DOI: https://doi.org/10.5564/mjas.v23i01.1012

THE IDENTIFICATION AND CLASSIFICATION OF ENDOGENOUS RETROVIRUSES IN THE HORSE GENOME

E.Batmagnai[1*], Erik Bongcam-Rudloff[2], Matthew Peter Kent[3], Göran Andersson[2]

[1]Institute of Veterinary and Medicine, Mongolian University of Life Sciences,
Ulaanbaatar, Mongolia
[2]Swedish Agricultural University, Sweden
[3]Norwegian University of Life Sciences, Norway

[*]Corresponding author: magnaishuudan@gmail.com

ABSTRACT

Endogenous retroviruses (ervs) are sequences that derived from ancient retroviral infections of germ cells and integrated in humans, mammals and other vertebrates millions years ago. These ervs are inherited according to Mendelian expectations as all other genes in the genome. Coding sequences are flanked by two ltrs (long terminal repeat sequences). Most ervs are defective however some ervs still have open reading frames in their genome. These ervs settle close to functional genes or within the genes and can influence or control functions of the host genes using their ltrs. Most integration has deleterious effects. However some integration could be example of positive co-adaptation as syncitin. The first equine endogenous beta retrovirus which is ecerv-beta1 has been found in 2011 by Antoinette C.van der Kuyl[1]. The first known beta retrovirus and few pol gene similar to foamy retrovirus were only known endogenous retroviruses fixed in the domestic horse (equuscaballus) genome. Our aim of the study was to identify other endogenous retrovirus sequences in an equine genome and classify them into groups. Based on the high number of sines (equine repetitive element) in the horse genome we hypothesized that certain ervs will be located sufficiently close to sines that they will be amplified using an unbiased sine-pcr approach with degenerate primers. The nearest sine element was located 5.5 kbp upstream at the 5'of the ecerv-beta1. Pan-pol pcr was also used to find novel ervs based on 640 bp long region of pol gene which is the most conserved region of ervs. 27 complete and novel ervs that are 13 beta, 13 gamma, 1 spuma and 249 candidate endogenous retroviruses have been revealed using ltr_struc tool and double checked by retrotector online tool and ncbi-blast tool. It was proven that ecerv-beta1, which has 2 ltrs with 1% divergence between ltrs has a polymorphism among 13 different breeds.

KEYWORDS: Endogenous retrovirus, LTR, SINES, horse, classification

INTRODUCTION

The infections of first exogenous retroviruses into the germ cell could have appeared at any time over an extended evolutionary time-scale between 2 to 70 million years ago [5]. Many extant species have been analyzed for their endogenous retroviral content, and even the extinct woolly mammoth has been shown to contain endogenous proviral fragments in its genome [6]. Surprisingly, information on endogenous retroviruses fixed in the domestic horse (Equus caballus) genome is scarce [1]. The first horse ERV, the full length beta retrovirus genome was retrieved from a horse chromosome 5 contig by Antoinette C. van der Kuyl and published in 2011. We pursued to find out all other EcERVs. SINEs was used as templates for identifying novel ERVs because it is likely that several ERVs are located in the vicinity of the SINEs. The idea of using SINE-PCR approach was rooted principally in the high density occurrence of SINE elements in the mammalian genome. Horses (Equus caballus) have abundant SINE elements as well as other mammals. Recent study has estimated that $5*10^4$ copies of Equine Repetitive Element-1 are in horse genome [2]. The location of *pol* gene allows us to determine novel ERV from horse genome. The Pan-PCR approach has universal, degenerate primers which are called 5'MOP-2 and 3'MOP-2 that can amplify approximately 640 bp product of *pol* gene which is the most conserved region. This approach has been successfully used in other species genome like human, swine, and avian genome etc. Recent integrations are likely to be polymorphic between different breeds. EcERV-beta1 has 2 LTRs with 1% divergence, which is relatively recent integration. It occured approximately 2.5 million years ago (mya). LTR divergence is the crucial factor for polymorphism. The classification of ERV using Retrotector tool is based on Pol nucleotide sequence similarity and Pol protein conservation (Jern et al., 2005). Pol protein is the most well conserved retrovirus protein therefore useful for classification. Integrated proviruses may activate cellular gene expression either in somatic cells or following germ-line infection. Most commonly, this has been detected as increased cellular growth associated with oncogenes. Syncytin is the best known example of co-adaptation between viruses and the host. Syncytin mediates placental cytotrophoblast fusion *in vivo*, and thus plays an important role in human placental morphogenesis [3].

MATERIALS AND METHODS

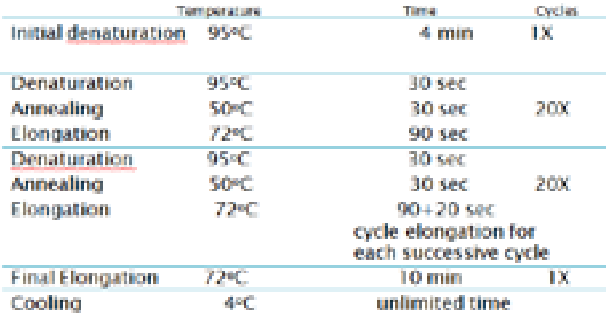
Our study consists of 2 sections, bioinformatics and experimental sections.

Experimental section SINE-PCR: The first step was to find the nearest SINE element in the flanking region of known equine beta endogenous retrovirus (EcERV-Beta1).To find the nearest SINE elements in the vicinity of EcERV-beta1 degenerate primers were designed using the multiple alignments of previously known SINEs. We have cloned the PCR products using TOPO TA cloning kit and sequenced. Degenerate primers were designed from the conserved region of all known horse SINE elements using the multiple alignments of all equine repetitive elements which were archived in GIRI database. Forward primer: CCRGBGTTTCGYTGGTTCV, where R stands for A or G; B stands for G, T or C; Y stands for C or T; V stands for G, A or C;

Reverse primer: CTAGAGAGGGGCAAAAACTTCTC

Table 1.

Cycle parameters of Touch down PCR



	Temperature	Time	Cycles
Initial denaturation	95°C	4 min	1X
Denaturation	95°C	30 sec	
Annealing	50°C	30 sec	20X
Elongation	72°C	90 sec	
Denaturation	95°C	30 sec	
Annealing	50°C	30 sec	20X
Elongation	72°C	90+20 sec cycle elongation for each successive cycle	
Final Elongation	72°C	10 min	1X
Cooling	4°C	unlimited time	

Pan-pol^{10}PCR was used to amplify 640 bp of conserved pol region.The degenerate oligonucleotides were used:

5‘ MOP-2 (5‘-CCWTGGAATACTCCYRTWTT-3‘)
3‘ MOP-2 (5‘-GTCKGAACCAATTWATATYYCC-3‘), where R stands for A or G, Y stands for C or T, K stands for G or T, and W stands for A or T.

PCR products were cloned and sequenced. 2 samples Thoroughbred horse were used in SINE-PCR and Pan-PCRs. Polymorphism of EcERV-beta1 were tested on 26 samples from 13 different breeds such as Shetland pony and Icelandic horse.

Cloning of PCR products: SINE and Pan pol PCR products were both cloned and sequenced. The products of the expected size were cut by scalpel from the 0.8% agarose gel after gel electroporation and gel purified using SNAP Mini prep Kit. After that the products were ligated with pCR2.1-TOPO vector. TOPO TA Cloning Kit for sequencing was used according to the protocol.

Sequencing of Cloned plasmid inserts

M13forward and M13 reverse primers were used to generate a nucleotide sequence of the DNA insert cloned into pCR-TOPO2.1.

M13 Forward (−20) 5′-GTAAAACGACGGCCAG-3′

M13 Reverse 5´-CAGGAAACAGCTATGAC-3´.

Plasmids DNAs were directly sequenced by BigDye® Direct Cycle Sequencing Kitaccording to the protocol. Since plasmid has its own M13 tails PCR amplification was not required before sequencing

Polymorphism of EcERV Beta1 region between 13different breeds

3360 bp segment has been amplified, which includes whole pol region. It was possible to analyze polymorphism between breeds. For that purpose 13 different horse breeds were used.

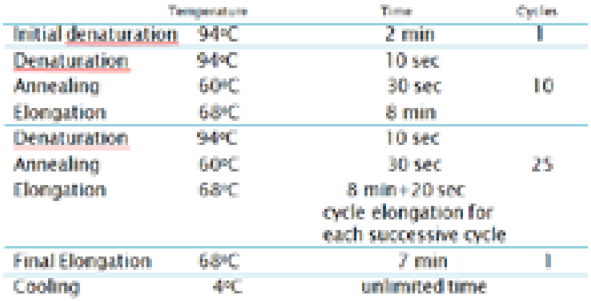
Left primer:
GTCTCAAGCCTCCTTCGAGC

Right primer:
TCCACAAAGGAGAGGAAGCG

Long range PCR amplification was used for pol region.

Table 2.

Cycle parameters of Long range PCR



	Temperature	Time	Cycles
Initial denaturation	94°C	2 min	1
Denaturation	94°C	10 sec	
Annealing	60°C	30 sec	10
Elongation	68°C	8 min	
Denaturation	94°C	10 sec	
Annealing	60°C	30 sec	25
Elongation	68°C	8 min+20 sec cycle elongation for each successive cycle	
Final Elongation	68°C	7 min	1
Cooling	4°C	unlimited time	

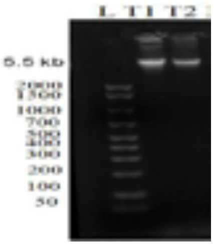
Bioinformatics approach

LTR_STRUC was the main tool on bioinformatics part of the study and the latest available version of the horse genome, EquCab2 sequence was used in the experiment. Repetitions were sorted out and excluded from further analysis. Retrotector online tool was used for scrutinizing the results of LTR_STRUC tool.

NCBI-BLAST

NCBI-BLAST search for endogenous retrovirus was used to double-check the candidates of ERVs.[7,8]BLAT (The BLAST-like Alignment Tool) searches through the Horse (*Equuscaballus*) Genome Browser Gateway of the Genome Bioinformatics Group of UC Santa Cruz.[9]

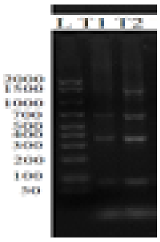
RESULTS



Experimental results:

SINE-PCR results

Figure 1.SINE-PCR result. L is ladder between 50-2000 bp. T represents thoroughbred horses.. Touch down PCR was used and the nearest SINE element has been found in 5.5 kb far from the ERV. It was quite large fragment between SINE element and EcERV Beta1. Through sequencing the fragment, we can find the location of the SINE element.



Pan-pol PCR results

Figure 2.Pan-PCR results. L is ladder between 50-2000 bp. T represents thoroughbred horses.

Touch down PCR was used and annealing temperature was at 45°C.

Sequencing results: Pan-pol sequence:
Pan-pol PCR products were cloned and sequenced. 7 endogenous retroviruses have been found from unplaced genomic scaffold and 3 ERVs from chromosome number 5 by sequence of Pan PCR products. These 640 bp products are overlapped with pol regions of EcERVs of chromosome 5.

SINE-sequence:
SINE-PCR products were sequenced but we sequenced 5.5 kb region between known beta retrovirus (EcERV beta1) and its SINE element instead of 230 bp SINE element. In order to find other ERVs near to SINE element of EcERV beta1 this experiment needs to be repeated.

Polymorphism of EcERV Beta1

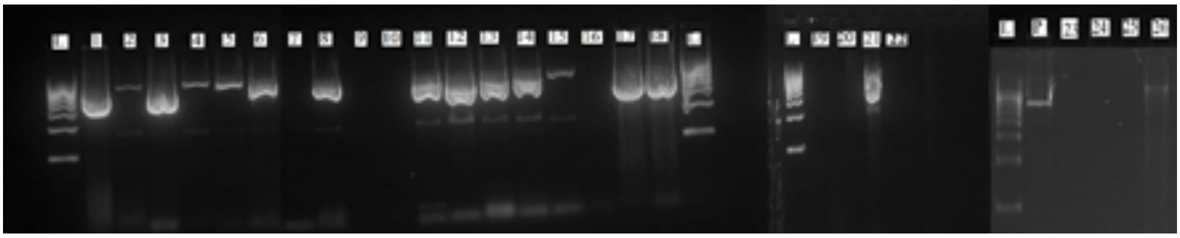


Figure 3 .Polymorphism of EcERV Beta1
L is 500 -5000 bp ladder, 1,3-Thoroughbred horse1, 2,11,-North Swedish warmblood , 4,8-,10Standardbred, 5,26-Shetland pony, 6-Morgan horse, 7,16-Gotland pony (Sweden), 9,19-Icelandic horse, 12,21-North Swedish draft, 13-Swedish Ardenne, 14, 15-Connemara pony, 17,18-Faeroe pony, 20-202, 22, 23-Knabstruber, 24, 25- Welsh pony, P is a sample of thoroughbred horse that was used as a positive control in the second PCR

From figure.3 we can see that the expected region was not amplified at the same size. And some breeds do not contain this EcERV Beta1 retrovirus or they have accumulated mutations on that region. Therefore we can say that there is a polymorphism between breeds. Some of them have 2 products and they could be possible candidates. There is definitely a polymorphism in different breeds because although some samples do not have the expected band but they have primer dimer in the bottom which can prove the PCR was performed well. Sample number 26 has slightly larger fragment than the positive control. The insertion could be the cause of different size of the products. Thoroughbred horses (1 and 3) have same products. Standardbreds (4 and 8) have 2 products but number 10 was not amplified. The reason of sample 10 could be bad DNA quality or fragmented DNA. Morgan (6) horse has the band. Shetland pony has the band (5 and 26) Connemara pony (14 and 15) Swedish Ardenne (13) has the ERV, Faeroe pony (17 and 18) have the ERV. Swedish draft (12 and 21) has the ERV, North Swedish warmblood (2) and Swedish Warmblood (11) have the ERV, Knabstruber does not have (22 and 23) Icelandic horses do not have this product (9,19). Gotland pony does not have (7 and 16) 202 (20) does not have. Welsh pony does not have (24 and 25). This polymorphism could be due to the geographic of different breed's distribution.

Bioinformatics results
Main result of the bioinformatics approach was that 27 complete and novel ERVs were found.

LTR_STRUC results
A total of 276 unique EcERVs were identified and every calculation and analysis on bioinformatics part were based on these selected endogenous retrovirus sequences from LTR_STRUC. Thoroughbred mare's (Twilight) genome was used by LTR_STRUC. The average EcERV is 8.3 kb long and the amount of 276 EcERV is 2299577 bp or 0, 085 % (based on the 276 chains real lengths in the horse genome that consist of 2.68 GB).

Score
It was proven that the highest scored ones occur more common in the genome by NCBI-BLAST tool. For example chr520000_RT3_B7_L7_8 on chromosome 5 has 99% similarity versions in other chromosomes as 11, 29, 1, 15, 6 etc.

27 complete endogenous retroviruses were discovered while all candidates were examined by Retrotector online tool. These ERVs were abundantly present on all other chromosomes except chromosome 29 and 31. 27 complete endogenous retroviruses that are 13 beta, 13 gamma, 1 spuma have been revealed from this study using LTR_STRUC tool. Previously known the EcERV Beta1 has also been found within these ERVs and we used it as a positive control in the *in silico* analysis. Retrotector score was quite high (1007.1) in average among complete 27 ERVs.

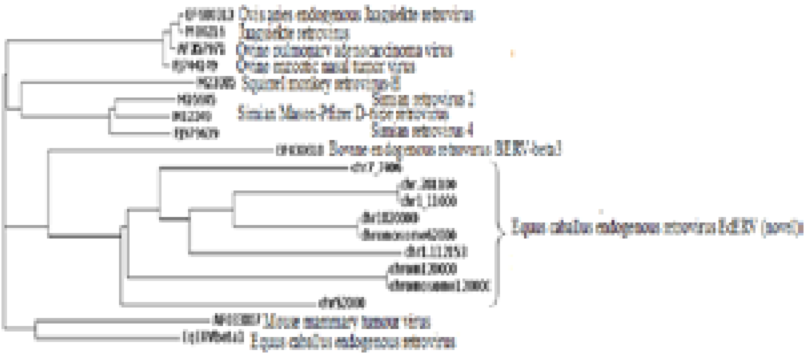

Ovis aries endogenous Jaagsiekte retrovirus
Jaagsiekte retrovirus
Ovine pulmonary adenocarcinoma virus
Ovine enzootic nasal tumor virus
Squirrel monkey retrovirus-H
Simian retrovirus 2
Simian Mason-Pfizer D-type retrovirus
Simian retrovirus 4
Bovine endogenous retrovirus BERV-beta3
Equus caballus endogenous retrovirus EcERV (novel)
Mouse mammary tumour virus
Equus caballus endogenous retrovirus

Figure 4. Phylogenetic tree of translated pol sequences of novel EcERVs, EcERV Beta1, beta- and delta retroviruses. Genbank accession numbers are indicated.

Novel EcERVs were branched closer to each other and resemble most closely with Bovine ERV-beta3 in the phylogenetic tree (Figure 4). And EcERV-beta1 was branched with murine retrovirus, Mouse mammary tumour virus. Beta retroviruses from sheep and primates are only distantly related to EcERVs and EcERV-beta1.

DISCUSSION

The integration polymorphism between breeds of EcERV Beta1 is an interesting result. The primers were designed on the bases of pol conserved gene of EcERV Beta1. Y chromosome has not been included in this analysis because the horse genome sequence that we mined was made from a mare. More ERVs could have been found if we included Y chromosome because it is assumed as “graveyard” of ERVs.[4] Chromosome 29 and 31 are lack of equine ERVs from this result. It remains to be confirmed that these chromosomes are essentially lacking e*quine ERV*s or whether it reflects annotation-bias of these chromosomes. Previous published research on equine endogenous retrovirus was limited. We have found several unique integrations within the functional genes that may be cause of cancer or imprint of co-adaptation between host and retrovirus. In The Retroviridae book volume 2 page 258. “A number of studies have probed various equine tissues for the presence of endogenous retroviruses. (Rice et al, 1978, 1989. Rasty et al 1990, O‘Rourke et al, 1991). None of the studies (Southern blots, PCR) have detected endogenous retrovirus sequences in tissues of equine origin, although more sensitive techniques such as nested PCR have not yet been used to search for equine retroelements. However, in 2011, 20 years later Van der Kuyl found the first Equine endogenous beta retrovirus using by Blast search. In the present study we have found 27 novel and complete ERVs with other candidates. The low amount of *EcERVs*in horse i.e. 276 unique chains scored more than 0.3 with LTR_STRUC in this study.

CONCLUSIONS

It was shown that SINE-PCR approach is available to find novel ERVs from horse genome by finding the nearest SINE element of the known beta retrovirus which belongs to ERE1 family. Pan-PCR has worked well on horse genome as well as other species genome. Seven EcERVs were found from unassembled region of horse genome and three ERVs were found from chromosome 5 as variants of EcERV beta1. 276 EcERV elements were discovered by LTR_STRUC tool based on the criterium that they should pass the lower limit of 0.3 score. 27 novel and complete EcERVs have been found which is about 10% of all candidates and the first beta ERV has also been found within them therefore we assumed it as a positive control for the *in silico* analysis. Nine equine ERVs located on the unassembled part of horse genome have been found by NCBI-BLAST tool. We have found 4 ERVs from chromosome 5 using LTR_STRUC tool and 9 ERVs from unassembled region using NCBI-BLAST tool and 3 out of 4 ERVs from chromosome 5 and 7 out of 9 ERVs from NCBI-BLAST results were same as what we have found from Pan-PCR result. The equine genome has been effective in protection from extensive retroviruses integration. We have studied the polymorphism between breeds on EcERV-Beta1. Integrations with 1% divergence between LTRs have polymorphism between breeds. The complexity of horse endogenous retroviruses was identified and classified to the retroviral genera. EcERVs were classified and characterized using a bioinformatics approach and experimental approaches. The highest scored chains were characterized according to their functional capacity. Using Retrotector© online tool, the following classes were identified in the horse genome: 53, 9% or 122 candidates were determined as gammaretroviruses, 34, 9% or 79 were determined as betaretroviruses. 3% or 7 were determined as deltaretroviruses, 7.9% or 18 were determined as spumaretroviruses from determined endogenous retroviruses.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude for kindly support and guidance to my supervisor Professor Göran Andersson, at the Department Animal Breeding and Genetics SLU, and Associate Professor Erik Bongcam-Rudloff at the Linnaeus Centre for Bioinformatics (LCB) Uppsala University, who encouraged and challenged me through my academic program. Great thanks for kindly support, guidance, to my supervisor Professor Matthew Peter Kent at the Department Animal Breeding and Genetics, NMBU and Professor B.Battsetseg at the Institute of Veterinary and Medicine of Mongolia, Professor Sofia Mikko at the Department Animal Breeding and Genetics SLU.

REFERENCES:

1. Van der Kuyl, A.C. “Characterization of a Full-Length Endogenous Beta-Retrovirus, EqERV-Beta1, in the Genome of the Horse (Equuscaballus).” Viruses 2011, 3, 620-628.
2. Richard Cordaux and Mark Batzer (October 2009). "The impact of retrotransposons on human genome evolution". Nature Reviews Genetics 10 (10): 691–703.
3. Mi S, Lee X, Li X, Veldman GM, et al. ”Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis.” Nature, Feb 2000, 403(6771):785–9, doi: 10.1038/35001608.
4. Kjellman C, Sjögren HO, Widegren B. “The Y chromosome: a graveyard for endogenous retroviruses.” 1995 Aug 19;161(2):163-70.
5. Ekerljung, Marie“Molecular systematics: data mining of canine endogenous retroviruses, CFERV.”Dept. of Animal Breeding and Genetics, SLU. 2007,Vol. 295.
6. Greenwood, A.D.; Lee, F.; Capelli, C.; DeSalle, R.; Tikhonov, A.; Marx, P.A.; MacPhee, R.D. “Evolution of endogenous retrovirus-like elements of the woolly mammoth (Mammuthusprimigenius) and its relatives.” Mol. Biol. Evol. 2001, 18, 840–847.
7. Horse Genome Resources, NCBI. Available online: http://www.ncbi.nlm.nih.gov/projects/genome/guide/horse/ (accessed on 4 April, 2012).
8. NCBI Basic Local Alignment Search Tool BLAST. Available online: http://blast.ncbi.nlm.nih.gov/ (accessed on 4 April, 2012).
9. Horse (Equuscaballus) Genome Browser Gateway of the Genome Bioinformatics Group of UC Santa Cruz. Available online: http://genome.ucsc.edu/cgi-bin/hgGateway?db=equCab2 (accessed on 4 January 2011).
10. Thomas Ericsson,Beth Oldmixon,Jonas Blomberg, Margaret Rosa, Clive Patience, and GöranAndersson “Identification of Novel Porcine Endogenous Betaretrovirus Sequences in Miniature Swine”J. Virol.March 2001 vol. 75 no. 6 2765-2770