

Joo-Mi Yi · Heui-Soo Kim

Evolutionary implication of human endogenous retrovirus HERV-H family

Received: 12 November 2003 / Accepted: 14 January 2004 / Published online: 9 March 2004
© The Japan Society of Human Genetics and Springer-Verlag 2004

Abstract The human endogenous retrovirus (HERV-H) family is most abundant and widely distributed in the human genome, with about 100–1,000 full-length or deleted elements and a similar number of solitary, long-terminal repeats. The HERV-H *env* ORF has been characterized in humans and in the course of primate evolution, indicating the increased possibility of biological roles in humans. Using the polymerase chain reaction approach with a human monochromosomal DNA panel, 70 *env* fragments belonging to the HERV-H family from chromosomes 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, X, and Y were identified and analyzed. They showed 82–99% sequence similarity to that of HERV-H (accession no. AF108843). We also identified other HERV-H *env* fragments in the DDBJ/EMBL/GenBank databases. The total of 120 fragments was evolutionarily analyzed. Phylogenetic analysis suggests that the HERV-H *env* family is divided into one major and two minor groups. The HERV-H members have been actively proliferated and evolved by intrachromosomal spread during hominid radiation.

Keywords Envelope gene · HERV-H family · Human monochromosome · Molecular evolution

Introduction

The human genome comprises approximately 8% of the human endogenous retroviruses (HERVs) and other long terminal repeat (LTR)-like elements (International Human Genome Sequencing Consortium 2001). Most HERV families have been inserted into the primate

genome and subjected to amplification on several occasions between the divergence of hominoids and Old World monkeys 30–45 million years ago (Sverdlov 2000). At least 22 distinct HERV families were identified in the human genome (Tristem 2000). HERVs are present in full-length or incomplete sequences with multiple stop codons, insertions, deletions, and frame shifts in them (Lee et al. 2000; de Parseval et al. 2001). However, structural genes from some HERV families are expressed preferentially in human placenta (Venables et al. 1995) and several cancer cell lines (Lower et al. 1993; Armbruster et al. 2002). A small minority of such sequences has acquired a role in regulating gene expression, and some of these may have recently been subject to retrotransposition and rearrangement (Sverdlov 1998).

HERV-H is one of the most abundant endogenous retroviral families in the human genome, consisting of full-length elements (about 100 copies), elements deleted in *pol* and *env* (800–900 copies), and solitary LTRs (about 1,000 copies) (Hirose et al. 1993). The HERV-H was inserted into primate genomes before the divergence of New World and Old World monkeys, and the elements lacking in *pol* and *env* were integrated with the Old World monkey lineage (Goodchild et al. 1993; Anderssen et al. 1997; de Parseval et al. 2001). The full-length HERV-H *env* contains a region encoding an amino acid sequence highly similar to the immunosuppressive peptide in murine leukaemia virus (MLV) p15E protein that may suppress maternal immunological rejection of the fetus and may also be involved in tumor development (Cianciolo et al. 1984, 1985). Expression of immunosuppressive protein-encoding HERV-H *env* transcripts has been confirmed in various normal and malignant cell types (Mangeny and Heidmann 1998; Mangeny et al. 2001). Recently, three HERV-H envelope genes (*HERV-H/env62*, *HERV-H/env60*, and *HERV-H/env59*) were analyzed in view of primate evolution (de Parseval et al. 2001).

In this report, we newly identified 70 HERV-H *env* sequences from human monochromosomes and analyzed them with those of HERV-H sequences in the

J.-M. Yi · H.-S. Kim (✉)
Division of Biological Sciences, College of Natural Sciences,
Pusan National University, Pusan, 609-735 South Korea
E-mail: khs307@pusan.ac.kr
Tel.: +82-51-5102259
Fax: +82-51-5812962

databases. We will discuss the evolutionary implication of HERV-H on the basis of our analysis.

Materials and methods

Using the polymerase chain reaction (PCR) approach, we identified the HERV-H *env* family from a human monochromosomal DNA panel purchased from the Coriell Cell Repositories (Coriell Institute, Camden, N.J., USA). New 596-bp *env* fragments of the HERV-H family were amplified by the primer pairs JM12 (5'-GTCGGTTTAGGACTTTCTGC-3', bases 1827–1846) and JM02 (5'-TGTGGGAACCTAGAGCGGA-3', bases 2401–2420) from the HERV-H (DDBJ/EMBL/GenBank, accession no. AF108843 from human chromosome 2). The PCR conditions followed were those of Kim et al. (1996), with an annealing temperature of 58°C. PCR products were separated on a 1.5% agarose gel, purified with the QIAEX II gel extraction kit (QUIAGEN, Chatsworth, Calif., USA) and cloned into the pGEM-T easy vector (Promega, Madison, Wis., USA). The cloned DNA was isolated by the alkali lysis method using the High Pure plasmid isolation kit (Roche, Indianapolis, Ind., USA). Individual plasmid DNA was digested by restriction enzyme *EcoRI*. Positive samples were subjected to sequence analyses on both strands with T7 and M13 reverse primers using an automated DNA sequencer (Model 373A) and the DyeDeoxy terminator kit (Applied Biosystems, Foster City, Calif., USA). Nucleotide sequence analyses were performed using the GAP, PILEUP, and PRETTY from the GCG software (Genetics Computer Group, University of Wisconsin, Madison, Wis., USA). The neighbor-joining tree was obtained with the MEGA2 program (Kumar et al. 2001). Bootstrap evaluation of the branching patterns was performed with 100 replications. Nucleotide sequences of the HERV-H family were retrieved from the DDBJ/EMBL/GenBank databases with the aid of the BLAST network server (Altschul et al. 1997) and the UCSC BLAT search (Kent 2002).

Results and discussion

The PCR products were found on human chromosomes 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, X, and Y of the human monochromosomal DNA panel (Fig. 1). Human genomic DNA also showed clear bands (lane A), whereas no bands appeared in mouse (lane B) and hamster DNAs (lane C). Each of the PCR products was cloned, and then 15 clones were selected randomly and sequenced. Among 15 clones of the HERV-H family, at least two independent clones showed same sequence identity. Therefore, 70 *env* gene sequences belonging to the HERV-H family were identified and analyzed (Table 1). They showed 82.2–99.0% sequence similarity to that of HERV-H (AF108843). We also retrieved the HERV-H family from the DDBJ/EMBL/GenBank databases in order to carry out comparative analysis with our data. The number of the retrieved family members was 50. The HERV-H family was not yet identified on human chromosomes 8, 13, 15, 17, 21, and 22 by searching databases. In the present study, we could obtain several HERV-H copies on chromosomes 7 (HE7-2, HE7-3, HE7-4, HE7-5, HE7-6), 15 (HE15-1, 2, 11), and 17 (HE17-2). However, we could not obtain additional HERV-H sequences that gave difference from those of databases (AL162151, AL356804,

AF108841 from chromosome 14, and AL360078 from chromosome 20), indicating that they may exist at low copy numbers on the human chromosomes 14 and 20. The nucleotide sequences of HERV-H *env* family reported in this paper appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession numbers AB100173–AB100242 from human monochromosomes.

Recently, the HERV-H *env* sequences were shown to have immunosuppressive properties (Mangeny et al. 2001). The amino acid sequence corresponding to the immunosuppressive domain of the envelope protein of HERV-H is LQNRRCDLLLTAEKGG (Mangeny et al. 2001). We also examined the corresponding amino acid sequence among the clones from the human monochromosomal DNA panel. The amino acid sequences in clones HE6-1, 3 on chromosome 6, HE9-4 on chromosome 9, HE18-5 on chromosome 18, and HE19-4 on chromosome 19 were similar to the immunosuppressive peptide and had no interruptions by deletion/insertion or stop codons. It is thus possible that they have a similar function to the immunosuppressive peptide. These observations could lead to a more clear understanding of the immunological role of HERV-H *env* ORF in the human genome.

Using all HERV-H *env* family members, including the DDBJ/EMBL/GenBank data, a dendrogram was constructed by the neighbor-joining method to examine their relationships. The HERV-H *env* sequences were divided into three groups, one major (group I) and two minor (groups II and III) through nucleotide distances (Fig. 2). All clones from chromosome 7 (HE7-2, HE7-3, HE7-4, HE7-5, HE7-6) belonged to group II, while the clones from chromosome 1 (HE1-5), chromosome 2 (HE2-1, HE2-3), and chromosome 17 (HE17-2) belonged to group III. Interestingly, high copy numbers of group I indicated that they continuously expanded by duplication and clustering on each chromosome. Within group I, the HERV-H *env* family could have been amplified at least nine times after the original integration into the hominid genome. Our result is in accordance with that of Goodchild et al. (1993), who classified HERV-H LTR elements into three classes: two classes that had been amplified before Old World monkey lineage, and one class that had been multiplied in a large copy numbers before gorilla diverged.

On the assumption that the dendrogram in Fig. 2 reflects the phylogenetic relationships of the HERV-H family, we computed the pairwise divergences for the three groups as group I = 4%, group II = 12%, and group III = 17%. We then estimated the divergence times of the three groups as 10 Myr (million years) (group I), 30 Myr (group II), and 43 Myr (group III), using the average evolutionary rate of 0.2% per million year (Anderssen et al. 1997). Approximately 10 Myr ago (gorilla lineage), the copy number of the HERV-H family was proliferated, suggesting that the elements

belonging to group I have integrated with primate genomes and evolved by intra-chromosomal duplication. This is in agreement with the result of PCR analysis that HERV-H elements with an open *env* reading frame (HERV-H10 and HERV-H18) are present in human and African great apes, but not in orangutan, gibbon, or the Old World monkey (Lindeskog et al.

1999). Therefore, the major expansion of the HERV-H elements belonging to group I in human and African great apes occurred after orangutan and gorilla diverged. The group I contains HERV-H10 (AF108841), HERV-H18 (AF108842), HERV-H19 (AF108843) (Lindeskog et al. 1999), and HERV-H/*env*62 (AJ289709), HERV-H/*env*60 (AJ289710) with the complete HERV-H *env* gene sequences (de Parseval et al. 2001). These HERV-H sequences show 4.5% divergence on the average, suggesting that they were amplified 11 Myr ago.

To summarize, we found new HERV-H *env* fragments on human chromosomes 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, X, and Y. These new members showed 82–99% sequence similarity to that of HERV-H (AF108843). We also identified other HERV-H *env* fragments in the DDBJ/EMBL/GenBank databases. The total of 120 fragments was evolutionarily analyzed. A neighbor-joining tree suggests that the HERV-H *env* family is divided into one major and two minor groups. The HERV-H members have been evolved by intra-chromosomal spread during hominid radiation.

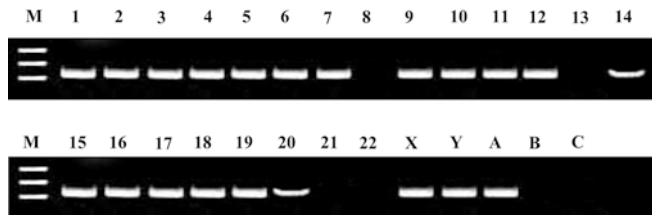
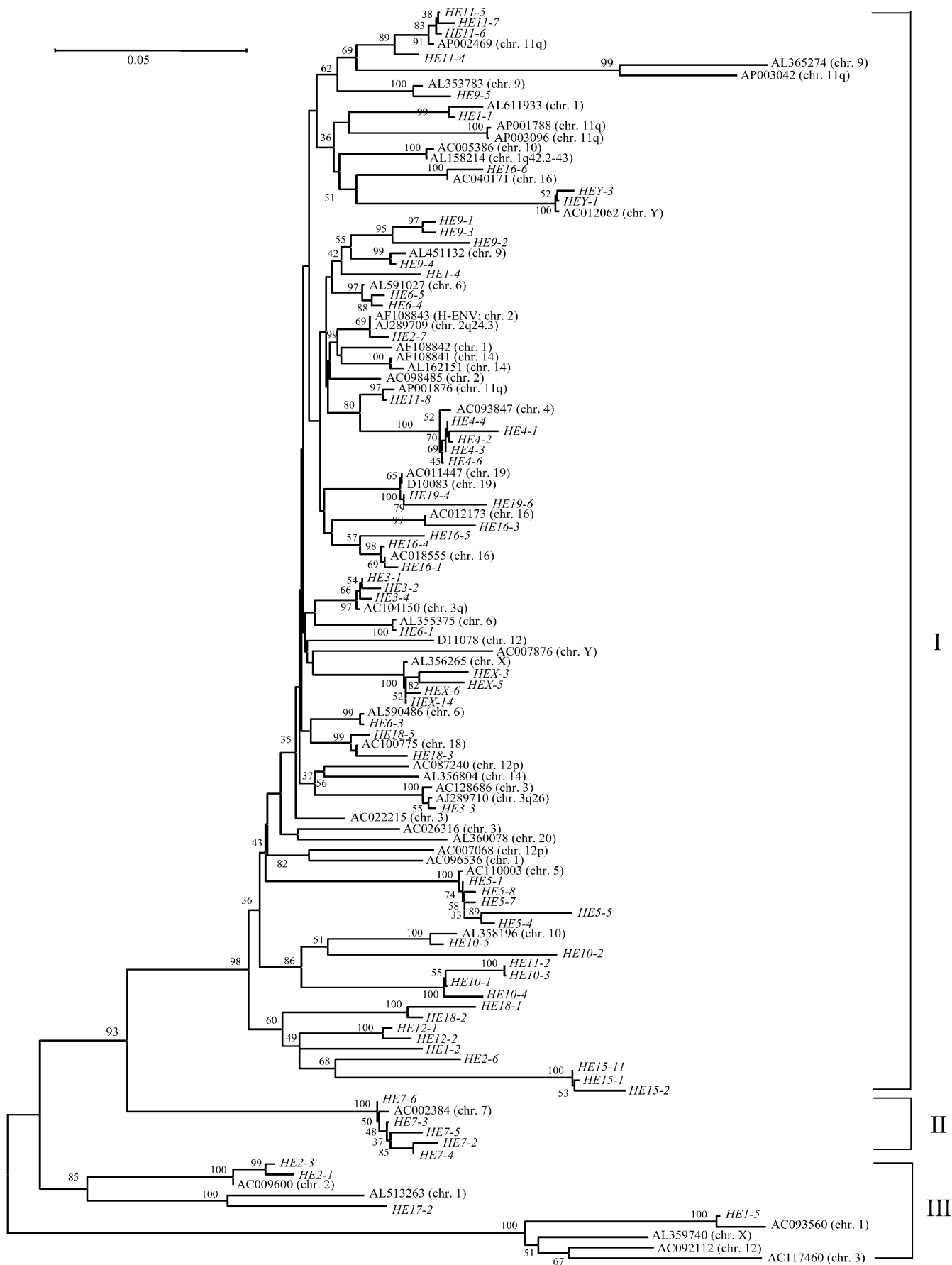


Fig. 1 Polymerase chain reaction analysis of genomic DNA for the presence of *env* fragments of the human endogenous retrovirus (HERV-H) family from human monochromosome DNA panel. The numbers designating the lines refer to the human monochromosome. *M* Marker (pUC18/*TaqI*), *A* human genomic DNA, *B* mouse DNA, *C* hamster DNA

Table 1 Chromosomal localization of the human endogenous retrovirus (HERV-H) *env* family

Chromosome	DDBJ/EMBL/GenBank data	Present data
1	AC096536, AL611933, AL513263, AC093560, AL158214(1q42.2-43), AF108842	HE1-1 (AB100173), HE1-2 (AB100174), HE1-4 (AB100175), HE1-5 (AB100176)
2	AC098485, AF108843, AJ289709(2q24.3), AC009600	HE2-1 (AB100177), HE2-3 (AB100178), HE2-6 (AB100179), HE2-7 (AB100180)
3	AC022215, AC026316, AC104150(3q), AC128686, AJ289710(3q26), AC117460	HE3-1 (AB100181), HE3-2 (AB100182), HE3-3 (AB100183), HE3-4 (AB100184)
4	AC093847	HE4-1 (AB100185), HE4-2 (AB100186), HE4-3 (AB100187), HE4-4 (AB100188), HE4-6 (AB100189)
5	AC110003	HE5-1 (AB100190), HE5-4 (AB100191), HE5-5 (AB100192), HE5-7 (AB100193), HE5-8 (AB100194)
6	AL355375, AL590486, AL591027	HE6-1 (AB100195), HE6-3 (AB100196), HE6-4 (AB100197), HE6-5 (AB100198)
7	AC002384	HE7-2 (AB100199), HE7-3 (AB100200), HE7-4 (AB100201), HE7-5 (AB100202), HE7-6 (AB100203)
8	-	-
9	AL353783, AL365274, AL451132	HE9-1 (AB100204), HE9-2 (AB100205), HE9-3 (AB100206), HE9-4 (AB100207), HE9-5 (AB100208)
10	AL358196, AC005386	HE10-1 (AB100209), HE10-2 (AB100210), HE10-3 (AB100211), HE10-4 (AB100212), HE10-5 (AB100213)
11	AP001788(11q), AP001876(11q) AP002469(11q), AP003042(11q) AP003096(11q)	HE11-2 (AB100214), HE11-4 (AB100215), HE11-5 (AB100216), HE11-6 (AB100217), HE11-7 (AB100218), HE11-8 (AB100219)
12	AC007068(12p), AC087240(12p), D11078, AC092112	HE12-1 (AB100220), HE12-2 (AB100221)
13	-	-
14	AL162151, AL356804, AF108841	-
15	-	HE15-1 (AB100222), HE15-2 (AB100223), HE15-11 (AB100224)
16	AC012173, AC018555, AC040171	HE16-1 (AB100225), HE16-3 (AB100226), HE16-4 (AB100227), HE16-5 (AB100228), HE16-6 (AB100229)
17	-	HE17-2 (AB100230)
18	AC100775	HE18-1 (AB100231), HE18-2 (AB100232), HE18-3 (AB100233), HE18-5 (AB100234)
19	AC011447, D10083	HE19-4 (AB100235), HE19-6 (AB100236)
20	AL360078	-
21	-	-
22	-	-
X	AL356265, AL359740	HEX-3 (AB100237), HEX-5 (AB100238), HEX-6 (AB100239), HEX-14 (AB100240)
Y	AC007876, AC012062	HEY-1 (AB100241), HEY-3 (AB100242)



◀ **Fig. 2** Neighbor-joining tree for the HERV-H *env* family on human chromosomes. The dendrogram is derived from the HERV-H *env* sequences identified in the GenBank database and this study. The *env*-containing sequences belonging to the HERV-H family are named according to the GenBank accession numbers. 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 replications. The nucleotide sequence data of the HERV-H family in *italics* reported in this article have been deposited in the DDBJ/EMBL/GenBank nucleotide sequence databases under accession numbers AB100173–AB100242 as shown in Table 1

Acknowledgments We thank Dr. Y. Tateno, National Institute of Genetics, Japan, for helpful comments on the manuscript. This study was supported by a grant of the Korea Health 21 R and D Project, Ministry of Health and Welfare, Republic of Korea (02-PJ1-PG3-21001-0004).

References

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman J (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* 25:3389–3402
- Anderssen S, Sjøttem E, Svineng G, Johansen T (1997) Comparative analyses of LTRs of the ERV-H family of primate-specific retrovirus-like elements isolated from marmoset, African green monkey, and man. *Virology* 234:14–30
- Armbruester V, Sauter M, Krautkraemer E, Meese E, Kleiman A, Best B, Roemer K, Mueller-Lantzsch N (2002) A novel gene from the human endogenous retrovirus K expressed in transformed cells. *Clin Cancer Res* 8:1800–1807
- Cianciolo GJ, Phipps D, Snyderman R (1984) Human malignant and mitogen-transformed cells contain retroviral P15E-related antigen. *J Exp Med* 159:964–969
- Cianciolo GJ, Copeland TD, Oroszlan S, Snyderman R (1985) Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins. *Science* 230:453–455
- Goodchild N, Wilkinson DA, Mager D (1993) Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements. *Virology* 196:778–788
- Hirose Y, Takamatsu M, Harada F (1993) Presence of *env* genes in members of the RTVL-H family of human endogenous retrovirus-like elements. *Virology* 192:52–61
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
- Kent WJ (2002) BLAT—The BLAST-like alignment tool. *Genome Res* 12:656–664
- Kim HS, Hirai H, Takenaka O (1996) Molecular features of the *TSPY* gene of gibbons and old world monkeys. *Chrom Res* 4:500–506
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245
- Lee JM, Choi JY, Kim JS, Hyun BH, Kim HS (2000) Identification and phylogeny of new human endogenous retroviral sequences belonging to the HERV-H family. *AIDS Res Hum Retroviruses* 16:2055–2058
- Lindeskog M, Mager DL, Blomberg J (1999) Isolation of a human endogenous retroviral HERV-H element with an open *env* reading frame. *Virology* 258:441–450
- Lower R, Boller K, Hasenmaier B, Korbmacher C, Muller-Lantzsch N, Lower J, Kurth R (1993) Identification of human endogenous retroviruses with complex mRNA expression and particle formation. *Proc Natl Acad Sci USA* 90:4480–4484
- Mangeny M, Heidmann T (1998) Tumor cells expressing a retroviral envelope escape immune rejection in vivo. *Proc Natl Acad Sci USA* 95:14920–14925
- Mangeny M, de Parseval N, Thomas G, Heidmann T (2001) The full-length envelope of an HERV-H human endogenous retrovirus has immunosuppressive properties. *J Gen Virol* 82:2515–2518
- de Parseval N, Casella J, Gressin L, Heidmann T (2001) Characterization of the three HERV-H proviruses with an open envelope reading frame encompassing the immunosuppressive domain and evolutionary history in primates. *Virology* 279:558–569
- Sverdlov ED (1998) Perpetually mobile footprints of ancient infections in human genome. *FEBS Lett* 428:1–6
- Sverdlov ED (2000) Retroviruses and primate evolution. *Bioessays* 22:161–171
- 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
- Venables PJ, Brookes SM, Griffiths D, Weiss RA, Boyd MT (1995) Abundance of an endogenous retroviral envelope protein in placental trophoblasts suggests a biological function. *Virology* 211:589–592