



Identification and Expression Analyses of Equine Endogenous Retroviruses in Horses

Jeong-An Gim^{1,2,3}, and Heui-Soo Kim^{1,2,*}

¹Department of Biological Sciences, College of Natural Sciences, Pusan National University, Busan 46241, Korea, ²Institute of Systems Biology, Pusan National University, Busan 46241, Korea, ³The Genomics Institute, Life Sciences Department, UNIST, Ulsan 44919, Korea

*Correspondence: khs307@pusan.ac.kr

<http://dx.doi.org/10.14348/molcells.2017.0141>

www.molcells.org

Endogenous retroviruses (ERVs) have been integrated into vertebrate genomes and have momentarily affected host organisms. Horses (*Equus caballus*) have been domesticated and selected for elite racing ability over centuries. ERVs played an important role in the evolutionary diversification of the horse genome. In the present study, we identified six equine ERV families (EqERVs-E1, I1, M2, P1, S1, and Y4), their full-length viral open reading frames (ORFs), and elucidated their phylogenetic relationships. The divergence time of EqERV families assuming an evolutionary rate of 0.2%/Myr indicated that EqERV-S3 (75.4 million years ago; mya) on chromosome 10 is an old EqERV family and EqERV-P5 (1.2 Mya) on chromosome 12 is a young member. During the evolutionary diversification of horses, the EqERV-I family diverged 1.7 Mya to 38.7 Mya. Reverse transcription quantitative real-time PCR (RT-qPCR) amplification of EqERV *pol* genes showed greater expression in the cerebellum of the Jeju horse than the Thoroughbred horse. These results could contribute further dynamic studies for horse genome in relation to EqERV gene function.

Keywords: EqERV, Jeju horse, *pol* gene, quantitative real-time RT-PCR, thoroughbred horse

INTRODUCTION

Modern horses (*Equus caballus*) have evolved for over 50

million years from the genus *Eohippus*. Since 4000 B.C., horses have been selected by humans on the basis of strength, speed, and endurance (Gu et al., 2009). A large number of horse subspecies have been domesticated and bred in regard to improvement of improve speed, endurance, and agility. Only one wild subspecies (Przewalski's horse) has survived to the present day (Lau et al., 2009; Orlando et al., 2013). Thoroughbred horses have been specifically bred for racing ability; therefore, many studies have examined genetic traits related to athletic performance (Bower et al., 2012; Hill et al., 2010; Park et al., 2012). The Jeju breed of horses has been developed over centuries by interbreeding the native horses of Jeju, Korea with the 160 Mongolian horses introduced by the invading Mongolian army in 1276. These horses have been bred for farming, as a food source, and for racing in the Jeju province in South Korea. Jeju horses are small; however, they are robust and possess a high level of endurance. Many studies have focused on elucidating the phylogenetic relationships as well as genetic characteristics of Jeju horses (Cho, 2007; Shin et al., 2002). Research on Thoroughbred horses has focused on improving racing ability. Therefore, genetic research on the Jeju horse may help predicting of phylogenetic relationships and identifying traits specific to the Jeju horse. New approaches are required for the investigation of racing ability, subspecies-specific traits, evolutionary traits, as well as the phylogenetic diversity of horse subspecies.

Retroviral infections resulting in germline integration have

Received 24 July, 2017; revised 20 August, 2017; accepted 24 August, 2017; published online 17 October, 2017

eISSN: 0219-1032

© The Korean Society for Molecular and Cellular Biology. All rights reserved.

©This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>.

an important role in vertebrate evolutionary diversification. Endogenous retrovirus (ERV) sequences account for about 5-10% of host genome and have been identified in all mammals and most vertebrates. Following integration into the host genome, retroviruses are transmitted from ancestors to progeny according to Mendelian inheritance principles and eventually form repeat elements (long terminal repeat (LTR) elements) (Blikstad et al., 2008; Jern and Coffin, 2008). The full-length ERV genome comprises *gag* (group-specific antigen), *prt* (protease), *pol* (polymerase), and *env* (envelope) genes as well as LTR elements at both ends. Primer binding sites (PBS) also exist between the 5'-LTR and the *gag* gene. Most ERVs display a truncated LTR, although some ERVs display a full-length sequence. These ERV sequences induce genomic instability via mutational events, such as retrotranspositions and genomic rearrangements (Esnault et al., 2005; Hughes and Coffin, 2001). When two LTR elements are integrated into the host genome, they initially display identical sequences that subsequently diverge at a rate of 0.13-0.21% over a period of a million years. Therefore, it is possible to estimate the point in time of integration by assessing the divergence between two LTR sequences (Tristem, 2000). ERVs are classified according to their PBS that are highly homologous to host tRNA complement sequences (Tristem, 2000).

Human endogenous retroviruses (HERVs) have many well-known functions, e.g., they are potential pathogens, alternative LTR promoters, and are involved in syncytin production in humans (Dunn et al., 2006; Malassiné et al., 2005; Ruprecht et al., 2008). Horses also have ERVs, referred to as equine ERVs (EqERVs) (Brown et al., 2012; Garcia-Etxebarria and Jugo, 2012; Jern and Coffin, 2008; van der Kuyl, 2011). In the equine genome, full-length beta-EqERV genomes have been identified (van der Kuyl, 2011). Fifteen EqERV types belonging to three classes were detected by *in silico* methods (Garcia-Etxebarria and Jugo, 2012). ERVs have the potential to alter the phenotype of different species, and EqERVs may also provide evolutionary information related to the traits of horses (Gifford and Tristem, 2003; Jern and Coffin, 2008). According to transcriptome analyses, EqERV-mRNAs are expressed in various horse tissues (Brown et al., 2012). The expression of ERVs has previously been identified by the detection of ERV-derived transcripts (Ahn and Kim, 2009; Perron et al., 2008). Some ERV coding regions (*gag*, *pol*, and *env*) have complete open reading frames (ORFs); interrupted ERV ORFs have also been detected. ERV genes are transcribed at high levels, and corresponding proteins are expressed more frequently in cancer than in normal tissues (Ahn and Kim, 2009; Kang et al., 2014; Ruprecht et al., 2008). Some ERV families tend to exhibit ubiquitous expression in various organs, while others show tissue-specific expression patterns (Ahn and Kim, 2009; Ahn et al., 2011b). We identified EqERVs in the horse genome and characterized their genomic organization. To date, no studies have been undertaken to evaluate the expression patterns of EqERVs in horse samples. Thus, the EqERV expression patterns were examined in the cerebrum and cerebellum of a Thoroughbred as well as a Jeju horse.

MATERIALS AND METHODS

Ethics statement

All animal trials approved by the Pusan National University-Institutional Animal Care and Use Committee (PNU-IACUC) (Approval Number for horse tissue samples: PNU-2013-0411). All biopsies were taken under the guidance of a veterinarian. Efforts were made to minimize animal stress and pain.

Horse samples

Cerebrum and Cerebellum tissue of individual Thoroughbred and Jeju horses obtained from the Subtropical Animal Experiment Station of the Rural Development Administration, South Korea. Total RNA was extracted using a standard TRIzol®-based purification protocol (Invitrogen, USA) and quantified using NanoDrop technology (Thermo Fisher Scientific Inc., USA). Then, RNA was adjusted to concentrations of 500 ng per samples. M-MLV (Moloney-Murine Leukemia Virus) reverse transcriptase (Promega, USA) was used for cDNA synthesis according to instructions of the manufacturer (Promega).

Identification of EqERV families

The *Equus caballus* (equCab2) genome assembly of a Thoroughbred horse was retrieved from NCBI Assembly (<http://www.ncbi.nlm.nih.gov/assembly/286598/>). In UCSC Genome Browser, BLAT was used to search for ERVs in the horse genome, and well-known full-length retrovirus DNA sequences were used as search queries. The resulting sequences were expanded to include upstream and downstream regions for total lengths of 1.5x, then retrieved ERV-sequences were queried using RetroTector (<http://retrotector.neuro.uu.se/>). EqERV was used to refer to the ERVs described in this work based on the previously described nomenclature (Garcia-Etxebarria and Jugo, 2012).

Bioinformatic analysis

EqERVs were classified according to the PBS and identified by length. Each LTR sequence was aligned using ClustalW in MEGA7 (<http://www.megasoftware.net/>). The point in time of integration of EqERVs was assessed based on the divergence of each LTR (Johnson and Coffin, 1999). The tRNA sequences were downloaded from the genomic tRNA database (<http://gtrnadb.ucsc.edu>) and the PBS sites were aligned with tRNA sequences in BioEdit (Hall, 1999). The PBS sites not matching with tRNA sequences were realigned using the RetroTector PBS site (Sperber et al., 2009). Sequence divergences between two LTR sequences were assessed by the Kimura two-parameter method in MEGA7 (Kimura, 1980). Neighbor-joining trees were constructed using MEGA7, with 100 bootstrap replicates (Kimura, 1980). The percentage of bootstrap replicates supporting the branch is revealed at each node. The ORFs of *gag*, *pol*, *env* region were retrieved by a BLASTx search (Johnson et al., 2008).

RT-qPCR amplification

EqERV *pol* transcripts were detected by reverse transcription

quantitative real-time PCR (RT-qPCR) using the primers shown in Table 1. We designed primer sets using the Primer3 program (<http://bioinfo.ut.ee/primer3-0.4.0/>). The RT regions of the EqERV *pol* gene were amplified. Each reaction only yielded one amplification product as determined by melting curve analyses. No-template controls were consistently tested negative. For normalization, the beta-2-microglobulin (*B2M*) gene was amplified as previously described (Ahn et al., 2011a). RT-qPCR amplification was performed as follows: 30 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 30 s, and a final elongation at 72°C for 7 min using a Mastercycler® Pro S Thermocycler (Effendorf, Hamburg, Germany). The conditions for the RT-qPCR amplification of EqERV elements and the *B2M* gene were 45 cycles of 95°C for 10 s, 58°C for 15 s, and 72°C for 15 s. A melting curve analysis was performed for 30 s at 55°C to 99°C using a Rotor-Gene Q (Qiagen, Hilden, Germany). All samples were amplified in triplicate to guarantee reproducibility.

RESULTS AND DISCUSSION

Identification and point of divergence of EqERVs in the horse genome

We analyzed EqERVs in the horse genome by an *in silico* method using the RetroTector. We confirmed the locations in the genome, retroviral genes, PBS regions, and LTR pairs of each EqERV family. In the horse whole genome, a total of 22 EqERV types were detected by PBS analysis (EqERV-E, I, M, P, S, and Y) (Fig. 1 and Table 2). Based on a comparative analysis of the reverse complement of PBS and tRNA sequences, we determined the first amino acid combined with the tRNA, and assigned the following EqERV family names: EqERV-E1, EqERV-I1, EqERV-M2, EqERV-P1, EqERV-S1, and EqERV-Y4 (Fig. 2). In all EqERV families, the surface (SU) component of the *env* gene was interrupted. The matrix (MA) domain of the *gag* gene was only detected, and full-length ORFs of viral genes (*gag* and *env*) were identified in EqERV-E1. The *prt* gene was truncated in EqERV-P1, EqERV-

S1, and EqERV-Y4. EqERV-P1 and EqERV-S1 showed a long capsid (CA) domain compared to viruses of other families, however full-length ORFs of the *gag* gene were not detected. Additionally, long integrase (IN) and transmembrane (TM) domains were detected in EqERV-P1, while a long reverse transcriptase (RT) domain in the *pol* region was observed for EqERV-I1. The EqERV-S1 has full-length ORFs of three viral genes (*gag*, *pol*, and *env*), but other EqERVs has at least one interrupted ORFs in these sequences (Fig. 1). In the host, the activation of restriction factors was not need to full-length ORFs, and interrupted ORFs could regulate immune systems or other biological roles (Malfavon-Borja and Feschotte, 2015). Thus, all these EqERV domains appear to have important biological roles in horses.

Both LTRs located in the 5' and 3' flanking regions are important for retroviral replication. We analyzed the LTR sequences to estimate the point in time of integration based on divergence values, because both LTR sequences can help to estimate their insertion time. The same LTR sequences were attached to each 5'- and 3'-end during retroviral replication, and differences in these LTR sequences arise during evolution. Therefore, it could be possible to estimate the moment of integration time of ERVs based on the divergence values (Johnson and Coffin, 1999). Garcia-Etxebarria and Jugo estimated the point in time of integration at 45-21 million years ago (Mya) and 6-3 Mya (Garcia-Etxebarria and Jugo, 2012). Thus, these two types are the youngest members of class I. Our estimated LTR divergence values indicated that EqERV-Y1 and Y3 proliferated approximately 35.9 Mya and 2.8 Mya based on average evolutionary rate of 0.2% per million years (Myr) (Table 2). This evolutionary rate was also applied to ERV-H divergence in primates (Anderssen et al., 1997). Depend on the species, different evolutionary rates (0.12%/Myr, 0.2%/Myr, and 0.26%/Myr) could be used for ERV studies in humans and apes (Lebedev et al., 2000). In our previous studies, the average evolutionary rate of 0.2%/Myr was used for humans and primates (Yi and Kim, 2006; Yi et al., 2007). After that, integration time

Table 1. List of RT-qPCR primers used for the amplification of the EqERV *pol* gene in Thoroughbred and Jeju horses

EqERV family	Direction	Primer sequences	Product size (bp)	Location in genome
EqERV-E1 <i>pol</i>	Forward	GGTACAGAGAGGGAGGCACA	132	chr1:183,840,430-183,840,561
	Reverse	TTGAGGCTGATTGCTTTTGA		
EqERV-I1 <i>pol</i>	Forward	ACCCATCTGCACTGAAATC	140	chrX:74,309,270-74,309,383
	Reverse	CCCAGTTGAGGGTTAGGG		
EqERV-P1 <i>pol</i>	Forward	TGTGGGTCCTTCTAGTTGTGG	134	chr7:49,472,758-49,472,891
	Reverse	CGAGGGGTTAAGTTCACACG		
EqERV-M2 <i>pol</i>	Forward	TGGA AAAAGGCAAAGACAAA	125	chr5:16,128,363-16,128,487
	Reverse	CTGAAGTGGCCACCAACTTT		
EqERV-S1 <i>pol</i>	Forward	CATGGCACTGCTCATCAAC	114	chrX:95,866,328-95,866,441
	Reverse	TTCTTTTTTCTCCCAAAGC		
EqERV-Y4 <i>pol</i>	Forward	GGGAGGTCAGAGCCTTGTTT	142	chr1:29,481,664-29,481,805
	Reverse	ACGTCCTTTCTTGCCATCC		
<i>B2M</i> (for normalization)	Forward	CCTGCTCGGGCTACTCTC	89	chr1:144,494,425-144,497,779
	Reverse	CATTCTGCTGGGTGACG		

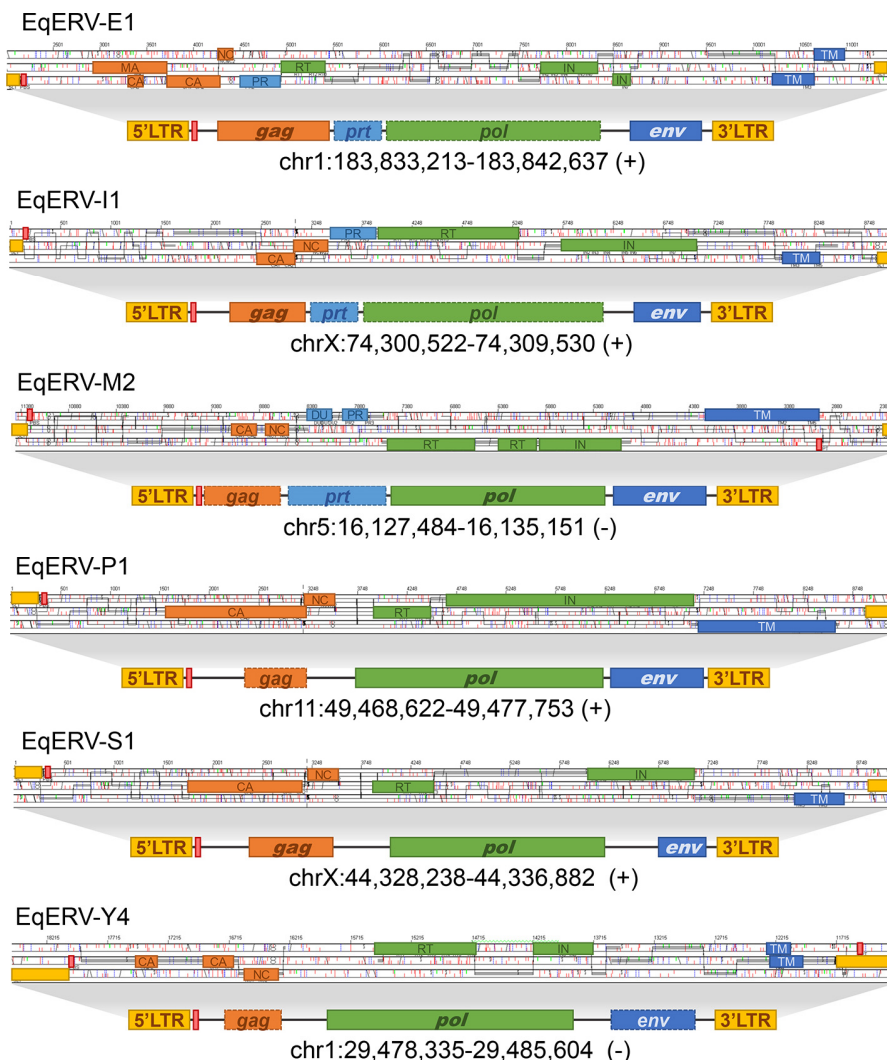


Fig. 1. Structural analysis of full-length EqERVs located at specific loci on horse chromosomes. The depicted structures were analyzed using EqERV sequences from the horse genome equCab2 assembly with the RetroTector10 program. The interrupted regions and full-length ORFs were indicated as dot lines and solid lines with bold type, respectively. These EqERVs harbor a 5' LTR-*gag*(*prt*)-*pol*-*env*-3' LTR structure.

based on ERV was estimated in novel methods (Martins and Villesen, 2011), and various species in mammals (Lee et al., 2013). In mammalian genome, insertion time is estimated by the divergence between 5'LTR and 3'LTR (Lee et al., 2013). As shown in Table 2, we estimated the divergence time of EqERV families assuming an evolutionary rate of 0.2%/Myr. Divergence time was estimated as 49.1 Mya in EqERV-Y2 and 5.6 Mya in EqERV-Y4. As an old EqERV family, EqERV-S3 (75.4 Mya) was detected on chromosome 10, and EqERV-P5 (1.2 Mya) was identified as a young member on chromosome 12. During the evolutionary diversification of horses, the diversification of the EqERV-I family took 1.7 Mya to 38.7 Mya (Table 2). The evolution of the horse species occurred over a period of 55 million years (Macfadden, 2005). According to our data and a previous report (Garcia-Etxebarria and Jugo, 2012), some EqERVs integrated into

the horse genome during the evolution of the horse lineage. The *Equus* genus consists of three clades. One clade is the domesticated horse, including Thoroughbred, Arab, Jeju, and Mongolian horse (*E. caballus*) breeds. The other two clades comprise zebras and donkeys. These clades split 3 million years ago (Macfadden, 2005; Oakenfull et al., 2000). Most EqERV integrations occurred earlier, as indicated by herein presented and other published data (Garcia-Etxebarria and Jugo, 2012). The HERV-S family has been detected in hominoids, Old World monkeys, and New World monkeys. However, it does not exist in prosimians, indicating that HERV-S integrated into the primate genome approximately 43 million years ago (Yi et al., 2004). Taken together, EqERV families integrated at different points in time, and therefore variant evolutionary patterns are observed in the equine genome.

Table 2. Divergence times and evolutionary rates of EqERV families

EqERV Family	*PBS	Location (equCab2)	Strand	Length (kb)	**Structure	Divergence (%)	Evolutionary time		
							r=0.3%/ Myr	r=0.2%/ Myr	r=0.15%/ Myr
EqERV-E1	Glu	chr1:183,833,213-183,842,637	(+)	9.4	Full	16.6	27.7	41.6	55.5
EqERV-I1	Ile	chrX:74,300,522-74,309,530	(+)	9.0	Full	2.9	4.8	7.2	9.5
EqERV-I2	Ile	chr11:46,777,889-46,786,573	(-)	8.7	Full	1.2	1.9	2.9	3.9
EqERV-I3	Ile	chr5:27,325,566-27,334,031	(-)	8.5	Full	0.7	1.1	1.7	2.2
EqERV-I4	Ile	chr7:47,466,052-47,474,526	(-)	8.5	Full	9.4	15.6	23.5	31.3
EqERV-I5	Ile	chr11:55,535,777-55,544,146	(-)	8.3	Full	15.5	25.8	38.7	51.6
EqERV-I6	Ile	chr1:111,609,131-111,617,366	(+)	8.2	Full	8.0	13.3	20.0	26.7
EqERV-I7	Ile	chr4:105,090,068-105,096,011	(-)	5.9	Δ3'LTR	Not detected	-	-	-
EqERV-M1	Met	chr7:43,693,479-43,702,540	(-)	9.1	Full	2.7	4.5	6.8	9.1
EqERV-M2	Met	chr5:16,127,484-16,135,151	(-)	7.7	Δ3'LTR	Not detected	-	-	-
EqERV-P1	Pro	chr11:49,468,622-49,477,753	(+)	9.1	Full	1.7	2.9	4.4	5.8
EqERV-P2	Pro	chr2:11,712,296-11,721,196	(+)	8.9	Full	2.4	4.1	6.1	8.1
EqERV-P3	Pro	chr4:26,026,629-26,034,772	(+)	8.1	Δ3'LTR	Not detected	-	-	-
EqERV-P4	Pro	chr10:26,653,364-26,660,774	(-)	7.4	Δ3'LTR	Not detected	-	-	-
EqERV-P5	Pro	chr12:15,551,778-15,559,089	(-)	7.3	Full	0.5	0.8	1.2	1.5
EqERV-S1	Ser	chrX:44,328,238-44,336,882	(+)	8.6	Full	1.3	2.1	3.1	4.2
EqERV-S2	Ser	chr20:32,762,746-32,770,917	(-)	8.2	Full	30.2	50.3	75.4	100.5
EqERV-S3	Ser	chr10:12,943,738-12,950,180	(+)	6.4	Δenv	10.6	17.7	26.6	35.4
EqERV-Y1	Tyr	chrX:53,124,155-53,133,636	(+)	9.5	Full	14.4	23.9	35.9	47.9
EqERV-Y2	Tyr	chrX:60,642,587-60,651,514	(+)	8.9	Full	19.6	32.7	49.1	65.5
EqERV-Y3	Tyr	chr1:27,393,645-27,402,113	(+)	8.5	Full	1.1	1.9	2.8	3.7
EqERV-Y4	Tyr	chr1:29,478,335-29,485,604	(-)	7.2	Full	2.2	3.7	5.6	7.4

*PBS: Primer binding site, **Full: full-length, Δdeletion. Myr: million years

Divergence was estimated by the Kimura-2-parameter method. Evolutionary time was calculated following as $T=d/2r$ (T: evolutionary time; d: divergence; r: evolutionary rate).

A. EqERV-E1

```

5LTR TGCCACAGGTCCTGCCGATGCTATTCTGAAATAAGGAGCACTACCAACGACCTTGAGAGTCCAGAAATCTTTC
3LTR TGCTCATGAATCCTGCCCATGCTATCCAGAGTCAAAAGAGCACTACCAACGACCTTGAGAGTCCAGAAATCTTTC
5LTR TTTTGACTCCTCTGGCTGCCAGCCCA
3LTR TTTTGACTCCTCTGGCTGCCAGCCCA
    
```

Reverse complement of PBS TCCCTGTaCAGGcgACCA
 Horse chr17 tRNA76-Glu (CTC) UCCCCGguCAGGguACCA

B. EqERV-I1

```

5LTR TGCCCATGTCTCCTGTCGCCATGCTATCCACACTATTTCTCTAAATAAAGAGCACTACTGCCAGATCTTGAGAGTCTCA
3LTR TGCCCATGGTCTCCTGCCATGCTATCCACACTATTTCTCTAAATAAAGAGCACTACTGCCAGATCTTGAGAGTCTCA
5LTR CAATCTTTCTTTTGGACTCTCTGGCTCA
3LTR GAAATCTTTCTTTGGACTCTCTGGCTCA
    
```

Reverse complement of PBS aGATCCCTGACGGGCCA
 Horse chr11 tRNA58-Ile (AAT) cGATCCCTGACGGGCCA

C. EqERV-M2

```

5LTR TGTAGA---AGCTTCGACATAGTTTTGAGATATCTA---TAAGTTCGCCGTGACTTGACTGTCTGACCCCGG----
3LTR TGTATACTTATTTCTTTTTTCTTGAAGATATAGTTTAACTTTTCAGCCCTTACTTACTTAACAATATGATAC
5LTR TGGCTACGGTACC CGGTGTGGCA-----
3LTR TTTTCTCCGGCAGTGTACTTAAGGGAACCGTTAGCTCTACA
    
```

Reverse complement of PBS CaTgaCAaetTgTAAcCa
 Horse chr20 tRNA39-Met (CAT) CgtTgCAgctTgaTAAcTc

D. EqERV-P1

```

5LTR TGAATCATTTAATGACTTTTTACATCATCTTTCTTTTTTCCAGTAAAAATAAGTCACTGACCCATGCCCTATAAA
3LTR TGAATCATTTAATGACTTTTTACATCATCTTTCTTTTTTCCAGTAAAAATAAGTCACTGACCCATGCCCTATAAA
5LTR ATTAGCCCTAACCTCAACTCAGGGCAGCAGGAGCTCTGACTGCCCTGGGTCTTGTCCCAACACAGCTCTGCC
3LTR ATTAGCCCTAACCTCAACTCAGGGCAGCAGGAGCTCTGACTGCCCTGGGTCTTGTCCCAACACAGCTCTGCC
5LTR TGCCCATGGTCTCCTGCCATGCCAGCGGGGAGCAGCAGGAGCTCTGACCCCGGTGGTCTCTGCCCA
3LTR TGCCCATGGTCTCCTGCCATGCCAGCGGGGAGCAGCAGGAGCTCTGACCCCGGTGGTCTCTGCCCA
    
```

Reverse complement of PBS GcaTGGGgacAgGaccCa
 Horse chr1 tRNA111-Pro (AGG) GueTGGGguAtGAuuCu

E. EqERV-S1

```

5LTR TGCCATTTACTCAATCCGGTTGATCTTATCTAAAAGTTGTGGGACCGCCAAATGGCCAGCCCTATGGGCACTGATAT
3LTR TGCCATTTACTCAATCCGGTTGATCTTATCTAAAAGTTGTGGGACCGCCAAATGGCCAGCCCTATGGGCACTGATAT
5LTR CATTTTAACTTTTACATATCTTTGCTTGTGAAAGATAACTCACATACCTATGCCCTAAATTTAGCTCTAACCTTC
3LTR CATTTTAACTTTTACATATCTTTGCTTGTGAAAGATAACTCACATACCTATGCCCTAAATTTAGCTCTAACCTTC
5LTR A
3LTR A
    
```

Reverse complement of PBS AgagCTTctGcCgCCa
 Horse chr1 tRNA82-Ser (GCT) AuccCaUccuGguCgCCa

F. EqERV-Y4

```

5LTR TGTAGCCCGCCCTGACACAGGAAGGGTAAATAACTCTGGAAAAAGCTGCCAACAACTGTGACCTGGAGTGGAGAAGG
3LTR TGTAGCCCGCCCTGACACAGGAAGGGTAAATAACTCTGGAAAAAGCTGCCAACAACTGTGACCTGGAGTGGAGAAGG
5LTR CCGGTTAGCCAAATGACGGTGGG-----GGGGGGGGGGGGCAACTAAGCCAGCGCGGCCCCGGGGGGCCAGAT
3LTR CCGGTTAGCCAAATGACAGTAAAGACCCCCGGGGGGGGGGGGCAACTAAGCCAGCGCGGCCCCGGGGGGCCAGAT
5LTR GGAGAAACGATATCGATACAGGACAGGAGGACTGTGCTTAAAGAGCTTGCCTGCTCCGAGATTAATAATAGAGG
3LTR GGAGAAACGATATCGATACAGGACAGGAGGACTGTGCTTAAAGAGCTTGCCTGCTCCGAGATTAATAATAGAGG
5LTR CACAGCAATAACATGATAATATAAACAAGAGGCAAGGAGGGCTCCTTGAAAACTCAAGAATCTTGGCATTCGC
3LTR CACAGCAATAACATGATAATATAAACAAGAGGCAAGGAGGGCTCCTTGAAAACTCAAGAATCTTGGCATTCGC
5LTR TAAATTACTTCAATCCAGAACCTGTGTATGTTTAAACAGATGTAAGCTACGTATATATTGAAACCTGGTTAATAAAT
3LTR TAAATTACTTCAATCCAGAACCTGTGTATGTTTAAACAGATGTAAGCTACGTATATATTGAAATCTGGTTAATAAAT
5LTR CAAAGTGCCATCAGCCCTCTCCGATCTATCTGATGTGACTTTGGATTGGACCCGACA
3LTR CAGAGTGCCATCAGCCCTCTCCGATCTATCTGATGTGACTTTGGATTGGACCCGACA
    
```

Reverse complement of PBS TCcctTgcGGTCCcCa
 Horse chr1 tRNA109-Pyr (GTA) UCaccCbaGGGTCCGcGg

Fig. 2. Alignment of LTR sequences and PBS with tRNA sequences of six novel EqERV families. EqERV LTR lengths differ depending on the family.

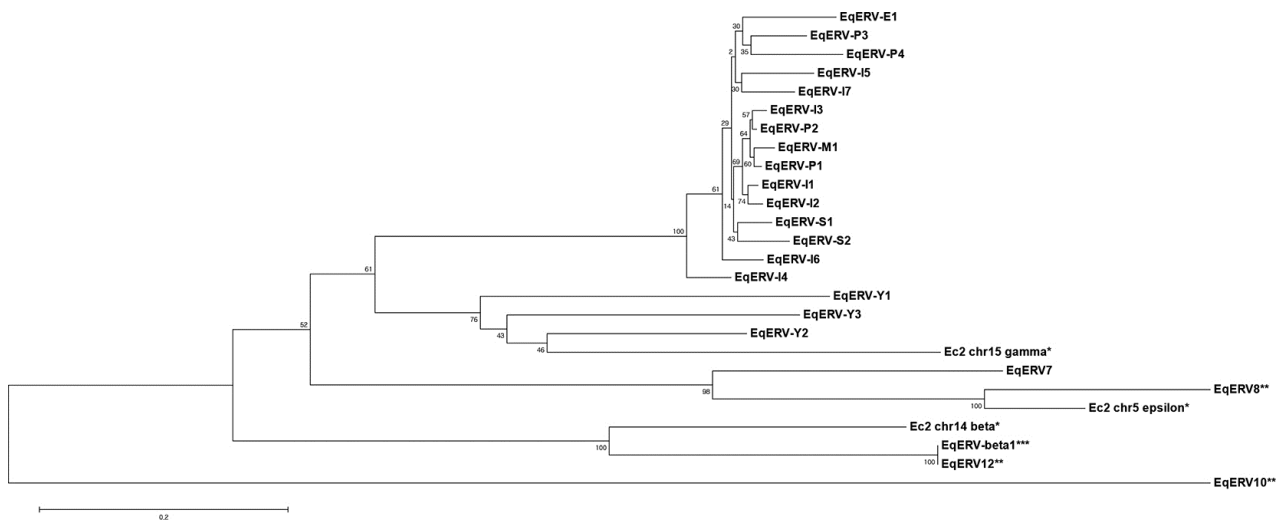


Fig. 3. Phylogenetic trees based on the *pol*/RT regions of identified EqERVs. The tree was generated by the neighbor-joining method. Bootstrap evaluation of the branching patterns was executed with 100 replicates. Branch lengths are matched to genetic distance. EqERVs identified in previous studies were indicated as asterisks. *(Brown et al., 2012), *(Garcia-Etxebarria and Jugo, 2012), *(van der Kuyl, 2011).

Phylogenetic analysis of identified EqERVs

We compared the relationship of the identified EqERV families with the EqERVs derived from previous studies using phylogenetic analysis (Fig. 3). Firstly, EqERV-beta1 was firstly identified in chromosome 5 (van der Kuyl, 2011). Then, Brown et al. identified the EqERV *pol* genes of beta-, gamma-, and epsilon-retroviruses (Brown et al., 2012), as well as Garcia-Etxebarria and Jugo identified the three classes of EqERVs in horse genome (Garcia-Etxebarria and Jugo, 2012). The phylogenetic tree was generated by the neighbor-joining method, with the EqERV RT regions, because the same families of ERV have similar RT regions (Tristem, 2000). EqERV-S3 was excluded from the phylogenetic analysis because of its short RT region. HERV-Y is recently identified (Gim et al., 2015), and EqERV-Y is also identified in the horse genome (Garcia-Etxebarria and Jugo, 2012). In the previous study, EqERV7, EqERV8 were class I, EqERV12 was class II, and EqERV10 was class III, respectively (Garcia-Etxebarria and Jugo, 2012). Class I EqERVs were included with epsilon-retrovirus, and our EqERV-Y families were included with gamma-retrovirus. Beta-retrovirus, EqERV-beta1, and class II EqERV12 were included in same group, and class III EqERV10 was out-grouped (Fig. 3). In the previous studies, the phylogenetic relationships of EqERVs were compared with other ERVs (Brown et al., 2012; Garcia-Etxebarria and Jugo, 2012; van der Kuyl, 2011). In this study, we revealed the phylogenetic relationship of six EqERV families, which can provide insights into the family-specific functions.

Expression analysis of EqERV *pol* in horse tissues of two subspecies

Many researchers have proposed ERV-derived transcripts or proteins as potential pathological factors. ERV *gag*, *pol*, and *env* gene transcripts and proteins are expressed in various

cancer tissues at higher levels than in normal tissues (Ahn and Kim, 2009; Reis et al., 2013). The expression of ERV genes has many effects in host organisms. In a previous study, 978 EqERV-derived *gag*, *pol*, and *env* transcripts were detected in horse tissues based on a whole transcriptome analysis (Brown et al., 2012). According to a phylogenetic analysis of EqERV transcripts, EqERVs are closely related to bovine ERVs. Therefore, it is necessary to elucidate their biological functions and expression patterns. In general, viral genes have potentially detrimental effects, and integrated viral genes in the host genome may also have deleterious effects on the host. The functions of HERV transcripts, HERV proteins, and even HERV-derived particles are extensively in various human tissues, most notably in cancer tissues. The best example of an ERV function is that of HERV-W *env*, which has an important role in placenta formation. In order to characterize ERV transcripts, it is important to identify tissue-specific expression patterns. Likewise, HERV-W *env* also presents expression patterns specific for brain diseases or brain cells, which suggests that the expression of ERVs can have immunopathogenic and neuroinflammatory roles in the host (Frank et al., 2005; Perron et al., 2012). Accordingly, the elucidation of ERV expression patterns in the brain is important to determine the exact role of the virus in this area. Therefore, we assessed the expression of EqERV in two brain specimens obtained from a Thoroughbred horse and a Jeju horse. Primers were designed to amplify the *pol*/RT regions in the whole transcriptome using the Tablet graphical viewer (Milne et al., 2013). As shown in Fig. 4, we examined the expression patterns of *pol* by RT-qPCR. EqERV-E1 *pol* was expressed at higher levels in the Jeju horse than in the Thoroughbred, particularly in the cerebellum (Fig. 4A). Each *pol* gene of EqERV-I1, M2, P1, S1, and Y4 showed higher expression in the cerebellum of the Jeju horse than the

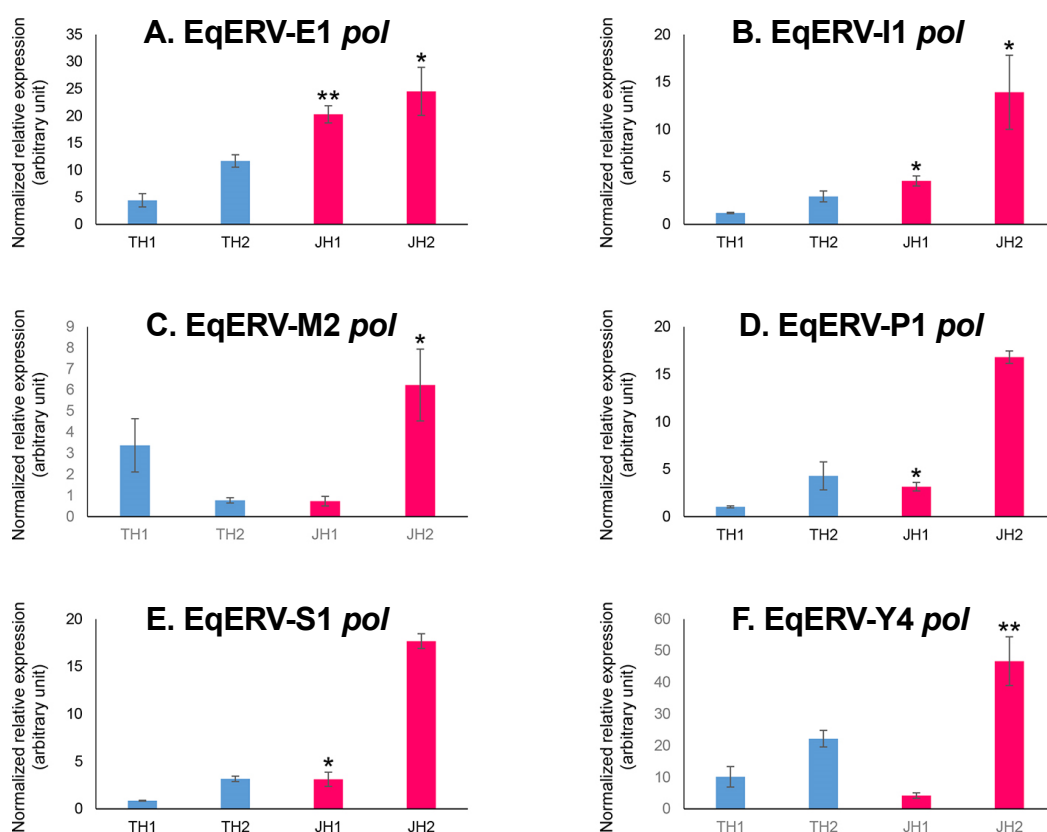


Fig. 4. Quantitative expression of six novel EqERV families in Thoroughbred and Jeju horse tissues by RT-qPCR. Relative expression levels of EqERV genes were normalized to the expression levels of horse *B2M* genes. TH1 indicates the Thoroughbred horse cerebrum and TH2 indicates the cerebellum. JH1 indicates the Jeju horse cerebrum and JH2 indicates the cerebellum. Compared between Thoroughbred and Jeju horse in each tissues, significantly different expression patterns were indicated as asterisks (* $p < 0.05$, ** $p < 0.005$).

Thoroughbred (Fig. 4). In humans, HERV-P *pol* expression was detected in brain and testis tissues (Yi et al., 2007), while HERV-S *pol* is specifically expressed patterns in the brain and thymus (Yi et al., 2004). The viral ERV genes can induce the proinflammatory cytokines in brain and immune system, and have crucial roles in the development and regulates central nervous system (CNS) (Kremer et al., 2013; Mortelmans et al., 2016). Then, pharmacological-induction of ERV genes could regulate the function of the brain or CNS. These expression data may provide the clues of Thoroughbred and Jeju horse specific pharmacological-induction of ERVs, and improve the breeding system of each horses. Thus, various expression patterns of EqERV *pol* in brain tissues could be of great use for further functional studies in horse research.

ACKNOWLEDGMENTS

This work was supported by a grant from the ministry of Education, Republic of Korea (NRF-2015R1D1A1A01056650).

REFERENCES

Ahn, K., Bae, J.-H., Nam, K.-H., Lee, C.-E., Park, K.-D., Lee, H.-K., Cho, B.-W., and Kim, H.-S. (2011a). Identification of reference genes for

normalization of gene expression in thoroughbred and Jeju native horse (Jeju pony) tissues. *Genes Genom.* *33*, 245-250.

Ahn, K., Han, K., and Kim, H.S. (2011b). Quantitative analysis of the HERV *pol* gene in human tissues. *Genes Genom.* *33*, 439-443.

Ahn, K., and Kim, H.S. (2009). Structural and quantitative expression analyses of HERV gene family in human tissues. *Mol. Cells* *28*, 99-103.

Anderssen, S., Sjøttem, E., Svineng, G., and 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.

Blikstad, V., Benachenhou, F., Sperber, G.O., and Blomberg, J. (2008). Evolution of human endogenous retroviral sequences: a conceptual account. *Cell. Mol. Life Sci.* *65*, 3348-3365.

Bower, M.A., McGivney, B.A., Campana, M.G., Gu, J., Andersson, L.S., Barrett, E., Davis, C.R., Mikko, S., Stock, F., Voronkova, V., et al. (2012). The genetic origin and history of speed in the Thoroughbred racehorse. *Nat. Commun.* *3*, 643.

Brown, K., Moreton, J., Malla, S., Aboobaker, A.A., Emes, R.D., and Tarlinton, R.E. (2012). Characterisation of retroviruses in the horse genome and their transcriptional activity via transcriptome sequencing. *Virology* *433*, 55-63.

Cho, G.J. (2007). Genetic relationship and characteristics using microsatellite. *J. Life Sci.* *17*, 699-705.

- Dunn, C.A., Romanish, M.T., Gutierrez, L.E., van de Lagemaat, L.N., and Mager, D.L. (2006). Transcription of two human genes from a bidirectional endogenous retrovirus promoter. *Gene* *366*, 335-342.
- Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A.J., Heidmann, T., and Schwartz, O. (2005). APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* *433*, 430-433.
- Frank, O., Giehl, M., Zheng, C., Hehlmann, R., Leib-Mösch, C., and Seifarth, W. (2005). Human endogenous retrovirus expression profiles in samples from brains of patients with schizophrenia and bipolar disorders. *J. Virol.* *79*, 10890-10901.
- Garcia-Etxebarria, K., and Jugo, B.M. (2012). Detection and characterization of endogenous retroviruses in the horse genome by in silico analysis. *Virology* *434*, 59-67.
- Gifford, R., and Tristem, M. (2003). The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* *26*, 291-315.
- Gim, J.-A., Han, K., and Kim, H.-S. (2015). Identification and expression analysis of human endogenous retrovirus Y (HERV-Y) in various human tissues. *Arch. Virol.* *160*, 2161-2168.
- Gu, J., Orr, N., Park, S.D., Katz, L.M., Sulimova, G., MacHugh, D.E., and Hill, E.W. (2009). A genome scan for positive selection in thoroughbred horses. *PLoS One* *4*, e5767.
- Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT; in *Nucleic Acids Symp. Ser.* Vol. 41, pp. 95-98.
- Hill, E.W., Gu, J., Eivers, S.S., Fonseca, R.G., McGivney, B.A., Govindarajan, P., Orr, N., Katz, L.M., and MacHugh, D.E. (2010). A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses. *PLoS One* *5*, e8645.
- Hughes, J.F., and Coffin, J.M. (2001). Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. *Nat. Genet.* *29*, 487-489.
- Jern, P., and Coffin, J.M. (2008). Effects of retroviruses on host genome function. *Annu. Rev. Genet.* *42*, 709-732.
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* *36*, W5-W9.
- Johnson, W.E., and Coffin, J.M. (1999). Constructing primate phylogenies from ancient retrovirus sequences. *Proc. Natl. Acad. Sci. USA* *96*, 10254-10260.
- Kang, Y.J., Jo, J.O., Ock, M.S., Chang, H.K., Baek, K.W., Lee, J.R., Choi, Y.H., Kim, W.J., Leem, S.H., Kim, H.S., et al. (2014). Human ERV3-1 env protein expression in various human tissues and tumours. *J. Clin. Pathol.* *67*, 86-90.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* *16*, 111-120.
- Kremer, D., Schichel, T., Förster, M., Tzekova, N., Bernard, C., Valk, P., Horszen, J., Hartung, H.P., Perron, H., and Küry, P. (2013). Human endogenous retrovirus type W envelope protein inhibits oligodendroglial precursor cell differentiation. *Ann. Neurol.* *74*, 721-732.
- Lau, A.N., Peng, L., Goto, H., Chemnick, L., Ryder, O.A., and Makova, K.D. (2009). Horse domestication and conservation genetics of Przewalski's horse inferred from sex chromosomal and autosomal sequences. *Mol. Biol. Evol.* *26*, 199-208.
- Lebedev, Y.B., Belonovitch, O.S., Zybrowa, N.V., Khil, P.P., Kurdyukov, S.G., Vinogradova, T.V., Hunsmann, G., and Sverdlov, E.D. (2000). Differences in HERV-K LTR insertions in orthologous loci of humans and great apes. *Gene* *247*, 265-277.
- Lee, A., Nolan, A., Watson, J., and Tristem, M. (2013). Identification of an ancient endogenous retrovirus, predating the divergence of the placental mammals. *Phil. Trans. R. Soc. B* *368*, 20120503.
- Macfadden, B.J. (2005). Evolution. Fossil horses—evidence for evolution. *Science* *307*, 1728-1730.
- Malassine, A., Handschuh, K., Tsatsaris, V., Gerbaud, P., Cheynet, V., Oriol, G., Mallet, F., and Evain-Brion, D. (2005). Expression of HERV-W Env glycoprotein (syncytin) in the extravillous trophoblast of first trimester human placenta. *Placenta* *26*, 556-562.
- Malfavon-Borja, R., and Feschotte, C. (2015). Fighting fire with fire: endogenous retrovirus envelopes as restriction factors. *J. Virol.* *89*, 4047-4050.
- Martins, H., and Villesen, P. (2011). Improved integration time estimation of endogenous retroviruses with phylogenetic data. *PLoS One* *6*, e14745.
- Milne, I., Stephen, G., Bayer, M., Cock, P.J., Pritchard, L., Cardle, L., Shaw, P.D., and Marshall, D. (2013). Using Tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* *14*, 193-202.
- Mortelmans, K., Wang-Johanning, F., and Johanning, G.L. (2016). The role of human endogenous retroviruses in brain development and function. *Apmis* *124*, 105-115.
- Oakenfull, E.A., Lim, H.N., and Ryder, O.A. (2000). A survey of equid mitochondrial DNA: Implications for the evolution, genetic diversity and conservation of Equus. *Conserv. Genet.* *1*, 341-355.
- Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., et al. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* *499*, 74-78.
- Park, K.D., Park, J., Ko, J., Kim, B.C., Kim, H.S., Ahn, K., Do, K.T., Choi, H., Kim, H.M., Song, S., et al. (2012). Whole transcriptome analyses of six thoroughbred horses before and after exercise using RNA-Seq. *BMC Genomics* *13*, 473.
- Perron, H., Mekaoui, L., Bernard, C., Veas, F., Stefas, I., and Leboyer, M. (2008). Endogenous retrovirus type W GAG and envelope protein antigenemia in serum of schizophrenic patients. *Biol. Psychiatry* *64*, 1019-1023.
- Perron, H., Germi, R., Bernard, C., Garcia-Montojo, M., Deluen, C., Farinelli, L., Faucard, R., Veas, F., Stefas, I., and Fabriek, B.O. (2012). Human endogenous retrovirus type W envelope expression in blood and brain cells provides new insights into multiple sclerosis disease. *Mult. Scler. J.* *18*, 1721-1736.
- Reis, B.S., Jungbluth, A.A., Frosina, D., Holz, M., Ritter, E., Nakayama, E., Ishida, T., Obata, Y., Carver, B., Scher, H., et al. (2013). Prostate cancer progression correlates with increased humoral immune response to a human endogenous retrovirus GAG protein. *Clin. Cancer Res.* *19*, 6112-6125.
- Ruprecht, K., Mayer, J., Sauter, M., Roemer, K., and Mueller-Lantzsch, N. (2008). Endogenous retroviruses and cancer. *Cell. Mol. Life Sci.* *65*, 3366-3382.
- Shin, J.A., Yang, Y.H., Kim, H.S., Yun, Y.M., and Lee, K.K. (2002). Genetic polymorphism of the serum proteins of horses in Jeju. *J. Vet. Sci.* *3*, 255-263.
- Sperber, G., Lovgren, A., Eriksson, N.E., Benachenhou, F., and Blomberg, J. (2009). RetroTector online, a rational tool for analysis of retroviral elements in small and medium size vertebrate genomic sequences. *BMC Bioinformatics* *10 Suppl 6*, S4.
- 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.

van der Kuyl, A.C. (2011). Characterization of a full-length endogenous beta-retrovirus, EqERV-beta1, in the genome of the horse (*Equus caballus*). *Viruses* **3**, 620-628.

Yi, J.M., and Kim, H.S. (2006). Molecular evolution of the HERV-E family in primates. *Arch. Virol.* **151**, 1107-1116.

Yi, J.M., Kim, T.H., Huh, J.W., Park, K.S., Jang, S.B., Kim, H.M., and Kim, H.S. (2004). Human endogenous retroviral elements belonging

to the HERV-S family from human tissues, cancer cells, and primates: expression, structure, phylogeny and evolution. *Gene* **342**, 283-292.

Yi, J.M., Schuebel, K., and Kim, H.S. (2007). Molecular genetic analyses of human endogenous retroviral elements belonging to the HERV-P family in primates, human tissues, and cancer cells. *Genomics* **89**, 1-9.