J Mol Evol (2001) 53:237–243
DOI: 10.1007/s002390010213

JOURNAL OF MOLECULAR EVOLUTION

Evolutionary Dynamics of the Human Endogenous Retrovirus Family HERV-K Inferred from Full-Length Proviral Genomes

Javier Costas

Departamento de Bioloxía Fundamental, Facultade de Bioloxía, Universidade de Santiago de Compostela, Campus Sur s/n E-15782 Santiago de Compostela, (A Coruña), Spain

Received: 5 February 2001 / Accepted: 22 March 2001

Abstract. Several distinct families of endogenous retroviruses exist in the genomes of primates. Most of them are remnants of ancient germ-line infections. The human endogenous retrovirus family HERV-K represents the unique known case of endogenous retrovirus that amplified in the human genome after the divergence of human and chimpanzee lineages. There are two types of HERV-K proviral genomes differing by the presence or absence of 292 bp in the *pol-env* boundary. Human-specific insertions exist for both types. The analyses shown in the present work reveal that several lineages of type 1 and type 2 HERV-K proviruses remained transpositionally active after the human/chimpanzee split. The data also reflect the important role of mosaic evolution (either by recombination or gene conversion) during the evolutionary history of HERV-K.

Key words: Human endogenous retrovirus — HERV — HML-2 — HERV-K — Master gene model — Retrovirus-like elements — Transposable elements — Interspersed elements — Mosaic evolution

Introduction

The human genome harbors a wide variety of endogenous retroviruses, representing at least 5% of total DNA (Smit 1996). They most likely stem from germ-line retroviral integrations at different times during primate evolution. Presumably, subsequent retrotransposition (although re-infection cannot be formally ruled out) often led to an increase in copy number (Löwer et al. 1996). The vast majority of these insertions persist within the human genome as solitary long terminal repeats (LTRs), created by homologous recombination between the 5′ and 3′ LTRs of an intact proviral element (Löwer et al. 1996).

HERV-K, also referred to as HML-2, is the biologically most active human endogenous retrovirus (HERV) family, retaining the capacity to be expressed at the RNA and protein levels, and to form virus-like particles (Löwer et al. 1993). A few proviruses have been detected preserving long open reading frames (ORFs) for several proteins (Barbulescu et al. 1999; Mayer et al. 1999; Tönjes et al. 1999). DNA hybridization data have suggested that HERV-K first entered the primate genome shortly after the split of New World and Old World monkeys (Mariani-Costantini et al. 1989). Phylogenetic analyses of HERV-K LTR sequences revealed the existence of distinct subgroups with different integration times, determined by locus-specific PCR of several members from each subgroup (Medstrand and Mager 1998). Remarkably, this and other studies have shown that the most recent insertions of HERV-K post-date the human/chimpanzee split, representing the unique known family of HERVs that include human-specific insertions (Medstrand and Mager 1998; Barbulescu et al. 1999; Mayer et al. 1999; Lebedev et al. 2000). The hallmark of this long period of activity is the existence of several thousand

Correspondence to: Javier Costas; *email:* bfcostas@usc.es

solitary HERV-K LTRs within the human genome (Leib-Mösch et al. 1993).

The existence of a correlation between sequence divergence of distinct HERV-K LTR subgroups and their time of integration (i.e., greater divergence between members of older subgroups) clearly supports an expansion of HERV-K elements similar to the "master gene" model (Medstrand and Mager 1998). This model proposes that most copies of a family of elements are produced by only one or a few source genes (the "master genes") at any time during evolution of the family. Accumulation of substitutions in the "master genes" or replacement of these "masters" by others from their own derivatives, gives rise to distinct subfamilies of pseudogene copies during the evolution of the element (Deininger et al. 1992). Although primarily applied to the evolution of mammalian SINEs (mainly, the primate *Alu* element) and LINEs (mainly, L1 in rodents and primates), the "master gene" model has been also suggested in the case of another mammalian retrovirus-like element, *mys* of the genus *Peromyscus* (Clough et al. 1996; Lee et al. 1996).

There are two types of HERV-K proviruses (Ono et al. 1986). Type 2 genomes supposedly more closely resembled the structure of the infectious retrovirus ancestor of the HERV-K family. Type 1, characterized by a 292 bp deletion within the *pol-env* boundary, has been detected only in the genomes of human, chimpanzee, gorilla, and orangutan but not in Old World primates, indicating that the mutational event giving rise to type 1 proviral genome occurred after the split of Old World primates and great apes (Mayer et al. 1998). Interestingly, human-specific insertions of both types of proviruses have been reported (Barbulescu et al. 1999). These facts suggest the coexistence of two active lineages of HERV-K genomes for a considerable period of time.

In order to get a better insight into the evolution of these two types of HERV-K proviruses, I analyzed the phylogenetic relationships of a significant portion of the almost intact full-length HERV-K elements present within the human genome. These analyses revealed several departures from a simple model of HERV-K evolution driven by two lineages, corresponding to type 1 and type 2 proviruses, since the appearance of the type 1. Instead, they support the existence of various lineages of HERV-K proviruses capable of transposition after the human/chimpanzee split. Besides, the data illustrate the important contribution of mosaic evolution (either by recombination or gene conversion) to the history of HERVs.

Material and Methods

Searches of full-length (not truncated) HERV-K proviruses were performed by screening the working draft sequence of the human genome and the GenBank nr and HTGs databases at the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) by using the BLASTN program (Altschul et al. 1990). Sequences HERV-K(C7), as representative of type 2 proviruses (GenBank AC: Y17832, Tönjes et al. 1999), and HERV-K10, as a type 1 provirus (GenBank AC: M14123, Ono et al. 1986) were the queries in the screenings. Duplicate HERV-K proviruses were eliminated on the basis of flanking sequences. ORF finder at NCBI was used to detect ORFs in the new identified proviruses.

All the full-length HERV-K nucleotide sequences were aligned with the help of the ClustalX program (Thompson et al. 1997) and slightly adjusted using the GeneDoc program (Nicholas and Nicholas 1997). ClustalX was also employed to infer phylogenetic relationships among sequences by the neighbor-joining method (Saitou and Nei 1987), and to obtain bootstrap values for each internal branch (1000 replicates). All the positions with gaps in one or more sequences were eliminated in these analyses, as well as the unique position with an undetermined nucleotide. Analysis of mosaicism was performed by a variant of the maximum χ^2 method (Maynard Smith 1992), as described by Robertson et al. (1995), based on the distribution of phylogenetic informative sites supporting the clustering of the putative mosaic sequence with each one of the putative "parental" sequences.

The DnaSP program (Rozas and Rozas 1999) was used to calculate nucleotide diversity (π); the number of synonymous substitutions per synonymous site (Ks) and the number of non-synonymous substitutions per non-synonymous site (Ka), using the method of Nei and Gojobori (1986); and the average number of nucleotide substitutions per site (Dxy) and the number of net nucleotide substitutions per site (Da), according to Nei (1987). In all cases, the Jukes and Cantor's correction was applied. Standard errors of Table 2, implemented with MEGA v.2 (Kumar et al. 2000), were calculated using the bootstrap method (500 replicates).

Results

Table 1 lists the 11 non-truncated proviruses identified in the GenBank database (as of December 1, 2000) as well as three others with almost intact full-length internal sequences but partially deleted LTRs, representing a total of seven type 1 and seven type 2 proviruses. Nine of them were previously sequenced after screenings of human genomic libraries to select full-length proviruses with long ORFs (Barbulescu et al. 1999; Mayer et al. 1999; Tönjes et al. 1999). The other five proviruses were sequenced for the first time by the Human Genome Project. HERV-K(C12) encodes full-length Gag and Env proteins; and HERV-K(C11c) encodes a full-length Pol and almost full-length Env proteins. HERV-K(C11a), HERV-K(C11b), and HERV-K(C3) present several stop codons and frameshift mutations, resulting in the shortening of their ORFs. On the basis of identity of the flanking sequences, several previously described insertions (HERV-K101, HERV-K102, HERV-103, and HERV-K109) were assigned to specific chromosomes (Table 1).

The alignment of these 14 sequences revealed the existence of 283 nucleotide positions (of 9472) at which a nucleotide difference from the consensus is shared by at least two elements, after exclusion of gaps (Fig. 1). Besides, there are 732 individual differences. In order to clarify the phylogenetic relationships between all the proviruses, a neighbor-joining tree of the internal region

Table 1. Full-length proviruses analyzed in the present work

Provirus[a]	GenBank	Position[b]	Duplicates	Bibliography names	LTRs difference (%)[g]	Genome type[h]
HERV-K(C1a)	AL121985	78059	AC068728	HERV-K18[c]	34 (3.51)	1
			Y18890	HERV-K110[d]		
HERV-K104	AF164612			HERV-K104[d]	16 (1.68)	2
HERV-K(C7)	AF164614		Y17832	HERV-K(C7)[c]	10 (1.03)	2
			AC072054	HERV-K108[d]		
				HERV-K108[d]		
				HERV-K(HML.2-HOM)[e]		
HERV-K(C1b)	AC044819	−130901	AL353807	HERV-K102[d]	3 (0.31)	1
			AF164610			
HERV-K(C5)	AC016577	−30110	M14123	HERV-K10[f]	2 (0.21)	1
			AC008535	HERV-K107[d]		
			AF164613			
HERV-K(C10)	AF164611		AL139404	HERV-K103[d]	7 (0.72)	1
HERV-K(C22)	AC007326	36875	AF164609	HERV-K101[d]	4 (0.41)	1
HERV-K(C6)	AF164615		AC055116	HERV-K109[d]	3 (0.31)	2
HERV-K(C19)	Y17833			HERV-K(C19)[c]	N.A.	2
HERV-K(C3)	AC024108	130739	AC021655	Present work	N.A.	1
			AC078785			
HERV-K(C11a)	AC015686	157484		Present work	30 (3.10)	2
HERV-K(C11b)	AP000831	−15315		Present work	5 (0.52)	1
HERV-K(C11c)	AP000776	34849		Present work	N.A.	2
HERV-K(C12)	AC074261	87074		Present work	18 (1.86)	2

[a] Proviruses are identified by its chromosomal assignment, followed by a letter if there are more than one provirus in the same chromosome. The only exception is HERV-K104, whose chromosomal assignment was not possible.

[b] Numbers correspond to the position of the first nucleotide of the 5′ LTR. They are only indicated in the case of sequences from the Human Genome Project because the other GenBank accession numbers correspond with specific HERV-K sequences. A minus sign indicates sequence orientation opposite to the provirus. Bibliography names according to [c] Tönjes et al. 1999; [d] Barbulescu et al. 1999; [e] Mayer et al. 1999; [f] Ono et al. 1986.

[g] Nucleotide differences (percentages in parentheses) between the 5′ and 3′ LTRs from the same provirus.

N.A.: not applicable. HERV-K(C19), HERV-K(C3), and HERV-K(C11c) begin at positions 946, 521, and 486, respectively. HERV-K(C3) finishes at position 8913.

[h] Based on the 292 bp deletion at the *pol-env* boundary.

(i.e., that flanked by the LTRs) was constructed. Taking into account that the two LTRs of a provirus are identical when the element first inserts into the host genome, an independent analysis of the LTRs may render additional information (see below). As shown in Fig. 2, the great majority of sequences are clustered into a main group, while HERV-K104 is placed at an intermediate position between this group and the more divergent sequences HERV-K(C11a), HERV-K(C11b), and HERV-K(C1a). The fact that HERV-K(C1a) and HERV-K(C11b) do not cluster with the other type 1 proviruses constitutes a strong indication of mosaicism, either by a recombination of gene conversion event around the 292 bp deletion characteristic of type 1 proviral genomes. Within the main group of sequences, consisting of five type 1 and five type 2 proviruses, those of type 1 are clustered together. The nucleotide diversity value (π) for these five type 1 proviruses is 0.01056 ± 0.00154; and 0.01181 ± 0.00172 for the five type 2 proviruses of the main cluster. The average number of nucleotide substitutions per site between these two groups (Dxy) is 0.01218 ± 0.00326; while the number of net nucleotide substitutions per site between the two groups (Da) is only 0.00100 ± 0.00340.

Comparison of synonymous (Ks) versus non-synonymous (Ka) substitutions per site can provide useful information into the nature of the forces acting on the HERV-K coding regions. Average values of these variables, as well as its ratio (Ks/Ka), were calculated for each of the different ORFs of HERV-K, between all members of the main cluster, as a whole or divided in type 1 and type 2 proviruses, and between this main cluster and the other elements (Table 2). In the great majority of comparisons, Ks was significantly higher than Ka.

As a preliminary analysis to search for other cases of mosaic evolution, in addition to that involving the 292 bp deletion, I reconstructed phylogenetic trees from the different HERV-K genes (data not shown). This approach revealed the mosaic nature of HERV-K(C1a), that clearly groups with HERV-K(C11a) in the *gag* tree (bootstrap value of 100%), and with HERV-K(C11b) in the *env* tree (bootstrap value of 87%). To more precisely localize the recombination break point in HERV-K(C1a), I analysed the distribution of informative sites supporting its cluster with HERV-K(C11a) or HERV-K(C11b) along the internal region. The location of the recombinational break point in any position between nucleotide 4158 and 4189 maximizes the 2×2 χ^2 value of this distribution. The ratio of sites supporting the cluster of HERV-K(C1a) with HERV-K(C11a) and with HERV-K(C11b) are 18:3 in the region between the end of the 5′ LTR and position 4158/4189, and 6:12 from this point to

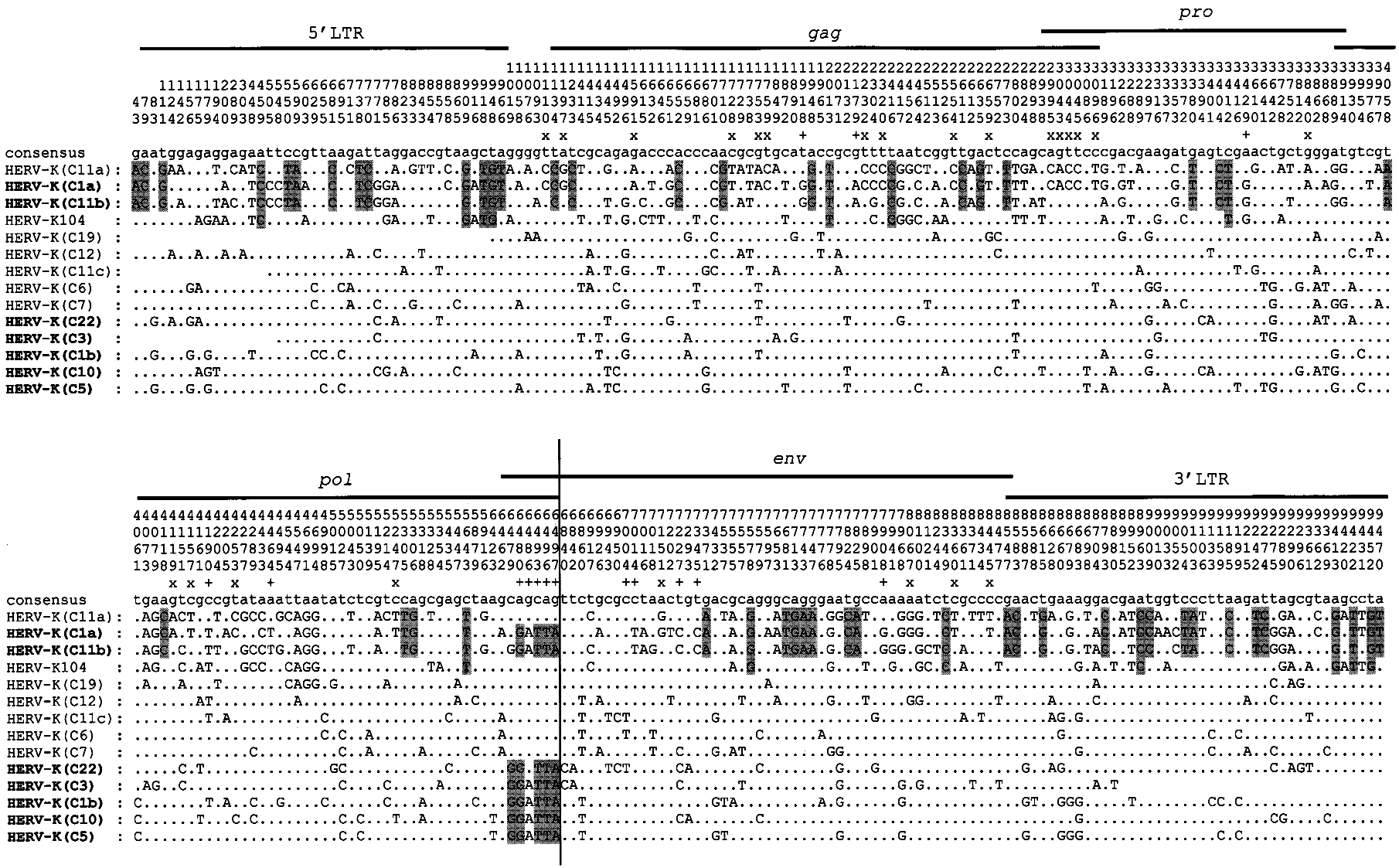

Fig. 1. Nucleotide polymorphisms shared by at least two sequences. Dots indicate identity with the consensus sequence. The positions are numbered across the top, referred to HERV-K(C7), a canonical type 2 provirus (9472 bp). Different regions of the proviral genome are indicated by horizontal lines. The very short *K-rev* gene is not shown. The vertical line indicates the location of the 292 bp deletion that distinguishes the two types of proviruses. Names of type 1 proviruses are in boldface. Positions characteristic of different groups (see text) are outlined in grey boxes. Sites supporting the cluster of HERV-K(C1a) with HERV-K(C11a) or with HERV-K(C11b) are indicated above the sequences as × or +, respectively (see text).

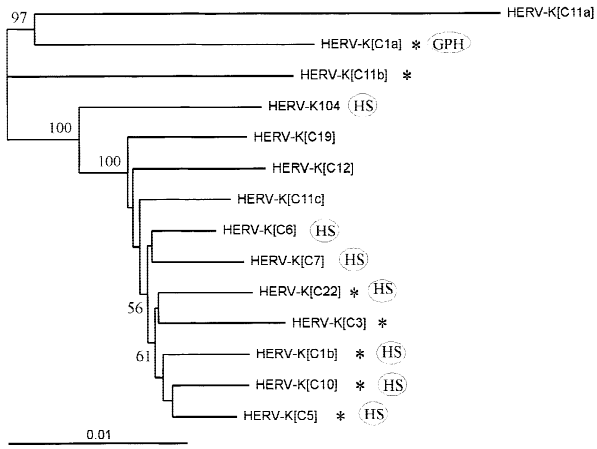

Fig. 2. Neighbor-joining tree of the full-length HERV-K proviruses based on internal nucleotide sequences. The value at a particular node indicates its percentage of appearance in 1000 bootstrap replicates. Only values above 50% are shown. An asterisk (*) identifies type 1 proviruses. HS and GPH represent human-specific insertions and insertions present in human, chimpanzee, and gorilla but not orangutan, respectively, as determined by Barbulescu et al. (1999).

the beginning of the 3′ LTR (Fig. 1). These ratios differ in a highly significant fashion ($\chi^2 = 11.236$, 1 d.f., $p < 0.001$).

To get further insight into the evolution of HERV-K, a neighbor-joining tree was constructed based on the LTRs of the detected proviruses (Fig. 3). This tree corroborated the existence of the same cluster of sequences revealed by the internal region tree, and HERV-K104 is located again in an intermediate place. Nevertheless, the type 1 proviruses of the main group do not form a cluster. It is interesting to note that, in all cases, the two LTRs of the same provirus are grouped together.

Discussion

The phylogenetic analyses of full-length HERV-K proviruses shown in the present work clearly revealed a main discrepancy with a simple model of HERV-K evolution driven by two lineages, corresponding to type 1 and type 2 proviruses, since the appearance of the type 1 preceding the split of the orangutan lineage from the other great apes (Mayer et al. 1998). Thus, under this assumption, we should expect a cluster formed by all the type 1 sequences in the phylogenetic trees, but this is not the case (Figs. 2 and 3). Taking into account that the analyses of mosaicism strongly supported the chimerical structure of the HERV-K(C1a) genome (the diagnostic 292 bp deletion of this provirus was acquired from HERV-K(C11b) or a similar provirus), this discrepancy may be explained by a unique gene conversion or recombination event in either an evolutionary forward or backward direction. In the first case, a young type 2 provirus should have acquired the region around the 292 bp deletion from an HERV-K(C11b)-like genome, giving rise to the young type 1 proviruses. In the second case, an old type 2 provirus should have acquired this diagnostic region from a young type 1 provirus, originating the HERV-K(C11b)-like genome. This option requires the appearance of young type 1 proviruses prior to the split of the orangutan lineage, around 14 Mya (Goodman et al. 1998). The average Ks value for the different ORFs between type I proviruses from the main cluster can be used to determine the approximate age of their most recent common ancestor. Using an average synonymous substitution rate in the primate lineage of 2.2 per site, per 10^9 years (Bulmer et al. 1991), this age has been calculated as 4.3 Myr. Although this value must be considered as a rough approximation, due mainly to the variation in synonymous substitution rate among genes (Matassi et al. 1999), it clearly suggests a more recent origin of young type 1 proviruses than the 14 Mya required in the case of a backward gene conversion/recombination event. Both this event involving the type 1/type 2 diagnostic region and the chimerical structure of the HERV-K(C1a) genome strongly strengthen the important role played by mosaic evolution in the evolutionary history of the different HERV families as described previously in the cases of HERV-H (Mager 1989) and ERV9 (Costas and Naveira 2000). We must take into account that retrovirus-like elements are expected to be especially prone to genetic rearrangements due to the possibility of recombination between two RNA genomes packaged within the same capsid (McDonald 1993), as in the case of exogenous retroviruses (Hu and Temin 1990).

The “master gene” model proposes that a family of elements evolve by successive replacement of existing “master genes” by novel ones, giving rise to different subgroups of elements characterized by the sequential accumulation of nucleotide differences (Deininger et al. 1992). My results suggest a similar situation until the split of the main cluster of sequences. Thus, there are 54 fixed differences between the main cluster and the more divergent ones, HERV-K(C11a), HERV-K(C11b), and HERV-K(C1a), along the 9472 bp of the alignment. HERV-K104, located at an intermediate position in the phylogenetic trees, shares 14 characteristic nucleotide positions with these three sequences and the other 40 differences with those from the main cluster (Fig. 1). Nevertheless, my data revealed several incongruities with HERV-K evolution through continuous replacement over time of the “master genes” leading to human-specific type 1 or type 2 proviruses since the split of the main cluster. First, in addition to the 292 bp diagnostic deletion, there are only five nucleotide differences in the human-specific insertions that distinguish type 1 and type 2 proviruses (Fig. 1). Furthermore, these five differences (located between positions 6480 and 6497) are closely linked to the 292 bp deletion (beginning at position 6501). Second, under the above scenario, we expect a higher level of divergence between the young type 1

Table 2. Synonymous and non-synonymous substitutions per site between HERV-K insertions, for the different ORFs of the proviruses[a]

Group	*gag*	*pro*	*pol*	*env*	*K-rev*[b]
Main cluster					
Ks	0.01652 (0.00338)	0.01794 (0.00433)	0.02227 (0.00366)	0.02418 (0.00413)	N.A.
Ka	0.00801 (0.00129)	0.01342 (0.00235)	0.00852 (0.00111)	0.01164 (0.00166)	N.A.
Ks/Ka	2.06**	1.34	2.61***	2.08**	
Main cluster type 1					
Ks	0.01564 (0.00390)	0.02441 (0.00642)	0.01867 (0.00370)	0.01813 (0.00434)	N.A.
Ka	0.00712 (0.00151)	0.01285 (0.00282)	0.00674 (0.00120)	0.00963 (0.00158)	N.A.
Ks/Ka	2.19*	1.90*	2.77**	1.88*	
Main cluster type 2					
Ks	0.01865 (0.00433)	0.01375 (0.00477)	0.02096 (0.00420)	0.02448 (0.00474)	0.05863 (0.02187)
Ka	0.00868 (0.00157)	0.01450 (0.00293)	0.00829 (0.00124)	0.00927 (0.00150)	0.00336 (0.00242)
Ks/Ka	2.15*	0.95	2.53**	2.64**	17.4**
Main cluster type 1 vs. 2					
Ks	0.01634 (0.00327)	0.01786 (0.00445)	0.02327 (0.00407)	0.02676 (0.00501)	N.A.
Ka	0.00822 (0.00140)	0.01333 (0.00227)	0.00967 (0.00138)	0.01296 (0.00211)	N.A.
Ks/Ka	1.99*	1.34	2.41**	2.06**	
"Old"[c] vs. Main cluster					
Ks	0.06515 (0.00919)	0.04691 (0.00927)	0.05815 (0.00717)	0.07437 (0.01046)	0.09038 (0.03184)
Ka	0.03090 (0.00339)	0.03836 (0.00486)	0.01698 (0.00179)	0.03054 (0.00277)	0.04947 (0.01428)
Ks/Ka	2.11***	1.22	3.42***	2.44***	1.83
K104 vs. Main cluster					
Ks	0.03764 (0.00882)	0.02987 (0.00933)	0.02868 (0.00575)	0.04300 (0.00880)	0.03680 (0.00141)
Ka	0.01648 (0.00291)	0.01663 (0.00349)	0.01159 (0.00190)	0.01820 (0.00323)	0.02294 (0.00918)
Ks/Ka	2.28*	1.80	2.47**	2.36**	1.60

[a] Standard errors are shown in parentheses.
Probability of the hypothesis that Ks = Ka established with a one-tailed *t* test: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.
[b] Calculated only for type 2 proviruses because the 292 bp deletion characteristic of type 1 proviruses removes almost all the *K-rev* gene.
N.A.: not applicable.
[c] For simplicity the more divergent sequences, namely HERV-K(C11a), HERV-K(C11b), and HERV-K(C1a), are grouped as "old".

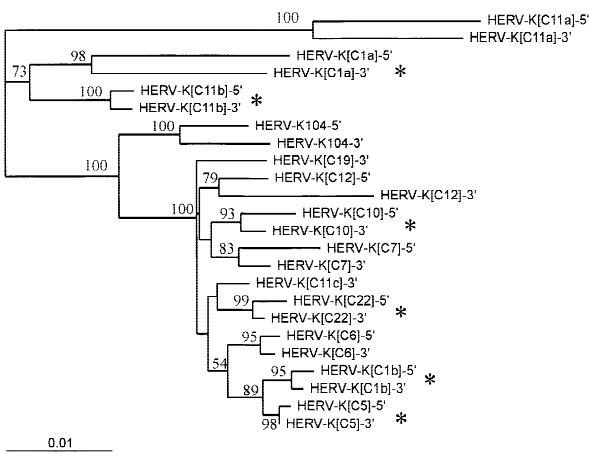

Fig. 3. Neighbor-joining tree of the full-length HERV-K proviruses based on LTR sequences. The value at a particular node indicates its percentage of appearance in 1000 bootstrap replicates. Only values above 50% are shown. An asterisk (*) identifies type 1 proviruses. Note that sequences HERV-K(C11c) and HERV-K(C19) present only one full-length LTR; while sequence HERV-K(C3) lacks the two LTRs. Exclusion of CpG dinucleotides prior to the analysis leads to a similar tree.

and type 2 proviruses than observed (Da = 0.00100). Third, the cluster of young type 1 proviruses based on the internal sequences is supported by a low bootstrap value (Fig. 2), while these proviruses are not grouped in the tree based on the LTRs (Fig. 3). These facts might be explained by a very recent origin of the young type 1 genome lineage, preventing the accumulation of diagnostic nucleotide differences. But, under this model, the Ks/Ka ratio might reflect lack of selective constrains within the young type 1 proviruses, and this is not the case (Table 2). Besides, if there was a "master gene" generating all the young type 1 proviruses in a short period of time, we should expect that the LTRs from all young type 1 proviruses show a "star" phylogeny. On the contrary, Fig. 3 reveals that the two LTRs from the same provirus always cluster together.

Finally, under the assumptions of the "master gene" model, all the type 2 human-specific insertions may correspond to proviruses that arose from the type 2 lineage after the type 2/type 1 split. Nevertheless, HERV-K104 (a human-specific insertion; Barbulescu et al. 1999) significantly appears in the phylogenetic trees as an outgroup of the main cluster of type 1 and type 2 elements (Figs. 2 and 3). All these facts strongly suggest that several proviruses, both of the type 2 and of the novel type 1 genome, do not merely become pseudogene copies of the "master genes." On the contrary, various new insertions have been subject to selection for transposability, giving rise to active lineages of HERV-K. Thus, several active HERV-K proviruses may be present within the human lineage after the split of human and

chimpanzee, approximately 6 Mya (Goodman et al. 1998).

Acknowledgments. I thank H. Naveira for his useful comments in an earlier version of the manuscript. The author is a recipient of a postdoctoral fellowship from the USC/Xunta de Galicia. This work was partially supported by Grant PGIDT99BIO10302 from Xunta de Galicia.

References

Altschul SF, Gish W, Miller W, Myers W, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Barbulescu M, Turner G, Seaman MI, Deinard AS, Kidd KK, Lenz J (1999) Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. Curr Biol 9:861–868

Bulmer M, Wolfe KH, Sharp PM (1991) Synonymous nucleotide substitution rates in mammalian genes: implications for the molecular clock and the relationship of mammalian orders. Proc Natl Acad Sci USA 88:5974–5978

Clough JE, Foster JA, Barnett M, Wichman HA (1996) Computer simulation of transposable element evolution: random template and strict master models. J Mol Evol 42:52–58

Costas J, Naveira H (2000) Evolutionary history of the human endogenous retrovirus family ERV9. Mol Biol Evol 17:320–330

Deininger PL, Batzer MA, Hutchison CA 3rd, Edgell MH (1992) Master genes in mammalian repetitive DNA amplification. Trends Genet 8:307–311

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP (1998) Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9:585–598

Hu W-S, Temin HM (1990) Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proc Natl Acad Sci USA 87:1556–1560

Kumar S, Tamura K, Jacobsen I, Nei M (2000) MEGA: molecular evolutionary genetics analysis, ver. 2. Distributed by the authors (http://www.megasoftware.net/)

Lebedev YB, Belonovitch OS, Zybrova NV, Khil PP, Kurdyukov SG, Vinogradova TV, Hunsmann G, Sverdlov ED (2000) Differences in HERV-K LTR insertions in orthologous loci of humans and great apes. Gene 247:265–277

Lee RN, Jaskula JC, Van Den Bussche RA, Baker RJ, Wichman HA (1996) Retrotransposon *mys* was active during evolution of the *Peromyscus leucopus-maniculatus* complex. J Mol Evol 42:44–51

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

Löwer R, Boller K, Hasenmaier B, Korbmacher C, Muller-Lantzsch N, Löwer J, Kurth R (1993) Identification of human endogenous retroviruses with complex mRNA expression and particle formation. Proc Natl Acad Sci USA 90:4480–4484

Löwer R, Löwer J, Kurth R (1996) The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. Proc Natl Acad Sci USA 93:5177–5184

Mager DL (1989) Polyadenylation function and sequence variability of the long terminal repeats of the human endogenous retrovirus-like family RTVL-H. Virol 173:591–599

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

Matassi G, Sharp PM, Gautier C (1999) Chromosomal location effects on gene sequence evolution in mammals. Curr Biol 9:786–791

Mayer J, Meese E, Mueller-Lantzsch N (1998) Human endogenous retrovirus K homologous sequences and their coding capacity in Old World primates. J Virol 72:1870–1875

Mayer J, Sauter M, Racz A, Scherer D, Mueller-Lantzsch N, Meese E (1999) An almost-intact human endogenous retrovirus K on human chromosome 7. Nat Genet 21:257–258

Maynard Smith J (1992) Analyzing the mosaic structure of genes. J Mol Evol 34:126–129

McDonald J (1993) Evolution and consequences of transposable elements. Curr Opin Genet Dev 3:855–864

Medstrand P, Mager DL (1998) Human-specific integrations of the HERV-K endogenous retrovirus family. J Virol 72:9782–9787

Nei M (1987) Molecular evolutionary genetics. Columbia Univ Press, New York, pp. 276–278

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nicholas KB, Nicholas HB Jr (1997) GeneDoc: a tool for editing and annotating multiple sequence alignment. Distributed by the authors (http://www.cris.com/~ketchup/genedoc.shtml)

Ono M, Yasunaga T, Miyata T, Ushikubo H (1986) Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. J Virol 60:589–598

Robertson DL, Hahn BH, Sharp PM (1995) Recombination in AIDS viruses. J Mol Evol 40:249–359

Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Smit AF (1996) The origin of interspersed repeats in the human genome. Curr Opin Genet Dev 6:743–748

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Tonjes RR, Czauderna F, Kurth R (1999) Genome-wide screening, cloning, chromosomal assignment, and expression of full-length human endogenous retrovirus type K. J Virol 73:9187–9195

Verneau O, Catzeflis F, Furano AV (1998) Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. Proc Natl Acad Sci USA 95:11284–11289