GENOME LETTERS 2002, 1, 1–5
www.aspbs.com/genomelett

RESEARCH ARTICLE

A Biologically Active Family of Human Endogenous Retroviruses Evolved from an Ancient Inactive Lineage

I. King Jordan[1] and John F. McDonald[2,*]

[1] *National Center for Biological Information, National Library of Medicine, National Institutes of Health, Bldg. 38A, Bethesda, Maryland, USA*
[2] *Department of Genetics, University of Georgia, Athens, Georgia, USA*

(Received: xxxx; accepted: xxxx)

ABSTRACT: Human endogenous retroviruses (HERVs) are remnants of ancient germ line infections that now make up a substantial fraction of the human genome. While most HERVs are inactive, there is a growing body of evidence that implicates some members of the HERV-K family of elements as being transpositionally active. Here we report the results of a phylogenetic survey of HERV-K LTR sequences. We have elucidated the evolutionary relationships among the youngest, most recently active (human specific) lineage of HERV-K elements and a number of more ancient lineages. Levels of sequence variation were used to estimate the ages of the different phylogenetic groups of element sequences. Our results suggest that a burst of transpositional activity led to the emergence of the human specific lineage of HERV-K elements and coincided with the time humans and chimps are believed to have diverged from a common ancestor ~6 million years ago. In addition, as noted previously, the youngest HERV-K subfamily shows a within-group pattern of variation where younger, more recently active subgroups are successively derived (evolve) from older subgroups. However, among HERV-K subfamilies there is no such correlation between phylogenetic relationship and the age of the groups. In fact, the oldest subfamily of HERV-K elements studied here appears to have given rise to the youngest and most recently active group of elements. This suggests that ancient families of HERVs may be capable of retaining the potential for biological activity over long spans of evolutionary time.

Keywords: Human Endogenous Retroviruses, HERV-K, LTR Retrotransposons, Evolution.

1. INTRODUCTION

A significant fraction of the human genome (~8%) comprises retroviral-like elements [1]. These include nonautonomous mammalian apparent LTR retrotransposons (MaLRs) [2] as well as endogenous retroviruses (HERVs) [3] that are remnants of ancient germ line infections. Most families of HERVs appear to be transcriptionally and transpositionally silent and are transmitted through the germ line in a strict Mendelian fashion. One notable exception is the HERV-K (K denotes a lysine tRNA primer binding site) family of elements. HERV-K transcripts have been detected in a variety of tissues and have been associated with reverse transcriptase activity in retroviral particles, suggesting that at least some elements may be biologically active [3–5]. Consistent with this suggestion, an apparently intact HERV-K provirus that retains all open reading frames and encodes a potentially active reverse transcriptase [6, 7] has recently been found. In addition, very recent

*Author to whom correspondence should be addressed.

 1537-3053/2002/01/01/5/$17.00+.25 doi:10.1166/gl.2002.11

insertions of full-length HERV-K proviruses that are polymorphic among different human populations have been identified [8].

The chromosomal positions of HERV-K elements among various primate species indicate that many HERV-K elements integrated prior to the divergence of hominoids from Old World monkeys [9, 10]. However, a more recent analysis of 37 HERV-K long terminal repeats (LTRs) taken from random human clones identified eight HERV-K integrations that are specific to humans [11]. In this same analysis it was determined that a correlation exists between the relative age of different HERV-K element groups and the taxonomic distribution of group members among primates. In addition, the phylogenetic relationships among the different aged groups were found to be consistent with the hypothesis that new groups of elements arise sequentially by amplification of one or a few "master" elements present in the most closely related progenitor group (i.e., the master gene model [12]). In this paper, we report the results of an analysis of 140 full-length (not truncated) HERV-K LTR sequences. The pattern of HERV-K evolution revealed by analysis of this larger and more diverse data set is not consistent with predictions of the master gene model. Particularly striking is the finding that the most recently evolved HERV-K group derived from one of the demonstrably oldest groups of HERV-K elements present in the human genome.

2. MATERIALS AND METHODS

The Seqlab (Wisconsin GCG package) implementation of the FASTA [13] program (default settings) was used to search the Genbank database (release 116) for sequences homologous to the 5′ long terminal repeat (LTR) of the recently inserted [11] canonical HERV-K10 element [14]. A more recent search of the HERV database [15] was used to confirm the results obtained here. The initial Genbank search resulted in the identification of hundreds of accessions containing HERV-K homologous sequences. More than half of these sequences were redundant Genbank entries or the HERV-K homologous region of SINE-R retroposons [16]. Both of these classes of hits were removed prior to further analysis. The majority of nonredundant hits were truncated HERV-K LTR sequences. These sequences were also removed prior to phylogenetic analysis. A total of 140 full-length HERV-K LTR sequences (>90% length of the canonical query element) were aligned using ClustalX [17] with default options. Minor manual adjustments to the sequence alignment were made as necessary. The alignment was used with the PAUP* program [18] to calculate pairwise distances between sequences using the Kimura-2 parameter distance correction [19] with pairwise deletion of gaps. Pairwise distances were calculated before and after removal of rapidly evolving CpG sites. Distances were used to reconstruct HERV-K phylogenies with the neighbor-joining method [20] in PAUP* [18]. The results of phylogenetic analyses based on CpG containing (as reported here) and CpG removed alignments were virtually identical. Support for internal branches of the HERV-K phylogenies was assessed with the use of 100 bootstrap replicates.

Subfamily ages were calculated by taking the average of the pairwise distances (K) between all sequences in a given subfamily (phylogenetic group). The average age for subfamilies was calculated using an estimate of 0.0016 changes per site per million years (my) for primate pseudogenes [average subfamily age (my) $= K/(0.0016 * 2)$] as described previously [21, 22]. Ages of individual full-length proviral elements were calculated similarly by using the pairwise distance between the 5′ and 3′ LTRs of individual proviral elements with the same calibration value as above. As with the phylogenetic reconstruction, element ages were calculated both before and after rapidly evolving CpG sites were removed. Removal of CpG sites resulted in a reduction of levels of variation and thus lower age estimates. However, the relative age estimates for the different groups remain exactly the same whether CpG-containing (as reported here) or CpG removed alignments were used. The age estimates obtained with CpG-containing alignments agreed better with previous results obtained for the age of HERV-K families [11, 23, 24].

3. RESULTS AND DISCUSSION

A phylogenetic tree generated from an alignment of 140 full-length HERV-K LTR sequences indicates that these elements fall into seven well-supported groups (Fig. 1). These include five monophyletic groups and two groups that are not monophyletic but rather are clearly delineated by well-supported internal branches on the main trunk of the tree. Comparison of relative branch lengths within groups indicates that group 1 elements share greater similarity than elements within any of the other groups. A more detailed analysis of the most recently evolved HERV-K group 1 demonstrates that it can be further divided into four additional subgroups (Fig. 2). Superposition of previous PCR results [11, 23] on the group 1 phylogeny indicates that subgroup 1a is a human-specific clade; that is, it consists of elements (* in Fig. 2) that were inserted subsequent to the separation of humans and chimps from a common ancestor. Subgroup 1b is a clade consisting of specific HERV-K insertions that are present in only humans and chimps (** in Fig. 2), or in human, chimps, and gorillas (*** in Fig. 2). Relative branch lengths within subgroups 1c and 1d indicate that these are likely even older clades with elements that integrated prior to those in subgroups 1a and 1b (Fig. 2).

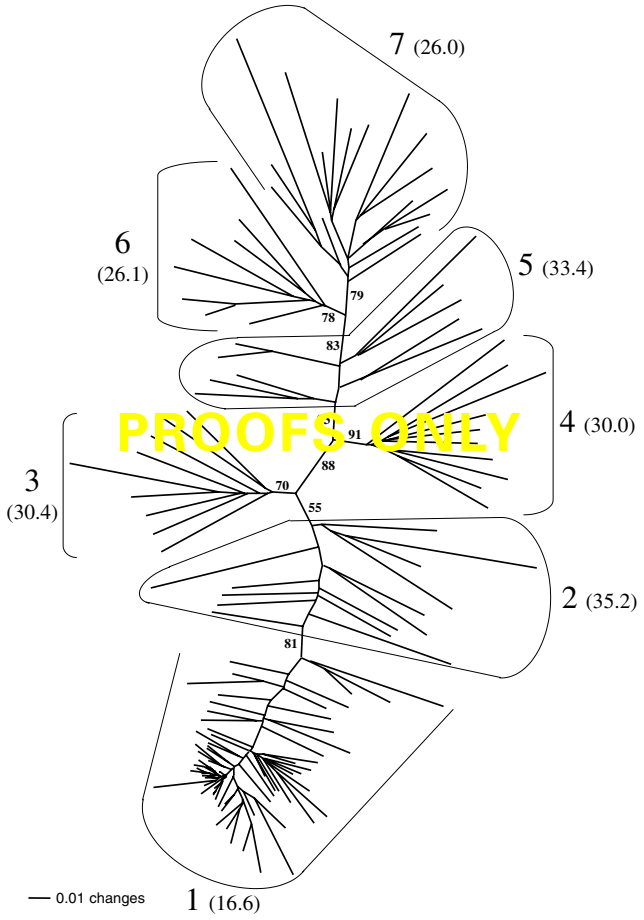

Figure 1. Phylogeny of the 140 HERV-K LTR sequences analyzed. There are seven well-supported HERV-K groups. Bootstrap values (100 replicates) are shown for the main branches separating the groups. Groups and their ages in millions of years (my) are indicated. Average group ages (my) were estimated before (1—16.6, 2—35.2, 3—30.4, 4—30.0, 5—33.4, 6—26.1, 7—26.0) and after (1—13.1, 2—30.0, 3—24.8, 4—24.7, 5—27.4, 6—21.9, 7—21.7) removal of CpG sites from the HERV-K LTR sequence alignment (see Materials and Methods). The two methods reveal the same relative ages among the groups. Scale bars indicate the number of changes per site that correspond to the indicated branch length.

Within group levels of HERV-K LTR sequence divergence can be used to estimate the age of the groups with the use of a pseudogene nucleotide substitution rate of 0.16% per million years [21, 22]. For example, an analysis of the nucleotide divergence among LTR sequences within HERV-K group 1 elements (Fig. 2) indicates that subgroup 1a is the most recently evolved, followed in order by subgroups 1b, 1c, and 1d. Interestingly, the age of subgroup 1a (~6 million years old) indicates that it evolved from subgroup 1b around the time humans and chimps diverged. Taken together with recent findings that support a role for HERVs in mediating genome rearrangement [25], the coincidence of the age of the element-rich group 1a with the human–chimp divergence date may suggest a role for these particular elements in driving primate speciation. The age estimates that we obtain are consistent with previous results [26], independent age estimates obtained here from intralement LTR data from full-length proviral elements (data not shown), as well as the taxonomic distribution of subgroup members (Fig. 2) determined by PCR [11, 23].

Our results indicate that the relative age of members of HERV-K group 1 correlates with their phylogenetic relationships. For example, the youngest and most recently active subgroup of elements (1a) is evolved from the next youngest subgroup (1b). This pattern holds for all of the subgroups (1a–d) in group 1 (Fig. 2). Such a correlation between the phylogenetic position and the age of different element subgroups was previously noted for HERV-K elements [11] and is consistent with the master gene model of transposable element evolution [12]. However, when our analysis was extended to all seven groups of HERV-K elements represented in the human genome, a surprisingly different result was obtained. In contrast to the expectation of the master gene model, no correlation was observed between the age of the seven HERV-K groups and their relative phylogenetic relationships (Fig. 1). For example, there are two pairs of distinct monophyletic groups that have approximately the same age (groups 6–7 and 3–4, Fig. 1). This is consistent with the multiple source model of transposition, where distantly related active elements coexist in the genome and serve as independent sources of amplification [27]. In addition, the two oldest groups analyzed here (2 and 5, Fig. 1) are approximately the same age and correspond roughly to a previously documented major wave of HERV-K integration into the genome [24]. These groups (2 and 5, Fig. 1) are somewhat anomalous in that they consist of multiple small monophyletic groups and/or multiple single lineages that branch off the main trunk of the tree. As such, the groups are more heterogeneous than the others and may include multiple active lineages within each group. Accordingly, the age estimates for these nonmonophyletic groups are less certain.

The topological and age relationships between groups 1 and 2 (Fig. 1) are most interesting. Contrary to the predictions of the master gene model [11, 12], the most recently active group of HERV-K elements (group 1) did not evolve from the next most recently active groups (groups 6 and 7). Rather, our results indicate that the most recently active group of HERV-K elements (group 1) evolved from what may be the oldest group of elements present within the human genome (group 2) some time between ~20 and 30 million years ago (Fig. 1). While the age of group 2 must be interpreted carefully for the reasons described above, it is clearly a relatively ancient and inactive lineage. Our findings suggest that even ancient families of retrotransposons are capable of retaining or otherwise reestablishing biological activity.

RESEARCH ARTICLE

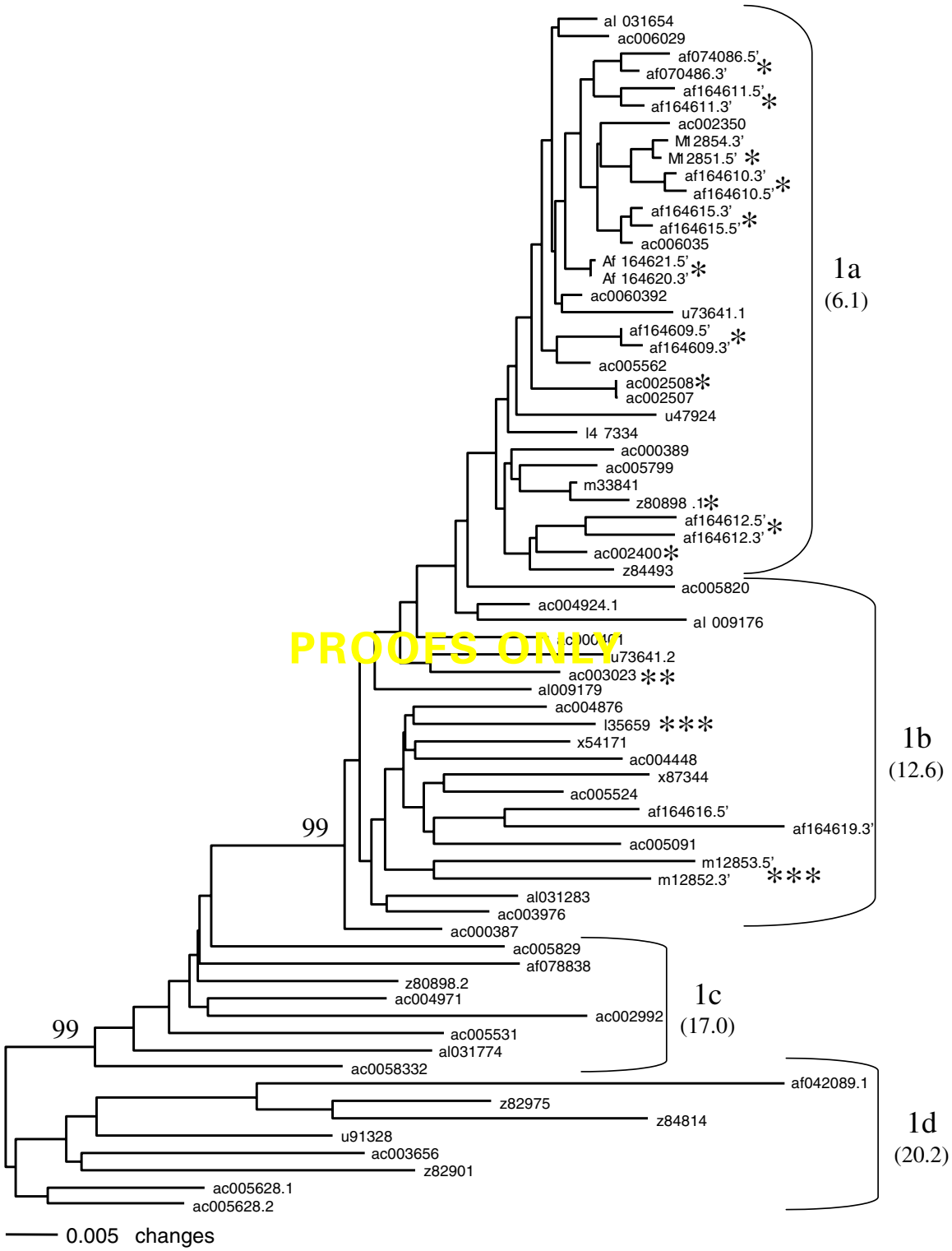

Figure 2. Phylogeny of HERV-K LTR group 1 sequences. The phylogeny is rooted using midpoint rooting. The four subgroups (1a–1d) and their ages, in millions of years (my), are indicated. Average subgroup ages (my) were estimated before (1a—6.1, 1b—12.6, 1c—17.0, 1d—20.2) and after (1a—4.4, 1b—9.4, 1c—13.5, 1d—16.5) removal of CpG sites from the HERV-K LTR sequence alignment (see Materials and Methods). The two methods reveal the same relative ages among the subgroups. Taxa names are Genbank accession numbers. LTR sequences from the ends of individual full-length proviral elements are indicated with a 5′ and 3′, respectively. The relative positions of HERV-K LTRs within accessions containing more than one element are indicated by .1 and .2. Scale bars indicate the number of changes per site that correspond to the indicated branch length. Results of previous PCR assays [11, 23] of the phyletic distributions of element insertions are indicated as follows: * indicates human specific insertions, ** indicates a human-chimp-specific insertion, and *** indicates human-chimp-gorilla-specific insertions.

References and Notes

1. E. S. Lander et al., Initial sequencing and analysis of the human genome. *Nature* 409, 860 **(2001)**.
2. A. F. Smit, Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res.* 21, 1863 **(1993)**.
3. R. Lower, J. Lower, and R. Kurth, The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl. Acad. Sci. USA* 93, 5177 **(1996)**.
4. R. R. Tonjes et al., HERV-K: the biologically most active human endogenous retrovirus family. *J. Acquired Immune Defic. Syndr. Hum. Retrovirol.* 13, S261 **(1996)**.
5. D. A. Wilkinson, D. L. Mager, and J. C. Leong, in edited by J. Levy, Plenum, New York **(1994)**, pp. 465–535.
6. J. Mayer, M. Sauter, A. Racz, D. Scherer, N. Mueller-Lantzsch, and E. Meese, An almost-intact human endogenous retrovirus K on human chromosome 7. *Nat. Genet.* 21, 257 **(1999)**.
7. K. Reus, J. Mayer, M. Sauter, D. Scherer, N. Muller-Lantzsch, and E. Meese, Genomic organization of the human endogenous retrovirus HERV-K(HML-2.HOM) (ERVK6) on chromosome 7. *Genomics* 72, 314 **(2001)**.
8. G. Turner, M. Barbulescu, M. Su, M. I. Jensen-Seaman, K. K. Kidd, and J. Lenz, Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr. Biol.* 11, 1531 **(2001)**.
9. R. Mariani-Costantini, T. M. Horn, and R. Callahan, Ancestry of a human endogenous retrovirus family. *J. Virol.* 63, 4982 **(1989)**.
10. S. Steinhuber, M. Brack, G. Hunsmann, H. Schwelberger, M. P. Dierich, and W. Vogetseder, Distribution of human endogenous retrovirus HERV-K genomes in humans and different primates. *Hum. Genet.* 96, 188 **(1995)**.
11. P. Medstrand and D. L. Mager, Human-specific integrations of the HERV-K endogenous retrovirus family. *J. Virol.* 72, 9782 **(1998)**.
12. P. L. Deininger, M. A. Batzer, C. A. Hutchison, III, and M. H. Edgell, Master genes in mammalian repetitive DNA amplification. *Trends Genet.* 8, 307 **(1992)**.
13. W. R. Pearson and D. J. Lipman, Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 85, 2444 **(1988)**.
14. M. Ono, Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. *J. Virol.* 58, 937 **(1986)**.
15. J. Paces, A. Pavlicek, and V. Paces, HERVd: database of human endogenous retroviruses. *Nucleic Acids Res.* 30, 205 **(2002)**.
16. M. Ono, M. Kawakami, and T. Takezawa, A novel human nonviral retroposon derived from an endogenous retrovirus. *Nucleic Acids Res.* 15, 8725 **(1987)**.
17. J. D. Thompson, T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876 **(1997)**.
18. D. L. Swofford, PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA **(1998)**.
19. M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111 **(1980)**.
20. N. Saitou and M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406 **(1987)**.
21. J. Costas and H. Naveira, Evolutionary history of the human endogenous retrovirus family ERV9. *Mol. Biol. Evol.* 17, 320 **(2000)**.
22. V. Kapitonov and J. Jurka, The age of Alu subfamilies. *J. Mol. Evol.* 42, 59 **(1996)**.
23. M. Barbulescu, G. Turner, M. I. Seaman, A. S. Deinard, K. K. Kidd, and J. Lenz, Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. *Curr. Biol.* 9, 861 **(1999)**.
24. E. D. Sverdlov, Retroviruses and primate evolution. *Bioessays* 22, 161 **(2000)**.
25. J. F. Hughes and J. M. Coffin, Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. *Nat. Genet.* 29, 487 **(2001)**.
26. Y. B. Lebedev, O. S. Belonovitch, N. V. Zybrova, P. P. Khil, S. G. Kurdyukov, T. V. Vinogradova, G. Hunsmann, and E. D. Sverdlov, Differences in HERV-K LTR insertions in orthologous loci of humans and great apes. *Gene* 247, 265 **(2000)**.
27. A. M. Shedlock and N. Okada, SINE insertions: powerful tools for molecular systematics. *Bioessays* 22, 148 **(2000)**.