames (Ono *et al.*, 1986). Third, there is now substantial evidence that HERV-K elements encode the retrovirus-like particles produced by the teratocarcinoma cell line GH and the particles detected in the placenta of primates (summarized in Lower *et al.*, 1996), and HERV-K10-like sequences were also found in particles produced by the mammary carcinoma cell line T47D (Patience *et al.*, 1996).

We set out to identify additional HERV-K members in order to analyse in more detail the phylogeny of this family of endogenous retroviruses. Early estimates on the copy numbers suggested that approximately 50 copies of HERV-K10-like elements are present in the haploid human genome (Ono *et al.*, 1986). Cross-hybridization studies revealed nine different groups related to mouse mammary tumour virus (MMTV) and to each other (Franklin *et al.*, 1988). More recently, PCR, in

**Author for correspondence:** Ben Berkhout.

Fax +31 20 691 6531. e-mail B.Berkhout@AMC.UVA.NL

**Table 1.** Endogenous retroviral sequences identified by PCR with the MOPs-2 primer set

Clone	Source*	Origin	Length (bp)†	Mostly related to:‡		
				Name	Length (bp)	%
M3.8	BM/T-ALL	RNA	586	HERV-K10	9179	95.1
HP.1	Hum. plac. cells	DNA	591	HERV-K10	9179	94.8
K1.1	BM/cALL rem.	RNA	594	HERV-(K)73	595	99.5
N8.4	BM/T-ALL	RNA	595	HERV-(K)55	594	99.9
	P1.2	BM/normal	RNA	595	HERV-(K)55	595
P1.1	BM/normal	RNA	595	HML 2.1	244	98.8
M3.5	BM/T-ALL	RNA	604	HML 2.1	244	97.5
	M3.6	BM/T-ALL	RNA	604	HML 2.1	244
P1.8	BM/normal	RNA	604	HML 2.1	244	98.0
P1.10	BM/normal	RNA	585	HML 2.5	240	98.8
D1.2	BM/cALL rem.	RNA	594	HML 2.6	244	98.0
M3.7	BM/T-ALL	RNA	593	HML 2.6	244	98.0
D1.3	BM/cALL rem.	RNA	554	HM16	2139	99.6
H3.1	BM/cALL	RNA	594	clone 5:30	244	99.6
St. 2	Sup T1 T-cells	DNA	595	HML 2.2	244	99.6
P1.4	BM/normal	RNA	585	HML 2.5	240	100
L4.4	BM/ANLL	RNA	595	clone 5:30	244	100
	L4.1	BM/ANLL	RNA	595	clone 5:30	244
C1.3	PB/HCL	RNA	554	HM16	2139	100
HP.3	Hum. plac. cells	DNA	554	HM16	2139	100

\* BM, bone marrow; PB, peripheral blood; T-ALL, T-cell acute lymphoblastic leukaemia; cALL, common acute lymphoblastic leukaemia; rem, in remission; ANLL, acute nonlymphoblastic leukaemia; HCL, hairy-cell leukaemia; hum. plac, human placenta.

† Excluding primers.

‡ Nucleotide sequence comparison with published elements.

combination with primers derived from conserved regions of the reverse transcriptase (RT) gene, was used to identify six HERV-K-related subgroups with a nucleotide sequence dissimilarity of about 25% between the groups (Medstrand & Blomberg, 1993). These authors termed the subgroups HML-1 to HML-6 (human endogenous MMTV-like) and, although there is some controversy regarding this nomenclature (Simpson *et al.*, 1996), we have used these names throughout this study. Most PCR studies to date have used a primer pair that amplifies an RT fragment of approximately 250 bp. In this study, we used a primer set designed by Li *et al.* (1996) that has the advantage of amplifying a much larger RT fragment of approximately 600 bp. A few known HERV-K10-like sequences were found and ten new members of the HML-2 subgroup were identified. The nucleotide sequence data of this larger RT fragment made it possible to clarify the structure of this subgroup of the HERV-K family.

## Methods

■ **Tissues and cells.** Bone marrow was obtained by routine diagnostic puncture from children with acute leukaemia at diagnosis [# N8 and M3 with acute T-cell lymphoblastic leukaemia (T-ALL), # H3 with acute

precursor B-cell lymphoblastic leukaemia (cALL), # L4 with acute non-lymphoblastic leukaemia FAB M4 (ANLL M4) and patients with cALL in complete remission (# K1 and D1)]. Normal bone marrow was obtained from one patient with a malignant solid tumour without bone marrow involvement (# P10). We also used a peripheral blood sample from a patient with hairy cell leukaemia with peripheral blood involvement (generously donated by J. C. Kluijn-Nelemans, Academic Hospital, University of Leiden, The Netherlands). Human placenta tissue was obtained from the Department of Obstetrics (AMC). The lymphoblastoid SupT1 T cell line (Smith *et al.*, 1984) was cultured as described previously (Das *et al.*, 1995).

■ **RNA/DNA isolation and preparation of cDNA.** Bone marrow and peripheral blood mononuclear cells were isolated directly after puncture by gradient centrifugation with Ficoll-Paque (Pharmacia), and washed twice with PBS. Cells were then either stored at  $-70^{\circ}\text{C}$  or lysed directly in 4 M guanidinium thiocyanate solution and total cellular RNA was purified on a CsCl gradient as described previously (Chirgwin *et al.*, 1979). Cellular DNA was isolated by standard phenol-chloroform extraction (Sambrook *et al.*, 1989) from the SupT1 T-cell line and human placenta cells. RNA and DNA samples were solubilized in 30  $\mu\text{l}$  TE buffer and stored at  $-70^{\circ}\text{C}$ . In the RT-PCR protocol we synthesized cDNA as follows: 2  $\mu\text{l}$  RNA was incubated for 10 min at  $65^{\circ}\text{C}$ , and then incubated for 2 h at  $38^{\circ}\text{C}$  with 5  $\mu\text{g}$  random hexaoligonucleotides (Boehringer) in a total volume of 20  $\mu\text{l}$  annealing-buffer (50 mM Tris-HCl pH 7.4, 10 mM  $\text{MgCl}_2$ , 1 mM DDT, 0.2 mg/ml BSA) with 20 U RNasin

	N8.4	<i>HERV-(K)55</i>	<i>HERV-(K)27</i>	L4.4	H3.1	P1.1	<i>HERV-(K)67</i>	T47D	P1.8	M3.5	St.2	M3.8	HP.1	P1.10	P1.4	K1.1	D1.2	<i>HERV-(K)73</i>	D1.3	HML6
<i>HERV-K10</i>	99,2	99,0	98,8	98,3	98,2	98,7	98,0	97,5	97,8	97,3	96,8	95,1	94,8	91,8	91,8	89,1	88,7	89,2	89,0	89,4
N8.4	--	99,8	99,3	98,8	98,7	99,2	98,2	98,0	98,7	98,2	97,7	95,9	95,6	92,7	92,7	89,9	89,6	90,1	89,7	90,1
<i>HERV-(K)55</i>		--	98,3	98,8	98,5	99,0	98,2	98,0	98,7	98,2	97,5	95,6	95,3	91,5	91,5	89,7	89,4	90,1	89,5	89,9
<i>HERV-(K)27</i>			--	98,2	97,8	98,5	98,5	98,0	98,7	98,2	97,6	95,7	95,4	90,9	90,9	89,4	89,4	89,7	89,4	89,7
L4.4				--	99,6	98,7	97,5	97,5	97,5	97,0	97,1	95,4	95,1	91,5	91,5	90,1	89,7	90,3	89,2	89,5
H3.1					--	98,5	97,3	97,3	97,3	96,8	97,0	94,7	94,4	89,1	89,1	89,4	89,1	89,7	88,8	89,2
P1.1						--	98,5	97,7	97,8	97,3	97,3	95,7	95,4	92,1	92,1	89,1	88,7	89,2	88,8	89,2
<i>HERV-(K)67</i>							--	98,3	98,5	98,0	97,5	95,7	95,4	92,1	92,1	88,9	88,9	89,1	88,8	89,2
T47D								--	98,3	98,5	97,7	95,1	94,8	91,6	91,6	88,9	88,9	89,1	88,8	89,2
P1.8									--	99,2	97,7	96,3	96,1	93,0	93,0	89,7	89,7	89,8	89,7	90,1
M3.5										--	97,3	95,9	95,8	92,5	92,5	89,1	89,1	89,2	89,0	89,4
St.2											--	95,7	95,4	92,1	92,1	88,9	88,9	89,1	88,5	88,8
M3.8												--	99,5	87,5	88,4	87,9	88,4	87,9	86,3	86,6
HP.1													--	92,8	92,8	87,7	88,0	87,6	87,9	
P1.10														--	99,6	86,0	86,3	86,0	85,4	85,7
P1.4															--	86,0	86,3	86,0	85,2	85,6
K1.1																--	99,0	99,5	95,1	95,5
D1.2																	--	98,5	95,1	95,5
<i>HERV-(K)73</i>																		--	94,6	95,0
D1.3																			--	99,6
HML6																				--

Fig. 1. Percentage identity between the nucleotide sequences of HERV-K-related sequences. The analysis was performed on the approximately 600 bp RT fragment of HML-2-related sequences. Retroviral sequences obtained from the literature are shown in italics.

(Boehringer), 20 U AMV RT (Boehringer) and 10 mM deoxynucleoside triphosphates (dNTPs, Pharmacia).

**■ PCR amplification and cloning.** A 3  $\mu$ l sample of the RT reaction was amplified by PCR with the MOP-2 primer set consisting of degenerate oligonucleotides (Li *et al.*, 1996) in 100  $\mu$ l buffer (10 mM Tris-HCl pH 8.3, 2 mM MgCl<sub>2</sub>, 50 mM KCl) with 1  $\mu$ g of each MOP-2 primer, 250  $\mu$ M dNTPs and 2.5 U AmpliTaq DNA polymerase (Perkin Elmer). Samples were overlaid with 50  $\mu$ l of liquid paraffin and amplified in a DNA thermal cycler (Perkin Elmer) with the following reaction profile: 95 °C for 5 min and then 35 cycles of 95 °C for 1 min, 42 °C for 2 min and 72 °C for 2 min with a final extension step at 72 °C for 10 min. The PCR products were analysed on 1.5% Tris-borate-EDTA agarose gels and fragments of approximately 600 bp were extracted from the gel and purified by the QIAEX II method (Qiagen). The DNA was solubilized in 10  $\mu$ l, of which 3  $\mu$ l was used for ligation into the pCRII TA cloning vector (Invitrogen). The ligation mixture was transformed into competent *E. coli* (INV alpha, supplied with TA cloning kit). Individual clones were screened for inserts by PCR, and positive samples (purified DNA or PCR material) were subjected to sequence analyses on both strands with the T7 and M13 reverse primers on a 373 automated sequencer (ABI) according to the dye primer cycle sequencing protocol.

**■ Sequence analyses.** To study phylogenetic relationships within the HML-2 subgroup of endogenous retroviruses, nucleotide sequences were aligned with PCGENE software (IntelliGenetics). The alignment

was corrected manually. All positions with an alignment gap in at least one sequence were excluded from any pairwise sequence comparison. Nucleotide distances were calculated with PCGENE software. Synonymous and nonsynonymous nucleotide *p*-distances ( $d_s$  and  $d_a$ , respectively) were calculated with the MEGA program (Kumar *et al.*, 1993). The  $d_s/d_a$  ratio for a group of sequences was calculated as described previously (Lukashov *et al.*, 1995) according to the formula  $d_s/d_a = (\Sigma M_{si}/\Sigma S_{si})/(\Sigma M_{ai}/\Sigma S_{ai})$ , where  $\Sigma M_{si}$  and  $\Sigma M_{ai}$  represent the sum of (weighted) mutation events at coding synonymous and nonsynonymous sites, respectively, and  $\Sigma S_{si}$  and  $\Sigma S_{ai}$  the sum of coding synonymous and nonsynonymous sites, respectively. The neighbour-joining phylogenetic analysis was performed with the MEGA package, the distance matrix was generated with Kimura's two-parameter model. The bootstrap analysis was performed with 100 replications. The maximum parsimony analysis was performed with the PHYLIP package (Felsenstein, 1990).

**■ Nucleotide sequence accession numbers.** The nucleotide sequences of ten new members of the HML-2 subgroup presented in this study have been deposited in the GenBank database under the following accession numbers: clone M3.8 (U87587), HP.1 (U87588), K1.1 (U87589), N8.4 (U87590), P1.1 (U87591), M3.5 (U87592), P1.8 (U87593), P1.10 (U87594), D1.2 (U87595), D1.3 (U87596). The reference and accession numbers of published retroviral sequences that we included in this study are as follows: HML-2.1 (accession number U35104), HML-

HERVK10  
 N8.4  
 P1.1  
 L4.4  
 H3.1  
 P1.8  
 M3.5  
 D1.3  
 D1.1  
 D1.2  
 P1.10  
 P1.4  
 HP.1  
 M3.8  
 St.2

HERVK10  
 N8.4  
 P1.1  
 L4.4  
 H3.1  
 P1.8  
 M3.5  
 D1.3  
 D1.1  
 D1.2  
 P1.10  
 P1.4  
 HP.1  
 M3.8  
 St.2

HERVK10  
 N8.4  
 P1.1  
 L4.4  
 H3.1  
 P1.8  
 M3.5  
 D1.3  
 D1.1  
 D1.2  
 P1.10  
 P1.4  
 HP.1  
 M3.8  
 St.2

HERVK10  
 N8.4  
 P1.1  
 L4.4  
 H3.1  
 P1.8  
 M3.5  
 D1.3  
 D1.1  
 D1.2  
 P1.10  
 P1.4  
 HP.1  
 M3.8  
 St.2

Fig. 2.

HERVK10 \* V10KSKPHELTDIR...AVNAV10PMPLOPCLSPAM1PKDWELI1IDLDGCFPTPLAEOCEKFAFTIPAINNKSPATFQMKVLPQGMINSPTICQTFVGRALQVREFFSDCVI1HVDDILCAEFTDKLIDCVTFLO.ARVANGLAIASEKIQSTPFFHLYGQIENRRIKPKQIEIRKDLKINDPOLL

N8.4 \*  
 P1.1 \*  
 L4.4  
 H3.1  
 P1.8 \*  
 M3.5 \*  
 D1.3  
 D1.1  
 D1.2  
 P1.10  
 HP.1  
 M3.8  
 St.2 \*  
 HERVK55  
 HERVK27  
 HERVK67 \*  
 HERVK4  
 HM16  
 HERVK73  
 HM2.1  
 HM2.2 \*  
 HM2.3 \*  
 HM2.4 \*  
 HM2.5  
 G15.930 \*  
 RT16 \*  
 RT21  
 RT66 \*

Fig. 3.

Figs 2 and 3. For legends see facing page.

2.2 (U35105), HML-2.3 (U35106), HML-2.4 (U35107), HML-2.5 (U35158), HML-2.6 (U35152), RT 16 (U34348), RT 21 (U34394), RT 66 (U34395) from Medstrand & Blomberg (1993); T47D (U47118) from Patience *et al.* (1996); HERV-K10 (M14123) from Ono *et al.* (1986); HERV-(K)27, HERV-(K)67 and HERV-(K)73 from Li *et al.* (1996); HERV-(K)55 (U39936) from Li *et al.* (1995); clone 5:30 (S46404) from Medstrand *et al.* (1992); HM16 (M30520) from Deen & Sweet (1986).

## Results

### PCR amplification of novel HERV-K viruses belonging to the HML-2 group

Cellular RNA was isolated from leukaemic and normal bone marrow and peripheral blood samples. Total RNA was converted into cDNA by reverse transcription with a mixture of random hexamer primers and the AMV RT enzyme. In addition, we isolated DNA from human placenta tissue and the SupT1 T cell line. These cDNA and DNA samples were subsequently PCR-amplified with the MOPs-2 primer set that was designed to detect HERV-K10-like sequences (Li *et al.*, 1996). Most samples yielded products of the expected 650 bp size (including the PCR primers) on agarose gels. The PCR products were subcloned and multiple clones were sequenced for each RNA/DNA sample. A total of 25 HERV-K10-like RT sequences were identified. In this study, we will focus on those elements that belong to the HML-2 group (listed in Table 1). The other sequences are also HERV-K10-related and will be presented elsewhere (J. Zsíros & B. Berkhout, in preparation). Some sequences were obtained in duplicate (e.g. clones N8.4 and P1.2 are identical), and some sequences were identical to published elements [e.g. clones C1.3 and HP.3 are identical to each other and the published HM16 element (Deen & Sweet, 1986)]. Different HERV elements were obtained from individual cell samples (e.g. five different HERV elements were isolated from the bone marrow sample P1).

All clones amplified with the MOPs-2 primer set belong to the HERV-K family, consistent with the initial results of Li *et al.* (1996). Nucleotide sequence identity of the elements listed in Table 1 with the HERV-K10 prototype sequence varied between 88.7 to 99.2% (Fig. 1). Similarity among the new elements ranged from 85.2 to 99.6%. For sequences differing by only a few bases it is possible that the differences result from misincorporations during the initial reverse transcription reaction or the subsequent PCR amplification by *Taq* polymerase, which lacks proofreading exonuclease activity. To control for such possible errors, the following experiment was performed. Cells were transfected with the pLAI molecular

clone of HIV-1 and cellular RNA was used to amplify RT sequences with the same RT-PCR protocol. The cDNA products were cloned and six recombinant plasmids were sequenced. We found two mismatch mutations, indicating that on average 0.33 mutations are generated by the RT-PCR protocol for a 598 bp RT fragment. Thus, for our study of different HML-2 species, the clones differing by one nucleotide are likely to be generated from the same mRNA species, and are therefore considered to be identical.

Comparison of the nucleotide sequence of the new clones with published elements indicated the highest similarity with HERV-K10-like elements belonging to the HML-2 subgroup. These published HML-2 elements were obtained with a different PCR primer set that yields a significantly shorter RT fragment of only 244 bp. Thus, direct comparison of the new retroelements with these HML types is restricted to this small RT fragment. Within the small RT fragment, clones St.2 and P1.4 are identical to HML-2.2 and HML-2.5, respectively (Medstrand & Blomberg, 1993), and clones L4.4 and H3.1 are indistinguishable from the GenBank entry clone 5:30 (not shown). Obviously, we cannot exclude the possibility that these elements differ in sequences outside the 244 bp RT fragment. Most importantly, we identified ten novel members of the HML-2 subgroup. Nucleotide sequence identity between the new RT sequences and known members of the HML-2 subgroup is 84.2 to 98.8% for the small RT fragment. Clones N8.4, P1.1, P1.8, M3.5, M3.8 and HP.1 are mostly related to HML-2.1 with a similarity of 94.7 to 98.8%, P1.10 is closest to HML2.5 (98.8% similarity) and clones K1.1, D1.2 and D1.3 show the highest identity with HML-2.6, ranging from 92.7 to 98.8%.

The new HML-2 nucleotide sequences are compared with the prototype HERV-K10 in Fig. 2. The two clones P1.8 and M3.5 contain a nine nucleotide insert, and although the sequence of the inserts differs by two bases, the same AVN amino acid triplet is encoded (see Fig. 3). The insert forms an almost perfect repeat of the downstream nucleotide sequence, suggesting that it was generated by slippage of the RT enzyme during reverse transcription. No other insertions were detected, but several deletions were observed in individual clones. Two relatively extended deletions of 11 bp and 23 bp are present in clone D1.3, both resulting in frameshift mutations that interrupt the RT reading frame (see Fig. 3). Smaller deletions were observed in several clones. Comparison of the 5'- and 3'-most RT sequences did not provide evidence for recombination between different HML-2 members, although

**Fig. 2.** Nucleotide sequence alignment of new members of the HML-2 group. All sequences were compared with the prototype HERV-K10 sequence (top line) (Ono *et al.*, 1986). Dashes indicate base identities, dots indicate insertions and gaps indicate deletions in the sequences compared with HERV-K10.

**Fig. 3.** Alignment of the RT amino acid sequence of HML-2 endogenous retroviruses. The nucleotide sequences of Fig. 1 were translated and compared with the corresponding amino acid sequence of other HML-2 viruses. Dashes indicate identical amino acids compared with the prototype HERV-K10. Gaps are shown with dots. Stop codons are marked by an asterisk (\*) and frameshifts are indicated with the number 1 or 2, which indicates the number of remaining nucleotides in the corresponding codon. The conserved LPQG motif and the YIDD sequence corresponding to the catalytic core of the RT enzyme are overlined. Clones with an open reading frame are marked by an asterisk.

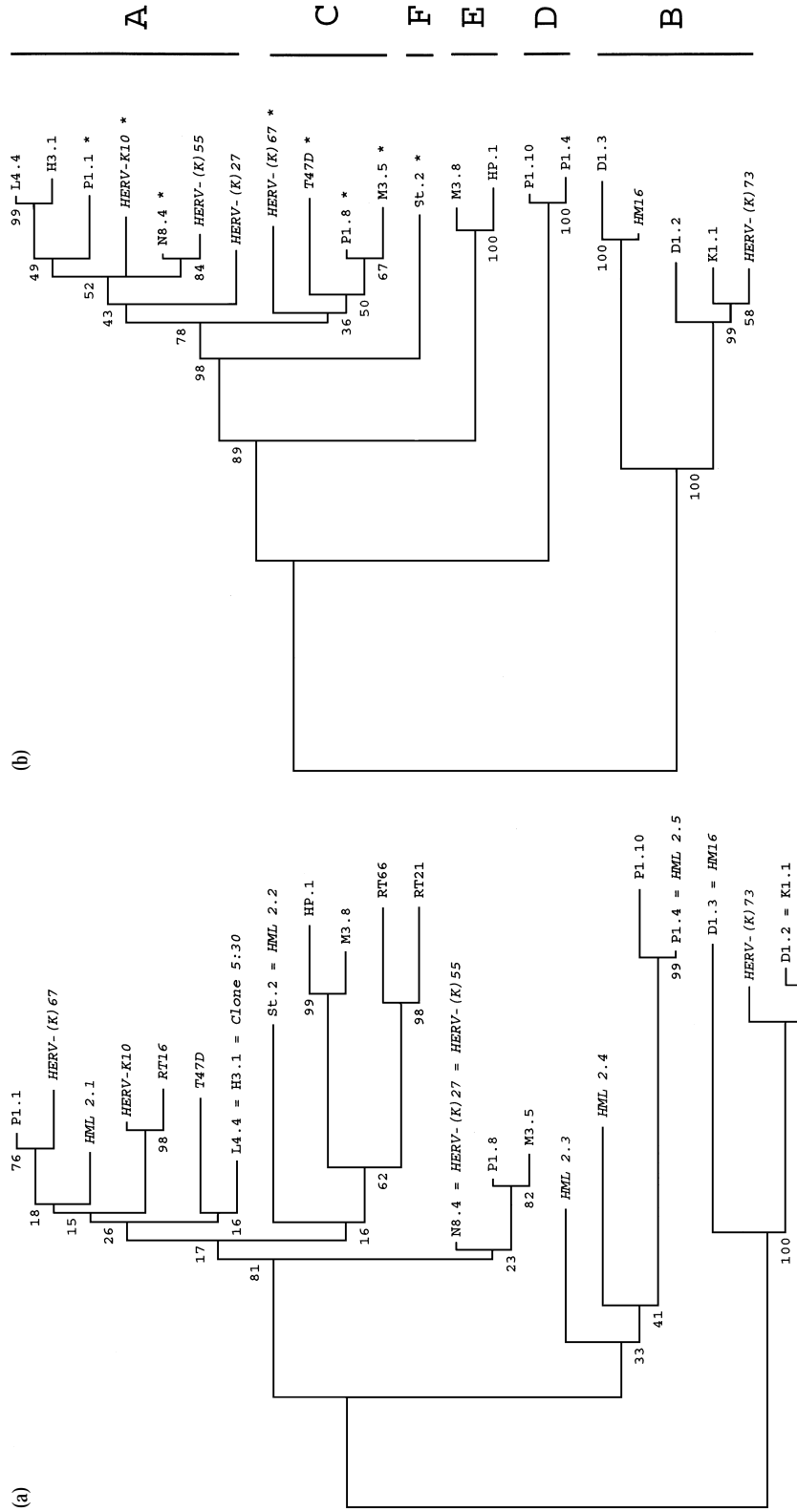
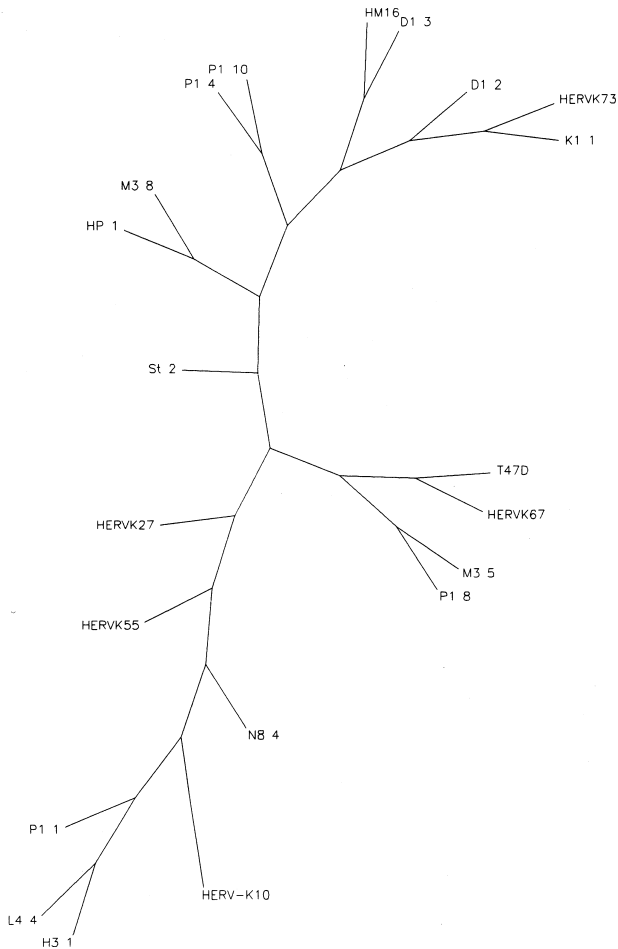


Fig. 4. Phylogenetic analysis of the HML-2 group by the neighbour-joining method. For this analysis we used either the short 244 bp or large 600 bp fragment of the RT gene (a and b, respectively). Sequences of published HERV-K10-like isolates are shown in italics; all other entries were identified in this study (see Table 1 for source of the samples). Branch lengths are proportional to the distances between the taxa. The values at the branch points indicate the percentage support for a particular node after 100 bootstrap replicates were performed.



**Fig. 5.** Phylogenetic analysis of the HML-2 group with the maximum parsimony method. Maximum parsimony tree of endogenous retroviral elements by alignment of a 600 bp RT fragment as shown in Fig. 2. The tree was constructed with the PHYLIP computer program (Felsenstein, 1989). Each new sequence is labelled with the clone name (Table 1).

this 600 bp fragment may be too small to detect such events.

We translated the putative RT genes and compared the amino acid sequences with that of other HML-2-related elements reported in the literature (Fig. 3). As discussed above, the nine nucleotide insertion in clones P1.8 and M3.5 results in duplication of the AVN triplet. Most deletions that were observed in the nucleotide sequence result in frameshifts. Five of the newly identified HML-2 elements contain an open reading frame for this part of the RT gene (these clones are marked by an asterisk in Fig. 3). Several conserved motifs that were previously used to design RT-specific primers can be recognized, e.g. the LPQG and YIDD elements (indicated in Fig. 3).

### Generation of a detailed phylogenetic tree for the HML-2 group

Phylogenetic trees were constructed with the neighbour-joining and maximum-parsimony methods. To study the

evolutionary relationships within this HML-2 group of retroelements, we analysed both the new and previously published HML-2 sequences. As noted above, most published HML-2 RT sequences are much smaller than the 600 bp fragment that was amplified in this study. We therefore generated phylogenies for two nucleotide sequence sets: one analysis included all previously identified HML-2 isolates and was performed on the 244 bp RT fragment (Fig. 4a); the second analysis was performed with a smaller number of sequences for which the 600 bp RT segment was available (Fig. 4b). The general topology of the trees obtained with the neighbour-joining method was similar for the small versus large RT fragments. For instance, in both cases subgroups containing the prototype HML-2.5 and HML-2.6 isolates are identified. However, detailed inspection of the two trees indicated some gross differences in the phylogeny of other clones. We think that the analysis of the larger RT fragment is more reliable for two reasons. First, this analysis provided much greater bootstrap values than the analysis of the small RT segment (compare panels A and B of Fig. 4). Second, the tree obtained with the parsimony analysis (Fig. 5) revealed the same overall topology as the distance analysis of the large RT fragment (Fig. 4B), testifying to the reliability of the outcome of the phylogenetic analyses.

We used the phylogeny of the 600 bp RT fragment for a detailed taxonomy of the HML-2 subgroup into eight branches as indicated in Fig. 4(b) (HML-2A to HML-2H). Most branches separate with high bootstrap values (98%) in the neighbour-joining trees of the larger RT fragment (Fig. 4b). Although the bootstrap value separating the A and C sequences is less impressive (78%), this branching is fully supported by the parsimony analysis (Fig. 5). The ranking order of the branches was determined arbitrarily by the number of isolates. For instance, the largest subgroup is HML-2A, which contains seven members including the prototype HERV-K10 element. Three subgroups with a single member were recognized. The HML-2F subgroup consists of the St.2 element of which the 600 bp RT fragment was identified in this study, but note that this clone is identical to the published HML-2.2 clone over the 244 bp RT fragment. Subgroups HML-2G and HML-2H correspond to clones for which only small RT fragments were reported: sequences HML-2.3 and HML-2.4, respectively.

In order to determine the evolutionary forces that could generate the sequence heterogeneity among the different members of the HML-2 group, we analysed separately the synonymous and nonsynonymous substitutions within the larger RT fragment. This analysis was performed for all sequences shown in Fig. 4(b), as well as for the subset that contains an RT open reading frame. The mean synonymous and nonsynonymous variation within a particular HML-2 subgroup was calculated as the mean of the synonymous and nonsynonymous distances between individual sequences belonging to this subgroup. To calculate the mean synonymous and nonsynonymous distances between two separate HML-2



	A	C	F	E	D	B
A	6.6	3.4	3.1	4.7	3.7	7.1
C		2.4	1.8	3.2	2.7	5.9
F			-	3.1	3.1	6.1
E				1.8	2.8	5.3
D					-	3.7
B						3.2

Fig. 6. Ratios of synonymous to nonsynonymous substitutions ( $d_s/d_a$ ) in HERV-K10-related sequences. The analysis was performed on the approximately 600 bp RT fragment of the HML-2 subgroups shown in Fig. 4 (b).

subgroups, each sequence belonging to one subgroup was compared with each sequence from the other subgroup. Interestingly, most sequence variation within and between HML-2 subgroups was due to synonymous nucleotide substitutions. For instance, the mean synonymous nucleotide distances between the most diverged HML-2 subgroup B and other HML-2 subgroups ranged from 25.1 to 30.5%, whereas the mean nonsynonymous nucleotide distances ranged from 4.2 to 6.8% (data not shown). The  $d_s/d_a$  ratios within and between subgroups ranged from 1.8 to 7.1 (Fig. 6). These results indicate that purifying selection is or was operating on this RT gene segment of these endogenous retroviruses. Furthermore, the topologies of the neighbour-joining phylogenetic trees obtained with these two sets of codon substitutions were similar to each other and matched the structure of the tree shown in Fig. 4 (b) (data not shown).

## Discussion

The MOPs-2 primer set was used to amplify an approximately 600 bp RT fragment of HERV-K10-related retroviral sequences. In this study, we identified ten new members of the HML-2 subgroup. The current HML-2 subgroup consists of 27 members (Fig. 4), which is considerably larger than anticipated from the approximately 15 signals that were observed with HML-2-specific probes on Southern blots (Medstrand & Blomberg, 1993). It is likely that the Southern technique did not detect all HML-2 members as individual hybridization signals because of the conservation of restriction sites among different HML-2 sequences. The phylogenies presented in Figs 4 (b) and 5 represent a detailed picture of this HML-2 subgroup of retroelements. We recognized eight subgroups that range in size from seven members (HML-2A) to the single member subgroups HML-2F, HML-2G and HML-2H. In general, the topology of the trees obtained by the distance and parsimony methods are similar, testifying to the reliability of this organization.

Among the 21 HML-2 isolates for which a 600 bp RT

fragment was obtained, eight sequences have an open reading frame. Interestingly, these putative RT-encoding clones are clustered in particular subgroups. Specifically, 8 of the 12 clones in groups HML-2A, HML-2C and HML-2F have an open RT reading frame (marked by an asterisk in Fig. 4 b). Analysis of full-length clones is under way to test whether these elements encode a functional RT enzyme. The original HERV-K10 isolate encodes a full-length RT protein, but no activity could be detected with recombinant forms of this protein [mentioned as unpublished data in the review article by Lower *et al.* (1996)]. Although these endogenous RT enzymes may require special assay conditions for activity, it is also possible that the HERV-K10 RT contains some inactivating mutations. Fig. 3 shows that new HML-2 elements differ at several amino acid positions from the prototype HERV-K10 sequence. For instance, comparison with other RT sequences (Xiong & Eickbush, 1990) reveals two relatively well-conserved amino acids at position 9 and 10 in all clones, except for the HERV-K10 prototype (Fig. 3: dipeptide HT, which is RM in most other isolates). To test this possibility, full-length HML-2 RT genes should be cloned and assayed for enzyme activity. RT activity has been recognized within virion particles of HERV-K-related endogenous retroviruses that are frequently observed in the placenta of primates and these particles were recently demonstrated to package HML-3 and HML-6 RNA sequences (Simpson *et al.*, 1996). This suggests that at least some members of the HML-3 and HML-6 groups encode a functional RT protein, although enzyme activity was not yet demonstrated for one of these clones (Simpson *et al.*, 1996). Given the transcriptional activity and complete RT open reading frame in several HERV-K elements, it remains possible that these elements are still active as intracellular retrotransposons.

Analysis of the synonymous and nonsynonymous mutation rates revealed a significant bias for the synonymous or silent codon changes ( $d_s/d_a$  ratios ranged from 1.8 to 7.1). Ratios > 1 were also measured for the equivalent RT region of HIV-1, which is an actively replicating, exogenous retrovirus. The strong prevalence of the synonymous nucleotide substitutions indicates that the genetic variation observed within the HML-2 group is for a large part the result of purifying (negative) selection. The target for this selection may have been an exogenously replicating ancestor of the HML-2 virus. Alternatively, the high  $d_s/d_a$  ratios may suggest that the contemporary, endogenized HML-2 retrovirus is still under selective pressure to maintain replicative capacity, leaving open the possibility that these elements still exhibit retrotransposition activity.

The idea that HML-2 elements have spread throughout the human genome by intracellular retrotransposition would predict a high similarity between the integrated proviruses because of a direct mother–daughter relationship. Instead, the phylogenies presented in this study (Figs 4 b and 5) do not form a continuous spectrum of variants, but rather form

clusters of related sequence elements. Another argument against spread via retrotransposition is the absence of recombination within the RT fragments, although such an analysis should be repeated with full-length proviral elements. These results may suggest that the current group did not result from intracellular amplification, although it cannot be excluded that many differences between the HML-2 elements are due to somatic mutations that occurred after provirus integration. Thus, the current HML-2 group is likely to represent independent introductions into the germline by infections with different members of an exogenous quasispecies of infectious retrovirus that existed many millions of years ago. An intriguing possibility is that the endogenized HML-2 proviruses provided the host cell with a mechanism to become immune to new infections. There is accumulating evidence that endogenous retroviral sequences can protect cells from infection by a related autonomous retrovirus (Craig *et al.*, 1991; Robinson *et al.*, 1981; Rubin, 1960; Best *et al.*, 1996; Goff, 1996; Coffin, 1996).

A variety of cell samples derived from children with leukaemia or with normal bone marrow were used in this study for the amplification of HERV-K10-like sequences. We note that the cause of childhood leukaemia is unknown for most cases. There is some epidemiological evidence that an infectious agent is involved (reviewed in Greaves & Alexander, 1993), but as yet no definite, or even candidate, agents have been identified. It has been suggested that an infective agent is involved in the aetiology of both acute lymphoblastic leukaemia (ALL) of childhood and ALL occurring at other ages (Greaves & Alexander, 1993; Mariani-Costantini *et al.*, 1989). The possible existence of an infectious agent in human leukaemia was reinforced by the discovery that mouse mammary tumour virus is an aetiological agent of murine mammary tumours and that retroviruses cause leukaemia in chickens as well as other animal species (reviewed in Weiss, 1985). Also, an immune response to HERV-K proteins was detected in human patients with certain tumours (Vogetseder *et al.*, 1993; Sauter *et al.*, 1995). Obviously, the number of samples analysed in this study is too small to evaluate whether the expression of certain HERV-K elements is altered in leukaemic bone marrow or peripheral blood.

**Note added in proof.** Recently, the HERV-K10 family of human endogenous retroviral elements was implicated in the induction of the autoimmune disease type 1 diabetes through expression of a superantigen domain within the viral Env protein [B. Conrad, R. N. Weissmahr, J. Böni, R. Arcari, J. Schupbach & B. Mach (1996). A human endogenous retroviral superantigen as candidate autoimmune gene in type 1 diabetes. *Cell* **90**, 303–313].

We thank Dr J. C. Kluin-Nelemans for the hairy cell leukaemia sample. This research was sponsored by the 'Stichting Kindergeneeskundig Kankeronderzoek' (SKK).

## References

- Best, S., Letissier, P., Towers, G. & Stoye, J. P. (1996). Positional cloning of the mouse retrovirus restriction gene Fv1. *Nature* **382**, 826–829.
- Chirgwin, J. M., Przybyla, A. E., McDonald, R. J. & Rutter, W. J. (1979). Isolation and biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**, 5294–5299.
- Coffin, J. M. (1996). Retrovirus restriction revealed. *Nature* **382**, 762–763.
- Craig, L. C., Pirtle, I. L., Gracy, R. W. & Pirtle, R. M. (1991). Characterization of the transcription unit and two processed pseudogenes of chimpanzee triosephosphate isomerase (TPI). *Gene* **99**, 217–227.
- Das, A. T., Klaver, B. & Berkhout, B. (1995). Reduced replication of human immunodeficiency virus type 1 mutants that use reverse transcription primers other than the natural tRNA(3Lys). *Journal of Virology* **69**, 3090–3097.
- Deen, K. C. & Sweet, R. W. (1986). Murine mammary tumor virus pol-related sequences in human DNA: characterization and sequence comparison with the complete murine mammary tumor virus pol gene. *Journal of Virology* **57**, 422–432.
- Felsenstein, J. (1989). PHYLIP—phylogeny inference package (version 3.2). *Cladistics* **5**, 164–166.
- Felsenstein, J. (1990). PHYLIP manual 3.2.
- Franklin, G. C., Chretien, S., Hanson, I. M., Rochefort, H., May, F. E. B. & Westley, B. R. (1988). Expression of human sequences related to those of mouse mammary tumor virus. *Journal of Virology* **62**, 1203–1210.
- Goff, S. P. (1996). Operating under a Gag order: a block against incoming virus by the Fv1 gene. *Cell* **86**, 691–693.
- Greaves, M. F. & Alexander, F. E. (1993). An infectious etiology for common acute lymphoblastic leukemia in childhood? *Leukemia* **7**, 349–360.
- Kumar, S., Tamura, K. & Nei, M. (1993). Molecular evolutionary genetics analysis (MEGA). Pennsylvania State University, University Park, Penn., USA.
- Li, M. D., Bronson, D. L., Lemke, T. D. & Faras, A. J. (1995). Restricted expression of new HERV-K members in human teratocarcinoma cells. *Virology* **208**, 733–741.
- Li, M. D., Lemke, T. D., Bronson, D. L. & Faras, A. J. (1996). Synthesis and analysis of a 640-bp pol region of novel human endogenous retroviral sequences and their evolutionary relationships. *Virology* **217**, 1–10.
- 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.
- Lukashov, V. V., Kuiken, C. L. & Goudsmit, J. (1995). Intrahost human immunodeficiency virus type 1 evolution is related to length of the immunocompetent period. *Journal of Virology* **69**, 6911–6916.
- Mariani-Costantini, R., Horn, T. & Callahan, R. (1989). Ancestry of a human endogenous retrovirus family. *Journal of Virology* **63**, 4982–4985.
- Medstrand, P. & Blomberg, J. (1993). Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: differential transcription in normal human tissues. *Journal of Virology* **67**, 6778–6787.
- Medstrand, P., Linderskog, M. & Blomberg, J. (1992). Expression of human endogenous retroviral sequences in peripheral blood mononuclear cells of healthy individuals. *Journal of General Virology* **73**, 2463–2466.

- Ono, M., Yasunaga, T., Miyata, T. & Ushikubo, H. (1986).** Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. *Journal of Virology* **60**, 589–598.
- Patience, C., Simpson, G. R., Colletta, A. A., Welch, H. M., Weiss, R. A. & Boyd, M. T. (1996).** Human endogenous retrovirus expression and reverse transcriptase activity in the T47D mammary carcinoma cell line. *Journal of Virology* **70**, 2654–2657.
- Robinson, H. L., Astrin, S. M., Senior, A. M. & Salazar, F. H. (1981).** Host susceptibility to endogenous viruses: defective, glycoprotein-expressing proviruses interfere with infections. *Journal of Virology* **40**, 745–751.
- Rubin, H. (1960).** A virus in chick embryos which induces resistance in vitro to infection with Rous sarcoma virus. *Proceedings of the National Academy of Sciences, USA* **46**, 1105–1119.
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989).** *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sauter, M., Schommer, S., Kremmer, E., Remberger, K., Dolken, G., Lemm, I., Buck, M., Best, B., Neumann-Haefelin, D. & Mueller-Lantzsch, N. (1995).** Human endogenous retrovirus K10: expression of Gag protein and detection of antibodies in patients with seminomas. *Journal of Virology* **69**, 414–421.
- Simpson, G. R., Patience, C., Lower, R., Tonjes, R. R., Moore, H. D. M., Weiss, R. A. & Boyd, M. T. (1996).** Endogenous D-type (HERV-K) related sequences are packaged into retroviral particles in the placenta and possess open reading frames for reverse transcriptase. *Virology* **222**, 451–456.
- Smith, S. D., Shatsky, M., Cohen, P. S., Warnke, R., Link, M. P. & Glader, B. E. (1984).** Monoclonal antibody and enzymatic profiles of human malignant T-lymphoid cells and derived cell lines. *Cancer Research* **44**, 5657–5662.
- Vogetseder, W., Dumfahrt, A., Mayersbach, P., Schonitzer, D. & Dierich, M. P. (1993).** Antibodies in human sera recognizing a recombinant outer membrane protein encoded by the envelope gene of the human endogenous retrovirus K. *AIDS Research and Human Retroviruses* **9**, 687–694.
- Weiss, R. (1985).** The search for human RNA tumor viruses. In *RNA Tumor Viruses*, 2nd edn, pp. 1205–1281. Edited by R. Weiss, N. Teich, H. Varmus & J. Coffin. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Wilkinson, D. A., Mager, D. L. & Leong, J.-A. (1994).** Endogenous human retroviruses. In *The Retroviridae*, pp. 465–535. Edited by J. A. Levy. New York: Plenum Press.
- Xiong, Y. & Eickbush, T. H. (1990).** Origin and evolution of retro elements based upon their reverse transcriptase sequences. *EMBO Journal* **9**, 3353–3362.

---

Received 15 May 1997; Accepted 4 September 1997