J. Microbiol. Biotechnol. (2002), **12**(3), 508–513

JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY

© The Korean Society for Microbiology and Biotechnology

Identification and Phylogeny of the Human Endogenous Retrovirus HERV-W LTR Family in Human Brain cDNA Library and Xq21.3 Region

KIM, HEUI-SOO[1]* AND TIMOTHY J. CROW[2]

[1]*Division of Biological Sciences, College of Natural Sciences, Pusan National University, Pusan 609-735, Korea*
[2]*POWIC, Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, OX3 7JX, U.K.*

Received: September 26, 2001
Accepted: April 8, 2002

Abstract Human endogenous retroviral long terminal repeats (LTRs) have been found to be coexpressed with sequences of genes located nearby. It has been suggested that the LTR elements have contributed to the structural change or genetic variation of human genome connected to various diseases. The HERV-W family has been identified in the cerebrospinal fluids and brains of individuals with schizophrenia. Using a cDNA library derived from a human brain, the HERV-W LTR elements were examined and five new LTR elements were identified. These elements were examined using a YAC clone panel from the Xq21.3 region linked to psychosis that was replicated on the Y chromosome after the separation of the chimpanzee and human lineages. Fourteen elements of the HERV-W LTR were identified in that region. Those LTR elements showed a high degree of sequence similarity (91.8 – 99.5%) with previously reported HERV-W LTR. A phylogenetic tree obtained from the neighbor-joining method revealed that new HERV-W LTR elements were closely related to the AX000960, AF072504, and AF072506 from the GenBank database. The data indicates that several copy numbers of the HERV-W LTR elements exist on the Xq21.3 region and are also expressed in the human brain. These LTR elements need to be further investigated as potential leads to neuropsychiatric diseases.

Key words: Brain cDNA library, HERV-W LTR elements, psychosis, phylogeny, Xq21.3 region

Approximately 1% of the human genome is represented by human endogenous retroviruses (HERVs) that are footprints of ancient germ-cell retroviral infections [29]. Full-length retroviral sequences may interact with cellular oncogenes [30], and retroviral long terminal repeat (LTR) sequences have the capacity to exert a regulatory influence as promoters and enhancers of cellular genes. Most HERV families encompass a relatively low copy number of per haploid genome [23], compared with others that are either high copy number or single copy retroviral elements [26]. These differences in the copy numbers could represent either multiple integration events or provirusly amplified after the integration by retrotransposition. Comparative analysis of the HERV LTR elements in the human genome could help us to understand the possible impact of HERVs on evolution and genome regulation.

Retroviral particles have been recovered from monocyte cultures from patients with multiple sclerosis [27], and virion-associated MSRV (multiple sclerosis associated retrovirus)-RNA has been reported in the serum of patients with the disease [9]. Expression of MSRV sequences in a normal placenta allowed the reconstruction of a 7.6-kb putative genomic retroviral RNA with RU5-gag-pol-env-U3R organization, with a polypurine binding site (PBS) showing similarity with avian retrovirus PBS used by tRNATrp [5]. Southern blot hybridizations using MSRV probes allowed characterization of a copy MSRV-related human endogenous retrovirus family named HERV-W [5]. In the previous study, HERV-W *pol* and *env* gene sequences in human monochromosomes were examined, and multiple frameshift and termination codons were found by deletion/insertion or point mutation [13, 15]. The HERV-W LTR elements were detected in hominoids, Old and New World monkeys, suggesting that they had inserted in the primate genome approximately 55 million years ago [17]. The expression, structure, and promoter activity of HERV-W LTR elements were examined in human cell lines [28]. Recently, the HERV-W family was identified in the cerebrospinal fluids and brains of individuals with schizophrenia [11]. We were interested in exploring the retroviral elements within the Xq21.3 region, because this region has been putatively linked to psychosis [24]. Here, the HERV-W LTR family within the Xq21.3 region and the cDNA library of the human brain were identified, and

*Corresponding author
Phone: 82-51-581-2962; Fax: 82-51-510-2259;
E-mail: khs307@pusan.ac.kr

phylogenetically analyzed with those sequences derived from GenBank database. Such retroviral elements could be of great potential relevance to neuropsychiatric diseases.

The cDNA synthesized from mRNA of human brain and YAC clone panel from the human Xq21.3 region were used as a template for PCR amplification. New 416-bp LTR elements of the HERV-W family were amplified by the primer pairs HS47 (5'-TGGTCCATGTTTCTTACGGCT-3', bases 127-147) and DS16 (5'-AAGATGGTGGTGAACC-ACTTC-3', bases 521-541) from the HERV-W (GenBank, accession no. AF072500). The PCR conditions were used as described by Kim *et al.* [16, 19] with an annealing temperature of 56°C. PCR products were separated on 2% agarose gel, purified with the QIAEX II gel extraction kit (Qiagen, Chatsworth, U.S.A.), and cloned into the T-khs307 vector [14]. The cloned DNA was isolated by the alkali lysis method using the High Pure plasmid isolation kit (Boehringer Mannheim, Germany). Individual plasmid DNAs were screened for inserts by PCR using the original primers designed for the locus. The plasmid DNAs 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 Biosystem, Foster City, U.S.A.). Nucleotide sequence analysis was performed using the GAP and PILEUP programs of the GCG software (Genetics Computer Group, University of Wisconsin, Madison, U.S.A.). The neighbor-joining phylogenetic analysis [18] was performed with the MEGA program [22]. Nucleotide sequences of HERV-W LTR elements were retrieved from the GenBank database using the BLAST network server [2]. The nucleotide sequences of new HERV-W LTR elements reported in this paper have been deposited in the DDBJ/EMBL/GenBank nucleotide sequence databases with the following accession numbers: clone HB-1 (AB066648), HB-3 (AB066649), HB-11 (AB066650), HB-12 (AB066651), HB-15 (AB066652), X3-1 (AB064529), X3-2 (AB064530), X3-3 (AB064531), X7-1 (AB064532), X7-2 (AB064533), X7-3 (AB064534), X7-4 (AB064535), X7-5 (AB064536), X11-1 (AB064537), X11-2 (AB064538), X13-1 (AB064539), X13-2 (AB064540), X13-3 (AB064541), and X13-4 (AB064542).

Retroviruses have been known to be one of the infectious agents involved in the pathogenesis of schizophrenia. Karlsson *et al.* [11] reported the identification of retroviral sequences in cerebrospinal fluids obtained from individuals with recent-onset schizophrenia, and the differential transcriptional upregulation of members of the HERV-W family of endogenous retroviruses in the postmortem frontal cortex of individuals with schizophrenia. HERV-W LTR elements in the human brain cDNA library were examined using the PCR approach, identifying five LTR elements (HB-1, HB-3, HB-11, HB-12, and HB-15). Those LTR elements showed a high degree of sequence similarity (91.8–99.5%) with that of HERV-W LTR (AF072500) (Table 1). In alignments analysis of HERV-W LTR elements, some clones (HB-11, HB-12, and HB-15) contained additional sequences, TGCCACCACCACTGTT, in the human brain cDNA (Fig. 1).

The Xq21.3 region that has been linked with psychosis was also examined [24]. Using a YAC clone panel, PCR amplification was performed and 14 HERV-W LTR elements (X3-1, X3-2, X3-3, X7-1, X7-2, X7-3, X7-4, X7-

Table 1. Percentage similarity of nucleotide sequences of HERV-W LTR elements.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. W-LTR	-																			
2. HB-1	99.5	-																		
3. HB-3	97.9	98.4	-																	
4. HB-11	91.8	92.3	93.3	-																
5. HB-12	91.8	92.3	93.8	99.0	-															
6. HB-15	92.3	92.8	93.8	99.5	99.5	-														
7. X3-1	99.5	100	98.4	92.3	92.3	92.8	-													
8. X3-2	99.2	99.7	98.1	92.0	92.0	92.5	99.7	-												
9. X3-3	99.2	99.7	98.1	92.0	92.0	92.5	99.7	99.5	-											
10. X7-1	96.2	96.8	97.3	91.3	91.3	91.8	96.8	96.5	96.5	-										
11. X7-2	99.5	100	98.4	92.3	92.3	92.8	100	99.7	99.7	96.8	-									
12. X7-3	98.4	98.9	97.3	91.3	91.3	91.8	98.9	98.7	98.7	95.7	98.9	-								
13. X7-4	99.2	99.7	98.1	92.0	92.0	92.5	99.7	99.5	99.5	96.5	99.7	98.7	-							
14. X7-5	98.9	99.5	97.9	91.8	91.8	92.3	99.5	99.2	99.2	96.2	99.5	98.4	99.2	-						
15. X11-1	99.2	99.7	98.1	92.0	92.0	92.5	99.7	99.5	99.5	96.5	99.7	98.7	99.5	99.2	-					
16. X11-2	99.5	100	98.4	92.3	92.3	92.8	100	99.7	99.7	96.8	100	98.9	99.7	99.5	99.7	-				
17. X13-1	99.5	100	98.4	92.3	92.3	92.8	100	99.7	99.7	96.8	100	98.9	99.7	99.5	99.7	100	-			
18. X13-2	98.7	99.2	97.6	91.5	91.5	92.0	99.2	98.9	98.9	96.0	99.2	98.1	98.9	98.7	98.9	99.2	99.2	-		
19. X13-3	98.9	99.5	97.9	91.8	91.8	92.3	99.5	99.2	99.2	96.2	99.5	98.4	99.2	98.9	99.2	99.5	99.5	98.7	-	
20. X13-4	98.9	99.5	97.9	91.8	91.8	92.3	99.5	99.2	99.2	96.2	99.5	98.4	99.2	98.9	99.2	99.5	99.5	98.7	98.9	-

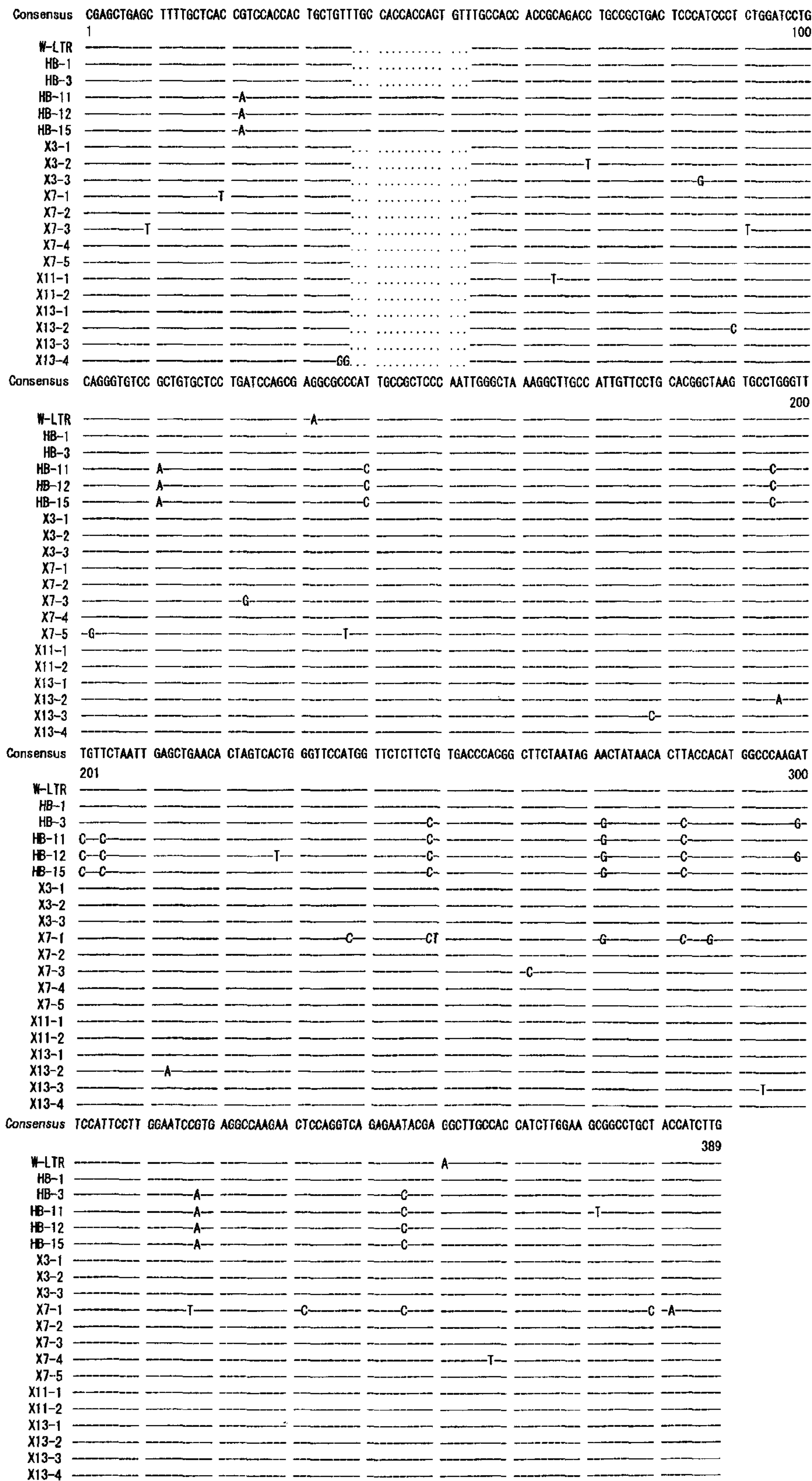

Fig. 1. Sequence alignments of the HERV-W LTR elements from the human brain cDNA library (HB-1, HB-3, HB-11, HB12, and HB-15) and the Xq21.3 region (X3-1, X3-2, X3-3, X7-1, X7-2, X7-3, X7-4, X7-5, X11-1, X11-2, X13-1, X13-2, X13-3, and X13-4). Consensus sequences are shown on the top row. Dashes indicate no change to the consensus sequences.

5, X11-1, X11-2, X13-1, X13-2, X13-3, and X13-4) were identified. Clones X3-1, X3-2, X3-3 were derived from the YAC clone 9AD1 and localized on STS v513R–sY23. Clones X7-1, X7-2, X7-3, X7-4, X7-5 were derived from the YAC clone 37GC10 and localized on STS 29GA2L–sY32. Clones X11-1, X11-2 were derived from the YAC clone 12AH2 and localized on STS 12AH2R–6EC2R. Clones X13-1, X13-2, X13-3, X13-4 were derived from the YAC clone 25BE9 and localized on STS sY46–sY47. Interestingly, although those clones were derived from different locations, clone X3-1 showed 100% sequence identity with those of clones X7-2, X11-2, and X13-1, suggesting that recent proliferation or intra-chromosomal translocation has occurred within the Xq21.3 region during hominid evolution. These LTR elements also showed a high degree of sequence similarity (96.2–99.5%) with that of HERV-W LTR (AF072500) (Table 1). In our previous study, only 4 HERV-K LTR elements in the Xq21.3 region could be identified [12], while 14 HERV-W LTR elements were identified in the same region, indicating that at least 3 times copies of the HERV-W LTR elements were higher than that of the HERV-K LTR elements in LTR frequency.

To understand the evolutionary relationship among HERV-W LTR elements, the LTRs were retrieved from the GenBank database and analyzed with new HERV-W LTR elements. A phylogenetic tree obtained by the neighbor-joining method revealed that the newly identified HERV-W LTR elements were closely related to the AX000960, AF072504, and AF072506 from the GenBank database (Fig. 2). Recently, several copy numbers of the HERV-W LTR elements were isolated from the human mammary carcinoma cell line T47D [28]. They divided into the two groups. The phylogenetic tree including GenBank data also shows a similar pattern, indicating that at least two different types of the HERV-W LTR elements were proliferated into the human genomes during hominid evolution. The examination of HERV-W LTR elements from the hominoid primates (chimpanzee, gorilla, orangutan, and gibbon) could be of great interest in further studies for understanding human evolution.

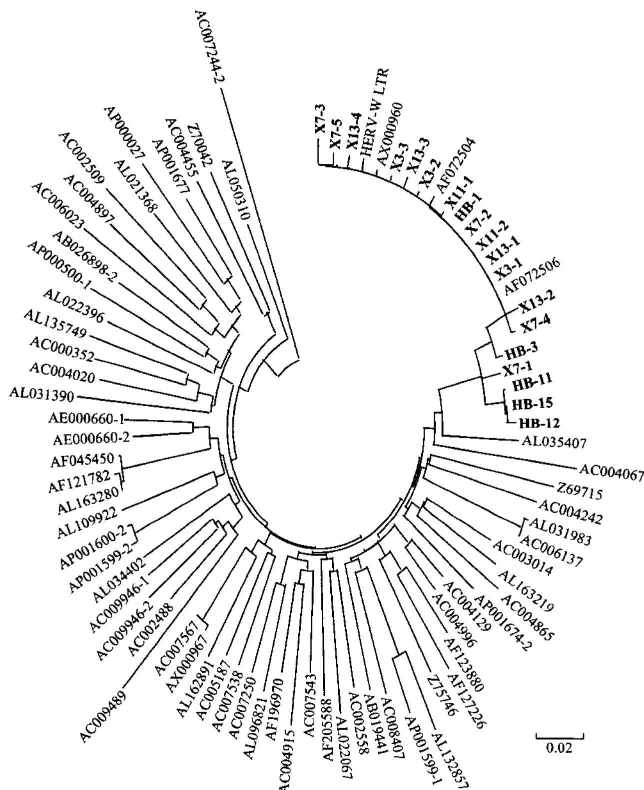

Fig. 2. Phylogenetic tree obtained from the neighbor-joining method for the LTR elements of the HERV-W family.
Branch lengths are proportional to the distances between the taxa.

Expression patterns of the HERV-W LTR elements varied in various cell lines (epidermal keratinocytes, liver cells, kidney cells, pancreatic cells, lymphocytes, and lung fibroblasts), in some cases showing strict cell type specificity [28]. The HERV-W LTR element (W8) showed highest promoter activity in LC5 cells, while the LTR element (W23) did not show activity in any cell line [28]. The HERV LTR elements could be useful for obtaining tissue-specific promoters. Akopov *et al.* [1] have noted that such sequences have the capacity to modify the expression of neighboring genes, and suggested that such modifications may have been acquired in the course of human evolution. From this point of view, investigation of the LTR elements is of great importance to understanding how those LTR elements and the neighboring genes were related to each other. Within the Xq21.3 region, the protocadherin (PCDHX) gene was identified recently [4]. The PCDHX gene was expressed predominantly in the human brain and located at STS sY43–GMGXY12. In this study, HERV-W LTR elements (clones X11-1, X11-2, X13-1, X13-2, X13-3, and X13-4) were very closely located to the PCDHX gene. The examination of the relationship between those LTR elements and the PCDHX gene should be clarified by further study.

The HERV-K-T47D-related LTR element mediates polyadenylation of cellular transcripts [3]. Such phenomenon was recently demonstrated in nucleosomal binding protein NSBP1 in Xq13.3 [20]. In the case of another retro-element (the HERV-F LTR element), a similar phenomenon was observed in relation to the Krüppel-related zinc finger gene ZNF195 [21]. A solitary HERV-K LTR element in the HLA DQ region (DQ-LTR3) resulted in type I diabetes mellitus affliction in 246 German and Belgian families [6]. The retroviral LTR element (DQLTR3) was a human-specific insertion [7]. This type of retroviral element also induced alternative splicing in the human leptin receptor [10]. The solitary HERV LTR elements showed that they retained detectable activity in human carcinoma cells, and

could direct the transcription in both orientations relative to the reporter gene [8]. Medstrand *et al.* [25] reported that LTR elements were used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. In this report, our new sequence data of the HERV-W LTR elements may contribute to an understanding of biological functions such as neuropsychiatric disease in the human brain and genomic instability of LTR integration in the human Xq21.3 region.

Acknowledgment

This study was supported by a grant of the Korea Health 21 R&D Project, Ministry of Health and Welfare, Republic of Korea (HMP-00-B-20200-0001).

REFERENCES

1. Akopov, S. B., L. G. Nikolaev, P. P. Khil, Y. B. Lebedev, and E. D. Sverdlov. 1998. Long terminal repeats of human endogenous retrovirus K family (HERV-K) specifically bind host cell nuclear proteins. *FEBS Lett.* **421:** 229–233.
2. Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and J. Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.
3. Baust, C., W. Seifarth, H. Germaier, R. Hehlmann, and C. Leib-Mosch. 2000. HERV-K-T47D-related long terminal repeats mediate polyadenylation of cellular transcripts. *Genomics* **66:** 98–103.
4. Blanco, P., C. A. Sargent, C. A. Boucher, M. Mitchel, and N. A. Affara. 2000. Conservation of PCDHX in mammals; expression of human X/Y genes predominantly in brain. *Mamm. Genome* **11:** 906–914.
5. Blond, J. L., F. Beseme, L. Duret, O. Bouton, F. Bedin, H. Perron, B. Mandran, and F. Mallet. 1999. Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family. *J. Virol.* **73:** 1175–1185.
6. Donner, H., R. R. Tonjes, B. Auwera, T. Siegmund, J. Braun, I. S. Weets, B. D. Registry, J. Herwig, R. Kurth, K. H. Usadel, and K. Badenhoop. 1999a. The presence or absence of a retroviral long terminal repeat influences the genetic risk for type I diabetes conferred by human leukocyte antigen DQ haplotypes. *J. Clin. Endocr. Metab.* **84:** 1404–1408.
7. Donner, H., R. R. Tonjes, R. E. Bontrop, R. Kurth, K. H. Usadel, and K. Badenhoop. 1999b. Intronic sequence motifs of HLA-DQB1 are shared between humans, apes and Old World monkeys, but a retroviral LTR element (DQLTR3) is human specific. *Tissue Antigens* **53:** 551–558.
8. Domansky, A. N., E. P. Kopantzev, E. V. Snezhkov, Y. B. Lebedev, C. Leib-Mosch, and E. D. Sverdlov. 2000. Solitary HERV-K LTR possess bi-directional promoter activity and contain a negative regulatory element in the U5 region. *FEBS Lett.* **472:** 191–195.
9. Garson, J. A., P. W. Tuke, P. Giraud, G. Paranhos-Baccaià, and P. Perron. 1998. Detection of virione-associated MSRV-RNA in serum of patients with multiple sclerosis. *Lancet* **351:** 33.
10. Kapitonov, V. V. and J. Jurka. 1999. The long terminal repeat of an endogenous retrovirus induces alternative splicing and encodes an additional carboxy-terminal sequence in the human leptin receptor. *J. Mol. Evol.* **48:** 248–251.
11. Karlsson, H., S. Bachmann, J. Schroder, J. McArther, E. F. Torrey, and R. H. Yolken. 2001. Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. *Proc. Natl. Acad. Sci. USA* **98:** 4634–4639.
12. Kim, H. S. and T. J. Crow. 2001. Cloning and phylogeny of endogenous retroviral elements belonging to the HERV-K LTR in cDNA library of human fetal brain and Xq21.3 region linked to psychosis. *Korean J. Genetics* **23:** 129–134.
13. Kim, H. S. 2001a. Sequence and phylogeny of HERV-W pol fragments. *AIDS Res. Hum. Retroviruses* **17:** 1665–1671.
14. Kim, H. S., Y. Chen, and P. Lonai. 1998. Complex regulation of multiple cytohesin like genes in murine tissues and cells. *FEBS Lett.* **433:** 312–316.
15. Kim, H. S. and W. H. Lee. 2001b. Human endogenous retrovirus HERV-W family: Chromosomal localization, identification, and phylogeny. *AIDS Res. Hum. Retroviruses* **17:** 643–648.
16. Kim, H. S., H. Hirai, and O. Takenaka. 1996. Molecular features of the TSPY gene of gibbons and Old World monkeys. *Chrom. Res.* **4:** 500–506.
17. Kim, H. S., O. Takenaka, and T. J. Crow. 1999. Isolation and phylogeny of endogenous retrovirus sequences belonging to the HERV-W family in primates. *J. Gen. Virol.* **80:** 2613–2619.
18. Kim, S. Y., H. S. Kang, and C. G. Lee. 2000. Phylogenetic analysis of hepatitis B virus genome isolated from Korean patient serum. *J. Microbiol. Biotechnol.* **10:** 823–828.
19. Kim, T. W., S. G. Min, D. H. Choi, J. S. Jo, and H. Y. Kim. 2000. Rapid identification of *Lactobacillus plantarum* in kimchi using polymerase chain reaction. *J. Microbiol. Biotechnol.* **10:** 881–884.
20. King, L. M. and C. A. Francomano. 2001. Characterization of a human gene encoding nucleosomal binding protein NSBP1. *Genomics* **71:** 163–173.
21. Kjellman, C., H. O. Sjogren, L. C. Salford, and B. Widegren. 1999. HERV-F (XA34) is a full-length human endogenous retrovirus expressed in placental and fetal tissues. *Gene* **239:** 99–107.
22. Kumar, S., K. Tamura, and N. Nei. 1993. MEGA: Molecular evolutionary genetics analysis, version 1.01. The Pennsylvania State University, University Park, PA 16802.
23. La Mantia, G., D. Maglione, G. Pengue, A. Di Cristofano, A. Simeone, L. Lanfrancone, and L. Lania. 1991. Identification and characterization of a novel human endogenous retroviral sequences preferentially expressed in undifferential embryonal carcinoma cells. *Nucleic Acids Res.* **19:** 1513–1520.

24. Laval, S. H., J. Dann, R. J. Butler, J. Loftus, J. Rue, S. J. Leask, N. Bass, M. Comazzi, A. Vita, S. Nanko, S. Shaw, P. Peterson, G. Shields, A. B. Smith, J. Stewart, L. E. DeLisi, and T. J. Crow. 1998. Evidence for linkage to psychosis and cerebral asymmetry (relative hand skill) on the X chromosome. *Am. J. Med. Genet. (Neuropsychiat. Genet.)* **81:** 420–427.
25. Medstrand, P., J. R. Landry, and D. L. Mager. 2001. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. *J. Biochem.* **276:** 1896–1903.
26. O'Connell, C., S. O'Brien, W. G. Nash, and M. Cohen. 1984. ERV3, a full-length human endogenous provirus: Chromosomal localization and evolutionary relationships. *Virology* **138:** 225–235.
27. Perron, H., J. A. Garson, F. Bedin, F. Beseme, G. Paranhos-Baccala, F. Komurian-Pradel, F. Mallet, P. W. Tuke, C. Voisset, J. L. Blond, B. Lalande, J. M. Seigneurin, B. Mandrand, and the collaborative research group on MS. 1997. Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. *Proc. Natl. Acad. Sci. USA* **94:** 7583–7588.
28. Schon, U., W. Seifarth, C. Baust, C. Hohenadl, V. Erfle, and C. Leib-Mosch. 2001. Cell type-specific expression and promoter activity of human endogenous retroviral long terminal repeats. *Virology* **279:** 280–291.
29. Sverdlov, E. D. 1998. Perpetually mobile footprints of ancient infections in the human genome. *FEBS Lett.* **428:** 1-6.
30. Varmus, H. E. 1982. Form and function of retroviral proviruses. *Science* **216:** 812–820.