CURRENT MICROBIOLOGY Vol. 44 (2002), pp. 216–220
DOI: 10.1007/s00284-001-0033-5

Current Microbiology
An International Journal

Molecular Cloning and Phylogenetic Analysis of New Human Endogenous Retrovirus HERV-W Family in Cancer Cells

Joo-Mi Yi,[1] Hwan-Mook Kim,[2] Won-Ho Lee,[1] Heui-Soo Kim[1]

[1]Division of Biological Sciences, College of Natural Sciences, Pusan National University, Pusan 609-735, Korea
[2]Laboratory of Biopotency Evaluation, Korea Research Institute of Bioscience and Biotechnology, Taejeon 305-600, Korea

Received: 12 March 2001 / Accepted: 12 July 2001

Abstract. A human endogenous retroviral family (HERV-W) has recently been described that is related to multiple sclerosis-associated retrovirus (MSRV) sequences. By using the PCR approach with human genomic DNA derived from cancer cell lines (HepG2, Jurkat, MCF7, UO-31), five *env* fragments of HERV-W family were newly identified and analyzed. They showed a high degree of nucleotide sequence similarity (94–99%) with that of the HERV-W. Translation of the *env* fragments showed no frameshift and termination codon by deletion/insertion or point mutation in clones HepG2-1 and JUR-3. The ratio of synonymous to non-synonymous substitutions indicated that negative selective pressure is acting on HepG2-1 and JUR-3 sequences. These *env* gene sequences could be associated with an active provirus in human cancer cells (HepG2 and Jurkat). The HepG2-1 and JUR-3 showed sister relationship with the HERV-W and W-7-1 derived from human Chromosome (Chr) 7. Phylogenetic analysis from the HERV-W family indicated close relationships of the *env* gene sequences across human chromosomes.

In humans and primates, endogenous retroviruses and related elements consist of an integral portion of the genome. Approximately 1% of the human genome is represented by human endogenous retroviruses (HERVs) and a much larger proportion by LINE and SINE elements, some of which have evolved from retroviral sequences [13]. They have been amplified in primates during evolution by repeated reintegration of reverse-transcribed mRNA into the genomic DNA of germline cells. Full-length retroviral sequences may interact with cellular oncogenes [14], and retroviral long terminal repeat (LTR) sequences have the capacity to exert a regulatory influence as promoters and enhancers of cellular genes [12].

A new family of endogenous retroviral sequences (HERV-W) has been identified by successive overlapping cDNA clones from human placenta [1]. *Gag* and *pol* open reading frames (ORFs) were interrupted by frameshifts and stop codons, whereas a complete ORF coding for an envelope was found. Homologies within the *pol* and *env* genes with murine type-C and simian type-D retroviruses, respectively, suggest a chimeric genome structure. In terms of its phylogenetic relationships, the HERV-W family is considered to be related to ERV-9 and RTLV-H families and to belong to endogenous retrovirus class 1 [3].

The HERV-W family may have relevance to the multiple sclerosis-associated retrovirus (MSRV) particles associated with multiple sclerosis [1]. The possibility that they have relevance to other neuropsychiatric conditions is raised by the finding of sequences homologous to HERV-W in a representational difference analysis of DNA from MZ twins discordant for schizophrenia [5], for which an endogenous retroviral etiology has previously been proposed [4]. Recently, chromosomal distributions of HERV-W *gag*, *pro*, and *env*-related sequences by Southern blot analysis were reported [15]. We determined fifteen members of *env* sequences of HERV-W family [7]. In this report, we additionally identified five HERV-W *env* families from human cancer cells and analyzed them with previous data.

Materials and Methods

Cell culture. Human cancer cells (HepG2, Jurkat, MCF7, UO-31) were grown in Dulbecco's medium (DMEM) supplemented with 10% heat-inactivated fetal calf serum, 2 mM glutamine, 1 mM nonessential amino

Correspondence to: H.-S. Kim; *email*: khs307@hyowan.cc.pusan.ac.kr

Table 1. Percentage similarity of nucleotide sequences in *env* fragments

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. HERV-W	—																				
2. W-1-1	92.8	—																			
3. W-3-6	92.8	89.8	—																		
4. W-3-7	94.8	92.2	91.9	—																	
5. W-3-8	93.3	90.4	91.3	92.4	—																
6. W-4-1	93.1	89.8	90.9	91.9	90.2	—															
7. W-5-6	93.5	90.0	90.7	91.9	91.1	91.3	—														
8. W-5-8	93.0	90.6	91.3	93.0	92.0	90.2	90.7	—													
9. W-6-3	95.0	92.4	92.8	95.2	93.3	92.2	92.0	92.4	—												
10. W-7-1	99.6	93.1	93.1	95.2	93.5	93.5	93.9	93.3	95.4	—											
11. W-12-2	94.6	92.6	90.9	93.3	91.3	92.6	91.9	91.5	93.5	95.0	—										
12. W-12-7	91.6	90.1	89.0	91.4	89.0	88.8	88.8	89.7	91.0	91.9	91.6	—									
13. W-14-1	94.8	91.9	92.2	94.3	92.2	91.9	92.6	92.2	93.9	95.2	93.5	91.0	—								
14. W-17-5	94.6	92.6	90.9	93.3	91.3	92.6	91.9	91.5	93.5	95.0	100	91.6	93.5	—							
15. W-20-9	93.7	92.0	90.9	93.1	90.7	90.9	90.7	91.5	93.3	94.1	93.7	91.2	93.7	93.7	—						
16. W-X-3	95.7	93.5	93.1	95.0	94.1	93.5	93.9	94.6	95.2	96.1	95.4	92.3	95.6	95.4	94.4	—					
17. **HepG2-1**	99.1	93.4	94.5	95.0	93.3	93.7	93.7	93.8	95.2	99.4	96.1	92.3	95.0	96.1	93.9	95.9	—				
18. **HepG2-2**	94.8	91.8	93.2	93.5	95.2	91.7	93.1	93.3	93.7	95.0	93.3	91.6	96.7	93.3	92.2	95.5	94.8	—			
19. **JUR-3**	98.3	94.2	94.3	95.3	93.7	93.7	93.7	94.2	95.5	98.7	96.8	93.6	95.0	96.8	94.4	96.3	99.3	94.8	—		
20. **MCF7-1**	94.6	97.9	92.6	94.0	91.7	91.9	91.4	92.8	93.8	94.9	95.9	93.2	93.4	95.9	93.4	95.3	95.5	92.9	96.2	—	
21. **UO31-3**	97.2	91.4	92.7	93.1	91.9	91.5	91.9	92.0	93.3	97.6	93.5	91.0	93.0	93.5	91.7	94.1	97.0	93.3	96.3	92.7	—

acids, 1 mM sodium pyruvate, 100 U/ml penicillin, 0.1 mg/ml streptomycin at 37°C, and 5% CO_2 incubator.

PCR amplification for HERV-W family. Using the polymerase chain reaction (PCR) approach, we identified five retroviral sequences belonging to the HERV-W *env* family from human cancer cells. New 583-bp *env* fragments of HERV-W family were amplified by the primer pair HS46 (5′-TCCCTGTACCTGAACAATGG-3′, bases 1360–1379) and HY76 (5′-CTTTCAGCGGTTAGCAAGTC-3′, bases 1920–1939) from the HERV-W (GenBank, accession no. AF072506). The PCR conditions followed were those of Kim et al. [8] with an annealing temperature of 56°C.

Molecular cloning of the PCR products. PCR products were separated on a 1.8% agarose gel, purified with the QIAEX II gel extraction kit (Qiagen) and cloned into the T-khs307 vector [9]. The cloned DNA was isolated by the alkali lysis method with the High Pure plasmid isolation kit (Boehringer Mannheim).

DNA sequencing and data analyses. Individual plasmid DNA was screened for inserts by PCR. Positive samples were subjected to sequence analyses on both strands with T7 and M13 reverse primers by using an automated DNA sequencer (Model 373A) and the DyeDeoxy terminator kit (Applied Biosystem). Nucleotide sequence analyses were performed with GAP, PILEUP, and PRETTY from the GCG package (University of Wisconsin). Neighbor-joining phylogenetic analysis was performed with the MEGA program [10].

Nucleotide sequence accession numbers. The nucleotide sequence data reported in this paper will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession numbers AB054085 (HepG2-1), AB054086 (HepG2-2), AB054087 (JUR-3), AB054088 (MCF7-1), AB054089 (UO31-1). The nucleotide sequences of HERV-W *env* gene, AB050996 (W-1-1), AB050997 (W-3-6), AB050998 (W-3-7), AB050999 (W-3-8), AB051000 (W-4-1), AB051001 (W-5-6), AB051002 (W-5-8), AB051003 (W-6-3), AB051004 (W-7-1), AB051005 (W-12-2), AB051006 (W-12-7), AB051007 (W-14-1), AB051008 (W-17-5), AB051009 (W-20-9), AB051010 (W-X-3), were taken from the GenBank database.

Results and Discussion

We identified five *env* gene sequences belonging to the HERV-W family from human cancer cells (HepG2, Jurkat, MCF7, UO-31), using a PCR approach. Two clones (HepG2-1 and HepG2-2) from HepG2, one clone (JUR-3) from Jurkat, one clone (MCF7-1) from MCF7, and one clone (UO31-3) from UO-31 were isolated and sequenced as HERV-W new members. As shown in Table 1, these sequences had 94.6–99.1% similarity to HERV-W *env* gene [1]. They also showed 91–99.4% sequence similarity to HERV-W *env* family identified from human monochromosomes in our previous study [7]. The HERV-W *env* family was detected on Chrs 1, 3, 4, 5, 6, 7, 12, 14, 17, 20, and X by analysis of PCR amplification by using the human monochromosomal DNA panel. Eight members of *env* fragments showed no frameshift and termination codon in their translation analysis. In the present study, among five *env* gene sequences of HERV-W from human cancer cells, no disruption by point mutations or insertions/deletions that inactivated the open reading frame by a frameshift or termination codon appeared in two sequences, HepG2-1 from HepG2 and JUR-3 from Jurkat. Table 2 shows the percentage identity of 180 amino acid sequences of *env* fragments among HERV-W family. The similarity ranged from

Table 2. Percentage identity of 180-amino-acid sequences of *env* fragments

	1	2	3	4	5	6	7	8	9	10
1. W-1-1	—									
2. W-3-8	84.9	—								
3. W-4-1	84.9	83.3	—							
4. W-7-1	91.1	86.1	88.9	—						
5. W-14-1	88.8	86.1	87.8	92.8	—					
6. W-17-5	89.9	85.5	88.8	91.6	91.1	—				
7. W-20-9	88.8	84.4	87.2	91.1	89.4	92.2	—			
8. W-X-3	89.9	90.6	91.1	93.3	93.9	92.7	91.1	—		
9. **HePG2-1**	91.6	86.6	89.4	99.4	93.3	92.2	91.6	93.9	—	
10. **JUR-3**	93.3	87.7	89.4	97.2	93.3	94.4	93.3	95.0	97.8	—

Table 3. Synonymous and non-synonymous substitutions in *env* fragments

	1	2	3	4	5	6	7	8	9	10
1. W-1-1	—	0.54	0.51	0.57	0.40	0.58	0.49	0.70	0.46	0.38
2. W-3-8	8.0/14.9	—	0.52	0.97	0.52	0.70	0.64	0.70	0.76	0.69
3. W-4-1	8.5/16.7	8.6/16.6	—	0.49	0.41	0.53	0.48	0.48	0.46	0.46
4. W-7-1	5.5/9.7	6.7/6.9	5.4/11.0	—	0.53	0.68	0.68	1.11	0.20	0.87
5. W-14-1	5.9/14.6	6.7/13.0	6.4/15.7	4.1/7.7	—	0.54	0.79	0.45	0.41	0.41
6. W-17-5	5.6/9.6	7.8/11.1	5.9/11.1	4.1/6.0	5.1/9.4	—	0.60	0.91	0.84	0.62
7. W-20-9	6.2/12.6	8.9/13.9	7.7/16.1	5.5/8.1	6.2/7.8	5.1/8.5	—	0.93	0.55	0.46
8. W-X-3	5.6/8.0	5.6/8.0	5.4/11.2	4.1/3.7	3.5/7.8	4.1/4.5	5.7/6.1	—	0.72	0.62
9. HepG2-1	5.2/11.4	6.5/8.5	5.1/11.0	0.3/1.5	3.8/9.3	3.8/4.5	5.3/9.7	3.8/5.3	—	0.00
10. JUR-3	4.2/11.2	5.9/8.5	5.1/11.0	1.3/1.5	3.8/9.3	2.8/4.5	4.5/9.7	3.3/5.3	0.5/0	—

Percentage of synonymous substitutions per site (Ks) and non-synonymous substitutions per site (Ka) are shown below the diagonal in the form Ka/Ks. The ratios are shown above the diagonal.

83.3% to 99.4%. We also analyzed synonymous and non-synonymous substitutions within the *env* fragments of the HERV-W family in order to discover the evolutionary forces at work. As shown in Table 3, the mean synonymous substitutions (Ks) ranged from 0% to 16.7%, whereas the mean non-synonymous substitutions (Ka) ranged from 0.3% to 8.9%. In terms of the Ka/Ks ratio, 97.8% of the values in pairwise comparisons were < 1. These *env* gene sequences could, therefore, be associated with an active provirus in human genomes and with pathological implications for neuropsychiatric diseases or cancers. Retroviral particles have been recovered from monocyte cultures from patients with multiple sclerosis [11]. Moreover, multiple sclerosis-associated retrovirus (MSRV) has been reported in the serum of patients with the disease [6]. The MSRV associated with multiple sclerosis was defined as HERV-W family [1]. An *env* cDNA of HERV-W containing a complete open reading frame was found to be specifically expressed in placenta cells [2]. In addition, the 538-amino acid HERV-W *env* gene was shown to encode a putative 80-kDa glycosylated protein by an in vitro transcription-translation assay [15].

In order to understand the evolutionary relationship within the HERV-W family, a phylogenetic tree was constructed by using the neighbor-joining method (Fig. 1). Clones HepG2-1 and JUR-3, which are not interrupted by a premature stop codon, showed very close relationship with HERV-W (clone cl.PH74, accession no. AF072507) on human Chr 7. Clone W-7-1, identified from Chr 7 in our previous study [7], is also grouped with HepG2-1, JUR-3, and HERV-W (AF072507). This allows us to speculate that there are at least four copies of open reading frame putatively encoding an envelope protein of HERV-W family on human Chr 7. Voisset et al. [15] suggested that the human haploid genome contains at least 30 copies of *env* gene by Southern blot analysis. We identified and sequenced 15 *env* fragments by using a human monochromosomal DNA panel [7]. In this report, we isolated five new *env* gene sequences belonging to the HERV-W family. Clone MCF7-1 is closely related to W-1-1 from Chr 1, while clone HepG2-2 is closely related to W-3-8 from Chr 3. The phylogenetic data indicated that the HERV-W family showed close relationships of the *env* gene sequences across human chromosomes.

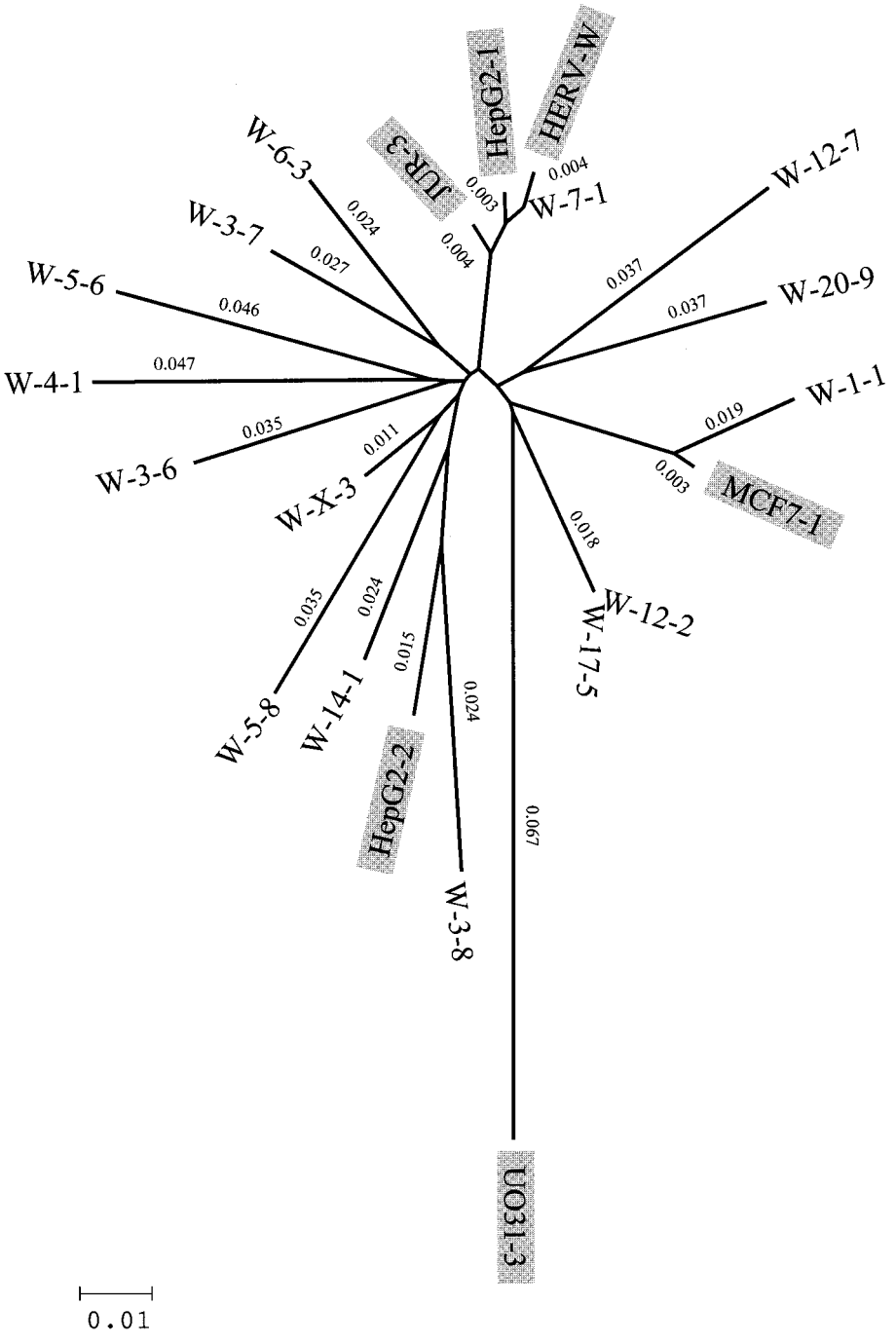

Fig. 1. Phylogenetic tree obtained by neighbor-joining method for the *env* fragments of the HERV-W family. Branch lengths are proportional to the distances between the taxa. The distance values are also represented on each branch.

ACKNOWLEDGMENT

This work was supported by Korea Research Foundation Grant (KRF-2000-015-DP0342).

Literature Cited

1. Blond JL, Beseme F, Duret L, Bouton O, Bedin F, Perron H, Mandran B, Mallet F (1999) Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family. J Virol 73:1175–1185
2. Blond JL, Lavillette D, Cheynet V, Bouton O, Oriol G, Chapel-Fernandes S, Mandrand B, Mallet F, Cosset FL (2000) An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. J Virol 74:3321–3329
3. Boeke JD, Stoye JP (1997) Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In: Coffin JM, Hughes, SH, Varmus HE, eds. Retroviruses Cold Spring Harbor Laboratory, NY: Cold Spring Harbor Laboratory Press, pp 343–435
4. Crow TJ (1984) A re-evaluation of the viral hypothesis: is psy-

chosis the result of retroviral integration at a site close to the cerebral dominance gene? Br J Psychiatr 145:243–253
5. Deb-Rinker P, Klempan TA, O'Reilly RL, Torrey EF, Singh SM (1999) Molecular characterization of a MSRV-like sequence identified by RDA from monozygotic twin pairs discordant for schizophrenia. Genomics 61:133–144
6. Garson JA, Tuke PW, Giraud P, Paranhos-Baccaia G, Perron P (1998) Detection of virion-associated MSRV-RNA in serum of patients with multiple sclerosis. Lancet 351:33
7. Kim H-S, Lee W-H (2001) Human endogenous retrovirus HERV-W family: chromosomal localization, identification, and phylogeny. AIDS Res Hum Retroviruses 17:643–648
8. Kim H-S, Hirai H, Takenaka O (1996) Molecular features of the TSPY gene of gibbons and Old World monkeys. Chromosome Res 4:500–506
9. Kim H-S, Chen Y, Lonai P (1998) Complex regulation of multiple cytohesin like genes in murine tissues and cells. FEBS Lett 433: 312–316
10. Kumar S, Tamura K, Nei N (1993) MEGA: Molecular evolutionary genetics analysis, version 1.01. University Park, PA: The Pennsylvania State University
11. Perron H, Garson JA, Bedin F, Beseme F, Paranhos-Baccala G, Komurian-Pradel F, Mallet F, Tuke PW, Voisset C, Blond JL, Lalande B, Seigneurin JM, Mandrand B, 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
12. Schon U, Seifarth W, Baust C, Hohenadl C, Erfle V, Leib-Mosch C (2001) Cell type-specific expression and promoter activity of human endogenous retroviral long terminal repeats. Virology 279: 280–291
13. Sverdlov ED (1988) Perpetually mobile footprints of ancient infections in the human genome. FEBS Lett 428:1–6
14. Varmus HE (1982) Form and function of retroviral proviruses. Science 216:812–820
15. Voisset C, Bouton O, Bedin F, Duret L, Mandrand B, Mallet F, Paranhos-Baccala G (2000) Chromosomal distribution and coding capacity of the human endogenous retrovirus HERV-W family. AIDS Res Hum Retroviruses 16:731–740