Avian sarcoma and leukosis virus *gag* gene in the *Anser anser domesticus* genome

F. Zhu[1], H. Jie[3], L. Lian[2], L.J. Qu[2], Z.C. Hou[2], J.X. Zheng[2], S.Y. Chen[1], N. Yang[2] and Y.P. Liu[1]

[1]Institute of Animal Genetics and Breeding,
College of Animal Science and Technology,
Sichuan Agricultural University, Ya'an, Sichuan, China
[2]Department of Animal Genetics and Breeding,
National Engineering Laboratory for Animal Breeding,
Ministry of Agriculture, Laboratory of Animal Genetics and Breeding,
College of Animal Science and Technology,
China Agricultural University, Beijing, China
[3]Laboratory of Medicinal Animal, Chongqing Institute of Medicinal Plant Cultivation, Chongqing, China

Corresponding author: Y.P. Liu
E-mail: liuyp578@yahoo.com

Genet. Mol. Res. 14 (4): 14379-14386 (2015)
Received June 10, 2015
Accepted September 28, 2015
Published November 18, 2015
DOI http://dx.doi.org/10.4238/2015.November.18.1

ABSTRACT. Endogenous retroviruses are regarded as ideal genetic markers for evolutionary analyses. Birds were some of the initial vertebrates found to contain endogenous retroviruses. However, few studies have investigated the presence and distribution of endogenous retroviruses in goose. In this study, we detected the avian sarcoma and leukosis virus *gag* gene in the genomic DNA of 8 Chinese native breeds using polymerase chain reaction method. The results indicated that a 1.2-kb avian sarcoma and leukosis virus *gag* sequence was integrated into all 8 goose breeds. The mean genetic pairwise distance was 0.918% among the investigated geese. To the best of our knowledge, this is the

first report demonstrating the presence of the endogenous retroviruses in the domestic goose genome. The genetic structure should be further examined in the domestic goose.

Key words: ASLV *gag*; Endogenous retroviruses; Goose

INTRODUCTION

Endogenous retroviruses (ERVs) are regarded as ideal genetic markers for evolutionary analyses; they are vertical transmitted intragenomic elements derived from integrated exogenous retroviruses. ERVs consist of 2 long terminal repeats and the canonical retroviral *gag*, *pol*, and *env* genes. Because retroviral integration can be considered random, each ERV is present at a unique chromosomal location (Stoye, 2012). The overwhelming majority of ERVs have been identified using hybridization (Hayward and Hanafusa, 1975; Martin et al., 1981; Chambers et al., 1986; Smith et al., 1999) and polymerase chain reaction (PCR) strategies by amplifying conserved viral sequences (Smith et al., 1999; Xiao et al., 2008) in various animal species. Herniou et al. (1998) identified ERVs in reptilian, amphibian, and piscine vertebrates using PCR to amply viral sequence-coding protease and reverse transcriptase. Subsequently, Dimcheff et al. (2000) amplified avian sarcoma and leukosis virus (ASLV) *gag* sequence in 26 galliform birds using PCR. In addition to genome data releases, more effective *in silicon* approaches were used to search the ERVs in the host genome, including *Gallus gallus* (Barr et al., 2005), *Bos taurus* (Garcia-Etxebarria and Jugo, 2010), and *Meleagris gallopavo* and *Poephila guttata* (Bolisetty et al., 2012). ERVs have been detected in major animals, including human (Martin et al., 1997), cattle (Zhang et al., 2008), chicken (Tereba et al., 1979), sheep (Klymiuk et al., 2003), and pig (Klymiuk et al., 2002), among others (Syomin and Ilyin, 2006). Although ERVs are important in evolution and disease studies, few studies have examined ERVs in the domestic goose. ERVs have been confirmed to be present in both the chicken (Barr et al., 2005) and zebra finch genomes (Bolisetty et al., 2012). The divergence time between chicken and zebra finch is approximately 100 million years, which is near the base of avian radiation (Hackett et al., 2008). Genes shared by both chicken and zebra finch genomes are quite likely to represent genes present in most, if not all, bird genomes (Hackett et al., 2008). Thus, we hypothesized that ERVs exist in the goose genome. Therefore, we detected ERVs in goose genome using the conserved *gag* sequence, which is sufficiently conserved to yield information from diverse host species and sufficiently variable to provide sequence differences among conspecific species (Dimcheff et al., 2000). The ASLV *gag* sequence was identified in 8 domestic goose breeds. Our results increase the knowledge of ERVs in goose.

MATERIAL AND METHODS

Sample collection

Genomic DNA was isolated by a simple lyses method from 8 Chinese local goose breeds, including Wanxi White goose, Yan goose, Xupu goose, Zhedong White goose, Huoyan goose, Zhi goose, Wugang Tong goose, and Gushi White goose. The geographic distribution of these geese is listed in Table 1.

Table 1. Geographic information for the 8 goose breeds.

Common name	Clone abbreviation	Geographic distribution
Wanxi White goose	Wan	Anhui, China
Yan goose	Yan	Anhui, China
Xupu goose	Xu	Hunan, China
Zhedong White goose	Zhe	Zhejiang, China
Huoyan goose	Huo	Shandong, China
Zhi goose	Zhi	Heilongjiang, China
Wugang Tong goose	Wu	Hunan, China
Gushi White goose	Gus	Henan, China

PCR amplification and sequencing

One pair of gene-special primers targeting the *gag* gene of ASLV described by Dimcheff et al. (2000) was used in this study. Primer locations within *gag* are illustrated in Figure 1 and had the following sequences: F, 5'-GCCGTCATAAAGGTGATTTCGTC-3'; R, 5'-AAGGACTCAGATGGTCCCTG-3'. Primers were synthesized by Shanghai Sangon Ltd. Company (Shanghai, China). The estimated length of the target fragment of the *gag* gene was 1.2 kb. PCR conditions were as follows: 95°C for 3 min followed by 35 cycles at 95°C for 30 s, 55°C for 30 s, 72°C for 1 min, and then 72°C for 5 min. Reaction conditions were as follows: 0.5-1 μg template DNA, 20 pmol of each primer, 200 μM deoxynucleotide triphosphates, PCR buffer (10 mM Tris, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), and 2 U *Taq* polymerase in a 20-μL reaction volume. Amplification fragments were electrophoresed on 2% agarose gels, and products of approximately 1200 bp were excised, purified, and cloned into the vector PMD19-T (Takara, Beijing, China). Cloned inserts were sequenced in both directions using an ABI Prism BigDye Terminator cycle sequencing ready reaction kit (Applied Biosystems, Foster City, CA, USA) on ABI 3700 automated sequencers (Applied Biosystems). The sequence of the Zhedong White goose was submitted to NCBI, with the accession No. KC466570.

Data analysis

Target sequences were analyzed by BLASTn (expected [E]-value $<10^{-10}$) against the GenBank nr database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). ASLV *gag* sequences were aligned based on the alignment of inferred amino acid sequences using ClusterX (Thompson et al., 1997). Pairwise genetic distances were calculated using the Tamura 3-parameters. Next, a phylogenetic tree was generated using a neighbor-joining model with 500 bootstraps implanted in the MEGA (v5.1) software (Tamura et al., 2007). The GenBank accession numbers of 3 published chicken ALV endogenous (ALVE), which were used for analysis, are KC610515, KC610516, and KC610517.

RESULTS

We designed primers based on conserved *gag* sequences from different species. Theoretically, the primer can be used to clone the 1st exon of the gap gene. We successfully obtained a 1.2-kb goose gap DNA sequence from all 8 investigated Chinese local goose breeds. To verify the accuracy of the goose gap sequence, we conduct BLAST (BLASTn,

E-value = 10^{-10}) against the GenBank nucleotide database. More than 100 unique ASLV-related sequences showed matches with obtained goose *gag* sequences. We then compared goose ASLV *gag* to chicken ALVE *gag* sequences. Three chicken ALVE sequences from BLAST results were used for alignment with 8 Chinese goose ASLV *gag* sequences. The total alignment length of goose sequences was 1203 bp (Figure 1), which was located in the first exon of the chicken gag gene, indicating that the obtained goose sequences were the *gag* fragment. The predicted amino acid changes of 8 goose ASLV *gag* sequences are shown in Figure 2. The number of variable sites within goose was 35, and 60% (21) of these sites were missense mutations.

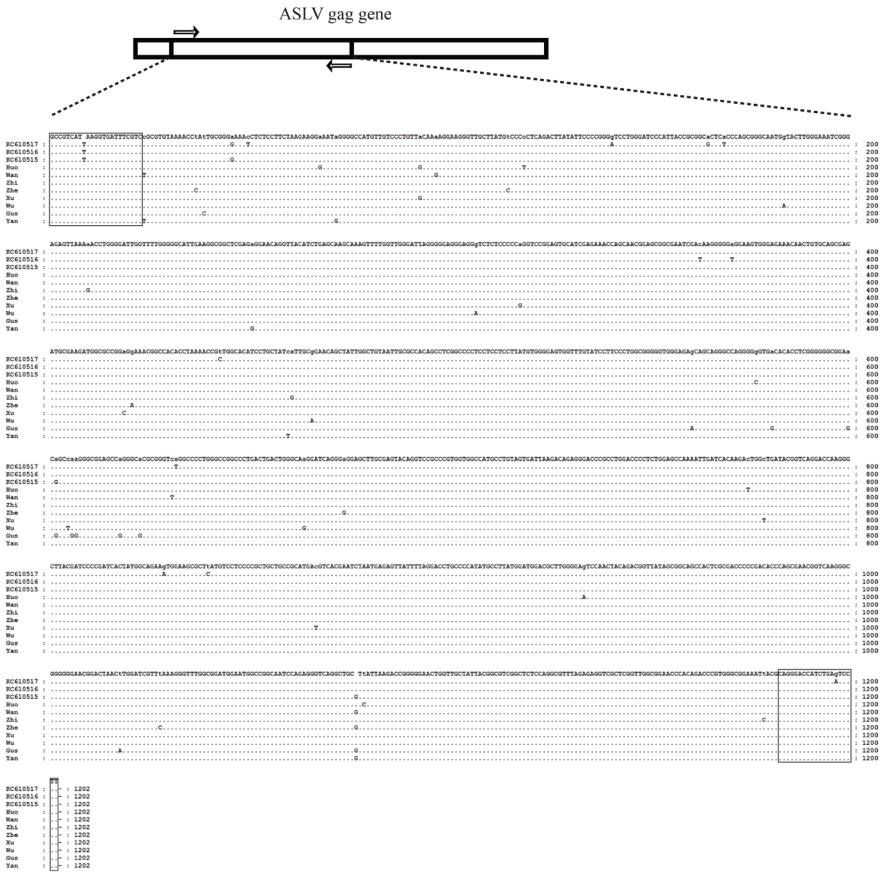

Figure 1. The entire sequence is a subset of the ASLV *gag* gene open-reading frame. The nucleotide sequence of forward primer was 5'-GCCGTCATAAAGGTGATTTCGTC-3', and the reverse one is 5'-AAGGACTCAGATGGTCCCTG-3'. Nucleotides in the box indicate the forward and reverse primers. Single nucleotide polymorphisms between obtained *gag* sequences from each goose breed and published *gag* ALVE sequences are illustrated in lower case. Dot is used to indicate the identities and dash to indicate deletions. Zhe, Gus, Huo, Xu, Yan, Wan, Wu, and Zhi represent Zhedong White goose, Gushi White goose, Huoyan goose, Xupu goose, Yan goose, Wanxi White goose, Wugangtong goose, and Zhi goose, respectively. Three chicken ASLV sequences (GenBank Nos.: KC610515, KC610516, and KC610517) were used for alignment. Consensus sequence is illustrated above the alignment.

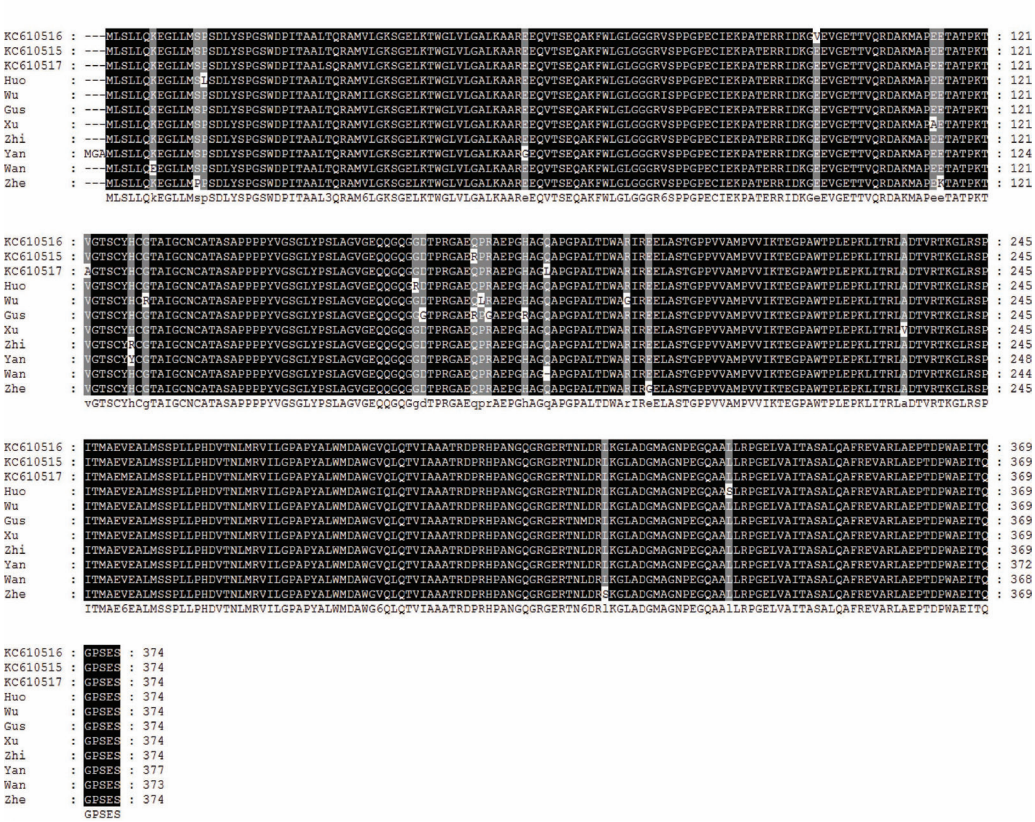

Figure 2. Alignment between goose ASLV *gag* and chicken ASLV *gag* based on the deduced amino acid sequences. Zhe, Gus, Huo, Xu, Yan, Wan, Wu, and Zhi represent Zhedong White goose, Gushi White goose, Huoyan goose, Xupu goose, Yan goose, Wanxi White goose, Wugangtong goose, and Zhi goose, respectively. Three chicken ASLV sequences (GenBank Nos.: KC610515, KC610516, and KC610517) were used for alignment. Black regions represent high consensus and gray represent partial consensus.

Particularly, a stop codon mutation, at the 180th amino acid location of the *gag* sequence was identified in WanXi White goose ASLV *gag*. We calculated the genetic distances among 3 chicken and 8 goose *gag* sequences. Within the chicken group, only 14 nucleotide positions were variable. The pairwise genetic distances among all sequences, which included 3 published ASLV *gag* sequences, was 0.925%. The mean pairwise distance of 0.918% was found for all clones isolated from goose and the mean pairwise distance of 0.782% was found from chicken, which indicated that *gag* fragment in goose was more variable than in chicken. The genetic pairwise distance between goose and chicken gag fragment was 0.952%. The neighbor-joining tree was illustrated in Figure 3. Although there was a lack of good phylogenetic resolution, midpoint-neighbor-joining tree showed a clean separation between chicken ALVE *gag* and goose ASLV *gag* (Figure 3).

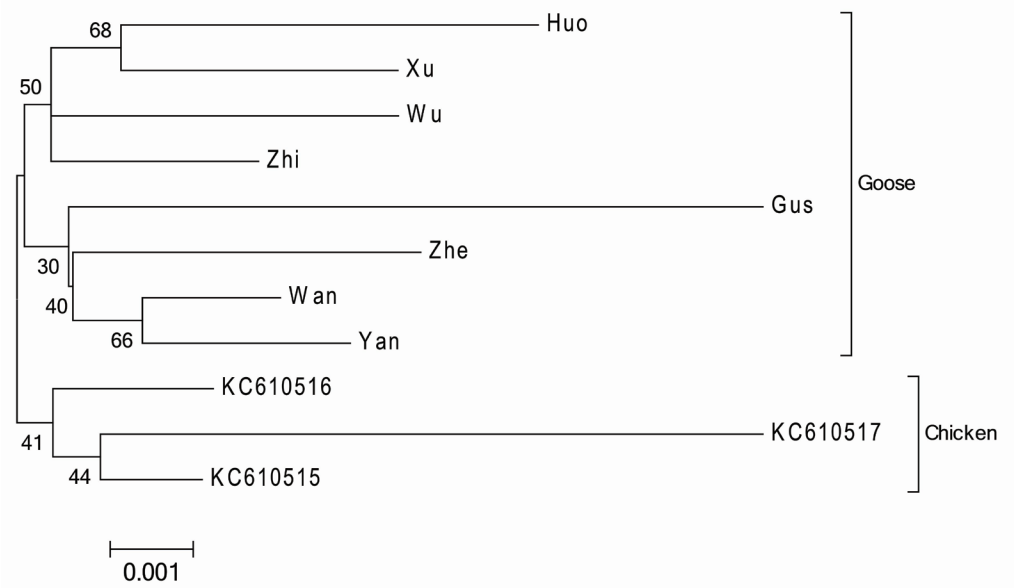

Figure 3. Phylogenetic midpoint neighbor-joining trees. Phylogenetic tree was constructed based on the deduced ASLV gag amino acid sequences using Tamura 3-parameter corrected distances with 500 bootstraps. Zhe, Gus, Huo, Xu, Yan, Wan, Wu, and Zhi represent Zhedong White goose, Gushi White goose, Huoyan goose, Xupu goose, Yan goose, Wanxi White goose, Wugangtong goose, and Zhi goose, respectively. Three chicken ASLV sequences (GenBank Nos.: KC610515, KC610516, and KC610517) were used for alignment. Numbers on each branch reflect percentage bootstrap value.

DISCUSSION

Abundant ERVs and its fragments have been reported in the genome of chicken, including ALV/RSV (Barr et al., 2005, Smith and Benkel, 2009) and MLV (Borysenko et al., 2008). In addition, studies have searched for ERVs in the domestic goose genome. In early studies, reverse transcriptase, a necessary enzyme for replicating ERVs, was detected in goose embryonic cells without the synthesis of viral DNA, indicating that ERVs exist in goose genome (Bauer and Temin, 1979). Grunder et al. (1993) did not identify the *pol* or *env* genes, which are regarded as essential markers of ALV/RSV in the goose genome using nucleic acid hybridization and indicating that endogenous ALV is not present in the goose genome. However, Gifford et al. (2005) identified class II-related ERVs, for which ALV/RSV are clustered in the genome of white-fronted goose and North American black duck, the close relatives of domestic goose, by PCR. This indicates that endogenous sarcoma-leukosis viruses are present in the domestic goose genome. Thus, the *gag* gene is present in the goose genome. Our results, together with those of Gifford et al. (2005) and Bauer and Temin (1979), indicate that endogenous sarcoma-leukosis virus fragments were integrated into the genome of goose at some evolutionary phases.

The results of Grunder et al. (1993) contrasted with our results and those of Gifford et al. (2005), which may be because of different methods used. Hybridization is a relatively low-sensitivity method for identifying ERVs, while PCR is sufficiently sensitive to detect only

several copies of target gene. The results showed that *gag* sequences have higher mutation rates in goose than in chicken. The higher mutation rate in goose than in chicken may be because chicken has undergone more intensive artificial selection than goose. However, the impact of selection on variations in the ASLV must be further investigated. The higher mutation rate in the goose ASLV *gag* than the chicken ASLV *gag* may reflect the general disease-resistant differences between chicken and goose. Further studies are necessary to provide insight into the relationship between variations in ASLV and general disease-resistance in birds.

In conclusion, our study supported that ERVs may be common in most of avian species based on our results in goose and other birds.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31172181), the Program for New Century Excellent Talent in University (#NCET-10-0889), and Sichuan Province (#2011JTD0032; #11TD007).

REFERENCES

Barr SD, Leipzig J, Shinn P, Ecker JR, et al. (2005). Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. *J. Virol.* 79: 12035-12044.

Bauer G and Temin HM (1979). RNA-directed DNA polymerase from particles released by normal goose cells. *J. Virol.* 29: 1006-1013.

Bolisetty M, Blomberg J, Benachenhou F, Sperber G, et al. (2012). Unexpected diversity and expression of avian endogenous retroviruses. *MBio* 3: e00344-12.

Borysenko L, Stepanets V and Rynditch AV (2008). Molecular characterization of full-length MLV-related endogenous retrovirus ChiRV1 from the chicken, *Gallus gallus*. *Virology* 376: 199-204.

Chambers J, Cywinski AA, Chen PJ and Taylor JM (1986). Characterization of Rous sarcoma virus-related sequences in the Japanese quail. *J. Virol.* 59: 354-362.

Dimcheff DE, Drovetski SV, Krishnan M and Mindell DP (2000). Cospeciation and horizontal transmission of avian sarcoma and leukosis virus gag genes in galliform birds. *J. Virol.* 74: 3984-3995.

Garcia-Etxebarria K and Jugo BM (2010). Genome-wide detection and characterization of endogenous retroviruses in *Bos taurus*. *J. Virol.* 84: 10852-10862.

Gifford R, Kabat P, Martin J, Lynch C, et al. (2005). Evolution and distribution of class II-related endogenous retroviruses. *J. Virol.* 79: 6478-6486.

Grunder AA, Benkel B, Sabour P and Gavora JS (1993). Research note: avian leukosis retroviral genes are not detected in geese. *Poultry Sci.* 72: 363-367.

Hackett SJ, Kimball RT, Reddy S, Bowie RC, et al. (2008). A phylogenomic study of birds reveals their evolutionary history. *Science* 320: 1763-1768.

Hayward W and Hanafusa H (1975). Recombination between endogenous and exogenous RNA tumor virus genes as analyzed by nucleic acid hybridization. *J. Virol.* 15: 1367-1377.

Herniou E, Martin J, Miller K, Cook J, et al. (1998). Retroviral diversity and distribution in vertebrates. *J. Virol.* 72: 5955-5966.

Klymiuk N, Müller M, Brem G and Aigner B (2002) Characterization of porcine endogenous retrovirus gamma pro-pol nucleotide sequences. *J. Virol.* 76: 11738-11743.

Klymiuk N, Müller M, Brem G and Aigner B (2003). Characterization of endogenous retroviruses in sheep. *J. Virol.* 77: 11268-11273.

Martin J, Herniou E, Cook J, O'Neill RW, et al. (1997). Human endogenous retrovirus type I-related viruses have an apparently widespread distribution within vertebrates. *J. Virol.* 71: 437-443.

Martin MA, Bryan T, Rasheed S and Khan AS (1981). Identification and cloning of endogenous retroviral sequences present in human DNA. *Proc. Natl. Acad. Sci. U. S. A.* 78: 4892-4896.

Smith A and Benkel BF (2009). Novel avian leukosis virus-related endogenous proviruses from layer chickens: characterization and development of locus-specific assays. *Poult Sci*. 88: 1580-1585.

Smith LM, Toye AA, Howes K, Bumstead N, et al. (1999). Novel endogenous retroviral sequences in the chicken genome closely related to HPRS-103 (subgroup J) avian leukosis virus. *J. Gen. Virol.* 80: 261-268.

Stoye JP (2012). Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat. Rev. Microbiol.* 10: 395-406.

Syomin BV and Ilyin YV (2006). Detection of the structural protein, Gag, of the endogenous insect retrovirus MDG4 (gypsy) in cultured cells. Doklady. *Dokl. Biochem. Biophys.* 408: 120-122.

Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.

Tereba A, Lai MM and Murti KG (1979). Chromosome 1 contains the endogenous RAV-0 retrovirus sequences in chicken cells. *Proc. Natl. Acad. Sci. U. S. A.* 76: 6486-6490.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, et al. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876-4882.

Xiao R, Park K, Lee H, Kim J, et al. (2008). Identification and classification of endogenous retroviruses in cattle. *J. Virol.* 82: 582-587.

Zhang Y, Maksakova IA, Gagnier L, van de Lagemaat LN, et al. (2008). Genome-wide assessments reveal extremely high levels of polymorphism of two active families of mouse endogenous retroviral elements. *PLoS Genet.* 4: e1000007.