

Letters to the Editor

Newly Identified Families of Human Endogenous Retroviruses

Human endogenous retroviruses (HERVs) make up approximately 8.3% of the human genome (12). HERVs have previously been classified into 31 distinct families based upon sequence alignment of reverse transcriptase (RT) and envelope domains and subsequent phylogenetic analyses (1, 9, 16). Using the data mining program LTR_STRUC (13) in conjunction with conventional sequence homology techniques, we recently completed an analysis of chimpanzee long terminal repeat (LTR) retrotransposon families (unpublished data). Since LTR_STRUC searches for LTR retrotransposons based on structure (e.g., the presence of LTRs, target site duplications, tRNA binding sites, etc.) rather than homology, elements can be identified that go undetected in traditional BLAST searches. We identified nine chimpanzee LTR retrotransposon families that are orthologous to HERV families not previously identified. These nine newly discovered HERV families are described and characterized in this letter.

LTR retrotransposons and retroviruses are grouped into three major classes (14). Class I contains elements related to gammaretroviruses, class II elements are related to betaretroviruses, and class III elements are related to spumaviruses. The RT-based phylogeny indicates that all the newly identified HERVs described here are class I elements (Fig. 1). The detailed characteristics of each of the newly discovered HERV families are presented in Table 1 and Table 2. All are low-abundance families, being composed of only one to seven full-length members with low homology to previously identified HERVs. This may, in part, explain why they have not been previously identified. The newly discovered full-length elements are of standard HERV length (7,198 to 10,675 bp with 359- to 682-bp LTRs) and display typically sized target site duplications (4 or 5 bp). With the exception of a few mutated copies, the newly identified elements have the same canonical dinucleotides terminating the LTRs as previously characterized HERVs (TG/CA). Since LTR_STRUC can only identify elements having two LTRs, we conducted BLAST searches by

using identified full-length elements as query sequences to identify solo LTRs and other fragmented elements. Consistent with what has been reported previously for other HERV families (15), we have found that each of the newly identified families is represented by significantly more solo LTRs and fragmented sequences than full-length elements (Table 1).

Because HERV LTRs are synthesized from the same RNA template during reverse transcription, they are identical in sequence at the time of integration (2). Using a primate pseudogene nucleotide substitution rate of 0.16% divergence/million years (5, 8, 10), the relative integration time or age of any full-length HERV can be estimated from the level of sequence divergence existing between the element's 5' and 3' LTRs. Using this method, the estimated ages of the new families of HERVs described here range from 18.0 to 49.5 million years, indicating that members of these families have not been transcriptionally active in the primate lineage since well before chimpanzees and humans diverged from a common ancestor (6 million years ago) (4). Although caution must be taken when using LTR divergence to estimate the ages of individual elements because of confounding processes such as recombination and conversion (e.g., see references 6 and 7), the method is able to provide useful age estimates, at least to a first approximation (e.g., see reference 3). Our estimated ages of the newly identified human elements fall within the median range of previously described families of HERVs (16). The possible contribution of these newly identified LTR retrotransposons to primate gene or genome evolution is currently under investigation.

This research was supported by a grant from the Georgia Institute of Technology Research Foundation.

REFERENCES

1. Benit, L., P. Dessen, and T. Heidmann. 2001. Identification, phylogeny, and evolution of retroviral elements based on their envelope genes. *J. Virol.* 75:11709–11719.

TABLE 1. Representative elements of human LTR retrotransposon families characterized in this study

Family name	Location on chromosome [chromosome no.; position (cytogenetic band) ^d]	5' and 3' LTR identity (%)	Lengths of 5'/3' LTRs (bp)	Target site repeats	Dinucleotides	Element length (bp)	tRNA primer	Age of element (10 ⁶ yr) ^a	Copy no. ^b
HERV 1	10; 42460513–42470423 (q11.21)	91.0	518/523	CCAC/CCAC	TG/CA	9,911	Pro	30.0	~21
HERV 2	1; 75864691–75868511 (p31.1)	76.0	328/329	TTTT/TTTT	TG/AA	3,821	ND ^c	49.5	~8
HERV 4	4; 75664267–75671483 (q13.3)	84	425/430	ACAG/ATAG	TG/CA	7,217	Glu	56.0	~170
HERV 5	3; 14055590–14063894 (p25.1)	90.7	359/361	TCAT/TCAT	TG/CA	8,305	Gln	31.0	~27
HERV 6	15; 82602937–82611762 (q25.3)	87.5	434/432	ND	AG/CT	8,826	ND	42.7	~40
HERV 7	3; 87738962–87748142 (p11.2)	88.7	659/682	ND	TG/CA	9,181	ND	38.0	~36
HERV 10	6; 120127640–120134837 (q22.31)	90.1	497/506	CAGT/CAGT	TG/CA	7,198	ND	33.0	~65
HERV 11	1; 31599889–31609246 (p35.2)	94.0	622/622	GCAA/GCAA	TG/CA	9,358	ND	19.3	~67
HERV 12	3; 168466792–168475512 (q26.1)	92.1	508/508	AGTT/AGTT	TG/CA	8,721	ND	25.9	~33

^a Determined by using a primate pseudogene nucleotide substitution rate of 0.16% divergence/million years (5, 8, 10), the relative integration time or age of full-length HERV was estimated from the level of sequence divergence existing between the element's 5' and 3' LTRs. The Jukes-Cantor model was used to correct for the presence of multiple mutations at the same site, back mutations, and convergent substitutions.

^b Includes full-length elements, fragmented copies, and solo LTRs.

^c ND, not determined. Because of the accumulation of substitutions, it was not possible to accurately determine target site repeats and tRNA binding sites.

^d Release hg17 of the University of California Santa Cruz (UCSC) genome browser.

CLASS I

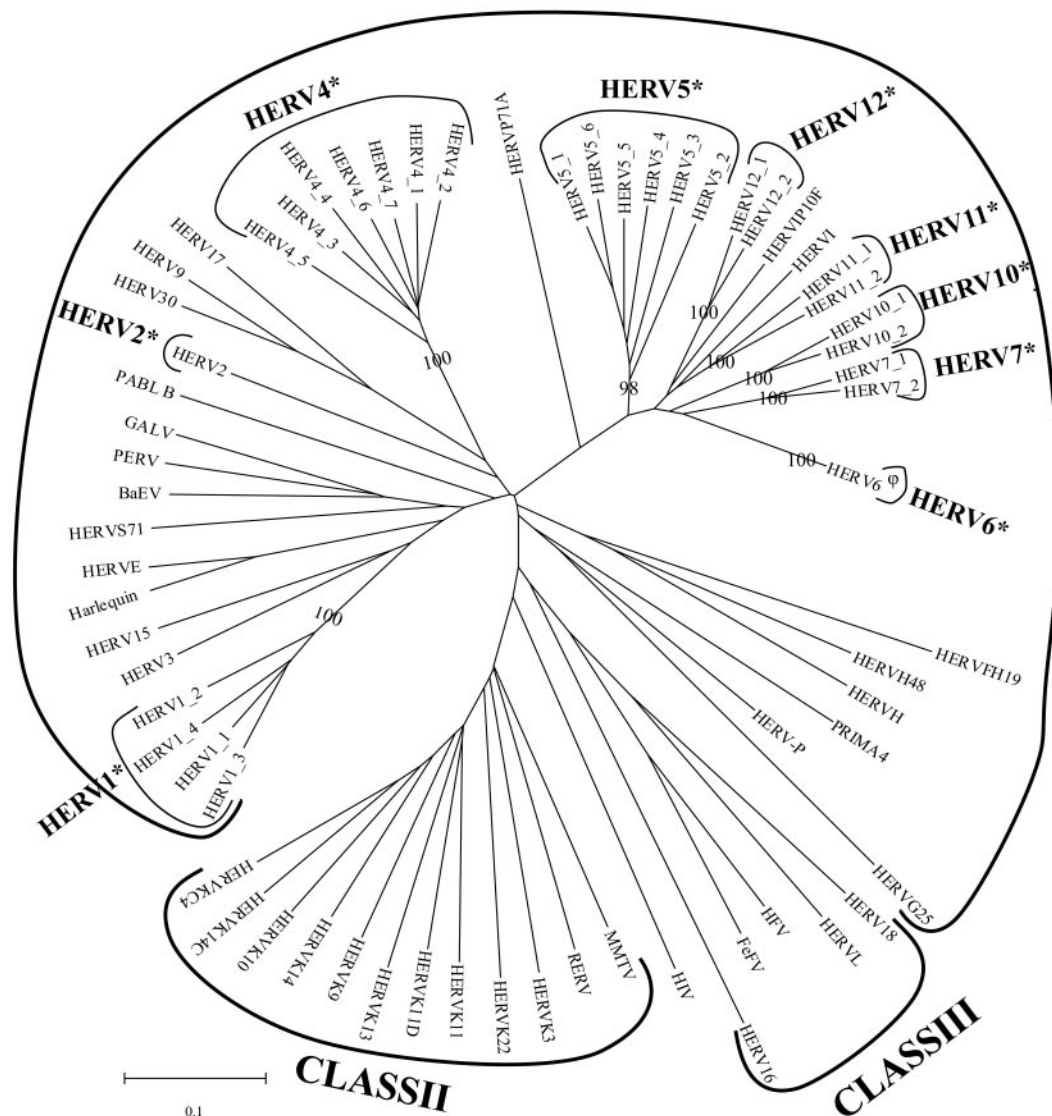


FIG. 1. Unrooted RT-based neighbor-joining tree of human LTR retrotransposon families. The phylogenetic tree is built from the DNA sequence (using MEGA3 software [11]) of the RTs taken from all the members of the newly identified HERV families together with representative members of previously identified families. (*, human LTR retrotransposon family characterized in this study. (†, this element is present in the duplicated region of the genome on chromosome 15; as a result, six elements of this family of greater than 97% identity are present in the human genome.) Elements are grouped into families based on bootstrap values (shown for families newly identified in this report). Previously identified families of retrotransposons and retroviruses are included for comparison (GALV, gibbon ape leukemia virus [accession no. M26927]; PERV, porcine endogenous retrovirus [accession no. AF038601]; BaEV, baboon endogenous virus [accession no. X05470]; HFV, human foamy virus [accession no. Y07725]; FeFV, feline foamy virus [accession no. AJ223851]; HIV, human immunodeficiency virus [accession no. K03454]; MMTV, mouse mammary tumor virus [accession no. NC_001503]; RERV, rabbit endogenous retrovirus [accession no. AF480925]).

TABLE 2. Characteristics of human LTR retrotransposon families identified in this study

Family name	Length of full element (bp)	Length of LTR (bp)	5' and 3' LTR identity	Age of element (10^6 yr)
HERV 1	9,671–10,665	380–526	91.0–96.19	12.2–29.78
HERV 2	3,821	328	76	49.52
HERV 4	7,217–9,766	249–489	82.5–92.62	24.25–62.24
HERV 5	8,059–10,724	359–637	87.9–92.14	25.92–41.22
HERV 6	8,820–8,825	430–434	87.0–88.0	42.0–43.0
HERV 7	9,181–9,668	563–682	88.29–88.7	38.01–39.75
HERV 10	7,198–9,400	495–509	90.1–94.4	18.0–33.0
HERV 11	9,358–9,916	552–622	90.0–94.0	19.38–30.16
HERV 12	8,721–8,794	484–508	88.42–92.4	24.97–25.99

2. **Boeke, J. D., and J. P. Stoye.** 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements, p. 343–435. *In* J. M. Coffin, S. H. Hughes, and H. Varmus (ed.), *Retroviruses*. Cold Spring Harbor Laboratory Press, Plainview, N.Y.
3. **Bowen, N. J., and J. F. McDonald.** 2001. *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res.* **11**:1527–1540.
4. **Chimpanzee Sequencing and Analysis Consortium.** 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**:69–87.
5. **Costas, J., and H. Naveira.** 2000. Evolutionary history of the human endogenous retrovirus family ERV9. *Mol. Biol. Evol.* **17**:320–330.
6. **Hughes, J. F., and J. M. Coffin.** 2005. Human endogenous retroviral elements as indicators of ectopic recombination events in the primate genome. *Genetics* **171**:1183–1194.
7. **Johnson, W. E., and J. M. Coffin.** 1999. Constructing primate phylogenies from ancient retrovirus sequences. *Proc. Natl. Acad. Sci. USA* **96**:10254–10260.
8. **Jordan, I. K., and J. F. McDonald.** 2002. A biologically active family of human endogenous retroviruses evolved from an ancient inactive lineage. *Genome Lett.* **1**:1–5.
9. **Jurka, J.** 2000. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16**:418–420.
10. **Kapitonov, V., and J. Jurka.** 1996. The age of Alu subfamilies. *J. Mol. Evol.* **42**:59–65.
11. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings Bioinformatics* **5**:150–163.
12. **Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, et al.** 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921.
13. **McCarthy, E. M., and J. F. McDonald.** 2003. LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics* **19**:362–367.
14. **Smit, A. F.** 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**:657–663.
15. **Stoye, J. P.** 2001. Endogenous retroviruses: still active after all these years? *Curr. Biol.* **11**:R914–R916.
16. **Tristem, M.** 2000. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. *J. Virol.* **74**:3715–3730.

Nalini Polavarapu
Nathan J. Bowen
John F. McDonald*
School of Biology
Georgia Institute of Technology
310 Ferst Dr.
Atlanta, GA 30332-0230

*Phone: (404) 385-6631
Fax: (404) 894-0519
E-mail: john.mcdonald@biology.gatech.edu