

Article

The Nature and Chromosomal Landscape of Endogenous Retroviruses (ERVs) Integrated in the Sheep Nuclear Genome

Sarbast Ihsan Mustafa ^{1,2} , Trude Schwarzacher ¹ and John S. Heslop-Harrison ^{1,*} 

¹ Department of Genetics and Genome Biology, University of Leicester, Leicester LE1 7RH, UK; sarbast.ihsan@uod.ac (S.I.M.); ts32@leicester.ac.uk (T.S.)

² Department of Animal Production, College of Agricultural Engineering Sciences, University of Duhok, Duhok 42001, Kurdistan Region, Iraq

* Correspondence: phh4@le.ac.uk

Abstract: Endogenous retroviruses (ERVs) represent genomic components of retroviral origin that are found integrated in the genomes of various species of vertebrates. These genomic elements have been widely characterized in model organisms and humans. However, composition and abundances of ERVs have not been categorized fully in all domestic animals. The advent of next generation sequencing technologies, development of bioinformatics tools, availability of genomic databases, and molecular cytogenetic techniques have revolutionized the exploration of the genome structure. Here, we investigated the nature, abundance, organization and assembly of ERVs and complete genomes of Jaagsiekte sheep retrovirus (JSRV) from high-throughput sequencing (HTS) data from two Iraqi domestic sheep breeds. We used graph-based read clustering (RepeatExplorer), frequency analysis of short motifs (k-mers), alignment to reference genome assemblies and fluorescent in situ hybridization (FISH). Three classes of ERVs were identified with the total genomic proportions of 0.55% from all analyzed whole genome sequencing raw reads, while FISH to ovine metaphase chromosomes exhibited abundant centromeric to dispersed distribution of these ERVs. Furthermore, the complete genomes of JSRV of two Iraqi sheep breeds were assembled and phylogenetically clustered with the known enJSRV proviruses in sheep worldwide. Characterization of partial and complete sequences of mammalian ERVs is valuable in providing insights into the genome landscape, to help with future genome assemblies, and to identify potential sources of disease when ERVs become active.

Keywords: chromosomes; eukaryotic genomes; endogenous retroviruses; fluorescent in situ hybridization; repetitive DNA; jaagsiekte sheep retrovirus (JSRV); high-throughput sequencing



Citation: Mustafa, S.I.; Schwarzacher, T.; Heslop-Harrison, J.S. The Nature and Chromosomal Landscape of Endogenous Retroviruses (ERVs) Integrated in the Sheep Nuclear Genome. *DNA* **2022**, *2*, 86–103. <https://doi.org/10.3390/dna2010007>

Academic Editor: Darren Griffin

Received: 11 January 2022

Accepted: 7 March 2022

Published: 16 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Mammalian nuclear genomes contain integrated fragments and entire DNA sequences of repetitive genetic elements that have homologous sequences to retroviral genomes known as endogenous retroviruses (ERVs) [1–5]. ERVs are genetic fossils descended from the infection and incorporation of ancient retroviruses into the chromosomal DNA of mammalian hosts [6,7]. ERVs were revealed in the late 1960s by observing that virological markers could be transmitted following patterns of Mendelian laws [8]. Within the repeat element database Repbase, ERV is one of the five superfamilies of LTR retrotransposons which is further subdivided into main five classes; ERV1, ERV2, ERV3, ERV4, and endogenous lentivirus [9]. ERVs are largely dispersed in vertebrate genomes in which 8.29% of the entire human genome represents sequences homology to the ERVs fragments [2,10]. ERVs are characterized by weak conservation of active sites in the encoded genes and the structure with an internal region including four genes in the order 5'-gag-pro-pol-env-3'. flanked by two Long Terminal Repeats (5' and 3' LTRs), which act as expression enhancers and/or promoters of cellular genes [7,11,12]. In human, most of the ERVs are truncated and silent, although some types of the human

endogenous retrovirus (HERV) such as HERV-H, HERV-K, and HERV-W are transcriptionally active [13,14]. In small ruminants, some diseases such as ovine pulmonary adenocarcinoma (OPA) (naturally occurring adenocarcinoma of sheep) are associated with exogenous pathogenic retroviruses known as the Jaagsiekte sheep retrovirus (JSRV) that cause neoplasms of the respiratory tract and abundantly expressed in the genital tract of the domestic sheep, predominantly in the trophoblast of the placenta and in the glandular epithelia and the endometrial luminal of the uterus [15,16]. In sheep, the Jaagsiekte sheep retrovirus belongs to the beta related exogenous class II within endogenous groups and therefore is called beta enJSRV [5,17]. In female sheep, enJSRV has been proposed to play crucial biological roles in the reproductive physiology [18]. Genomes from many organisms have been sequenced and assembled, but their genome assembly approaches lack efficient methods to classify and analyze diversity and genomic distribution of fragmented, degenerated and variable retroviral sequences. Thus, in this study, we aimed to analyze unassembled raw reads from whole genome sequencing of sheep using graph-based read clustering, k-mers frequency analysis and sequence alignments to references, to identify and characterize DNA sequences related to endogenous retroviruses. We aimed to reveal and understand the role of amplification, chromosomal localization and evolution of ERV related DNA repeats and complete genome of enJSRV in the context of genomic and cytogenetic evidence and phylogenetic relationships.

2. Materials and Methods

2.1. Animal Materials

Blood samples were collected from five individuals of sheep breeds representing two main breeds Karadi and Hamdani from Iraqi Kurdistan region (sample origin in Table S1). Total genomic DNA was extracted from whole blood using the Wizard Genomic DNA Purification kit (Promega, Southampton, UK). Five samples of genomic DNA were sequenced commercially at the University of Florida Interdisciplinary Center for Biotechnology Research, Gainesville, FL, USA using Illumina NextSeq500 mid-throughput with paired-end 2×150 bp cycles, generating 43 to 60 million raw reads ($2\text{--}3 \times$ coverage of the sheep genome) with 5 to 6 Gb total sequence for each DNA sample.

2.2. Discovery of Endogenous Retrovirus (ERV)-Related Repetitive Sequences

2.2.1. Graph-Based Read Clustering (RepeatExplorer)

Similarity-based read clustering implemented in RepeatExplorer [19,20] was used to identify major repetitive sequences in the whole genome sequence reads. Parameters of RepeatExplorer for clustering included a minimum read overlap length of $\geq 55\%$ of the read length and over 82 bases with 90% of similarity as edges to save the potential error of clustering reads with partial similarity among two unrelated ERV groups. The graphical output clusters with a genome proportion of $<0.01\%$ were analyzed to identify retroelement domains and homology hits for the major repeat classes.

2.2.2. k-mer Frequency Tool (Jellyfish)

As an alternative, reference-free, approach to identify abundant ERV-related sequences, a k-mer analysis of highly abundant sequence motifs k-bases long (k-mers) was carried out: the frequency of all short sequence motifs k-mers in the raw reads were calculated using Jellyfish version 2 [21], with $k = 22, 32$, and 44 . The most abundant k-mers in each group were assembled to generate longer contigs representing overlapping k-mers. Several thousand contigs were produced from assembly of short motifs and the top 100 contigs were analyzed.

2.2.3. Database Searching

The consensus sequence of each cluster from RepeatExplorer outcome, and assembled contigs of abundant k-mers, were compared with the Repbase database of repetitive DNA sequences (<http://www.girinst.org/repbase/> (accessed on 17 November 2017)) and NCBI

database to investigate and identify similar query sequences to ERV classes of sheep or other animal species.

2.3. Design and Amplification of FISH Probes, Chromosome Preparation and In Situ Hybridization

Consensus sequences of ERV related contigs were used for designing PCR primer using Geneious software [22] (Table 1). PCR amplifications were set up in 25 μ L total volume reaction mixture containing water (Sigma-Aldrich, St. Louis, MO, USA) (18.4 μ L), 10 \times Buffer A (Kapa Biosystems, Wilmington, MA, USA) (2.5 μ L), 10 mM dNTP Mix (1 μ L), 10 μ M forward and reverse primers; Sigma-Aldrich (each 0.5 μ L) and 5 U/ μ L KAPA Taq DNA Polymerase (Kapa Biosystems) (0.1 μ L) with 80–120 ng of genomic DNA. The PCR cycling conditions consisted of 3 min initial denaturation at 95 $^{\circ}$ C, followed by 35 cycles of denaturation (95 $^{\circ}$ C, 0.5 min), annealing (T_m -5 $^{\circ}$ C, 0.5 min) and primer extension (72 $^{\circ}$ C, 8 min). The final cycle was the 1 min final extension at 72 $^{\circ}$ C followed by indefinite hold time between 4–16 $^{\circ}$ C. Amplified PCR products were gel electrophoresed (1% *w/v* agarose) in 1 \times TAE buffer and purified using an E.Z.N.A. Cycle Pure Kit (Omega, Norcross, GA, USA) and then labelled with either biotin-16-dUTP or digoxigenin-11-dUTP (Roche Diagnostics) using the BioPrime Array CGH random priming kit (Invitrogen, Waltham, MA, USA). Probe designations use CL for cluster and ERV for endogenous retroviruses followed by its number. Whole sheep blood was collected from freshly slaughtered commercial sheep (Joseph Morris Butchers Ltd., Leicestershire, UK) in sterile 50 mL tubes containing heparin. About 43.5 mL of RPMI medium 1640 (GibcoTM, Fisher Scientific, Loughborough, UK); 0.5 mL of antibiotic antimycotic solution (10,000 lg/mL streptomycin, 10,000 U/mL penicillin G, and 25 lg/mL amphotericin B, HyCloneTM, GE Healthcare Life Sciences, Amersham, UK) and 6 mL of foetal calf serum were used to make lymphocyte short-term medium. Then, 7 mL medium containing either 0.5 or 0.75 mL of blood, 10–30 mg/mL phytohemagglutinin (PHA; Sigma-Aldrich) was incubated in 5% CO₂ incubator at 37 $^{\circ}$ C for 3–5 days. Metaphases were arrested by adding 50–90 mL of demecolcine solutions (10 lg/mL; Sigma-Aldrich) and left for further 1.5–2 h at 37 $^{\circ}$ C. Metaphase chromosome preparations were then made using hypotonic treatment with 0.075 M KCl and fixation in absolute methanol: glacial acetic acid 3:1. Fluorescent in situ hybridization (FISH) followed Schwarzacher and Heslop-Harrison (2000). The hybridization mixture contained formamide 50% (*v/v*), dextran sulphate 20% (*w/v*), saline sodium citrate (0.3 M NaCl, 0.03 M sodium citrate) 2 \times SSC (Saline-Sodium Citrate), FISH probe 50–100 ng, sheared salmon sperm DNA 20 μ g (Sigma-Aldrich), SDS (sodium dodecyl sulphate) 0.3% (*w/v*) and EDTA 0.12 mM. After overnight hybridization at 37 $^{\circ}$ C, low stringency (20% formamide and 0.1 \times SSC) washes were used enabling probe-target hybrids with more than 70% homology to remain stably hybridized. Biotin-labelled probes were detected with 2.0 μ g/mL streptavidin conjugated to Alexa594 (Molecular Probes), and digoxigenin probes were detected with 4 μ g/mL anti-digoxigenin conjugated to FITC (fluorescein isothiocyanate, Roche). Slides were simultaneously counterstained and mounted by applying antifade mixture [6 μ L DAPI (4',6-diamidino-2-phenylindole diluted in McIlvaines buffer pH 7.0; 100 μ g/mL), 97 μ L Citifluor antifade mountant solution (Citifluor, Agar Scientific, Stansted, UK) and 97 μ L ddH₂O]. Preparations were analyzed on a Nikon Eclipse N80i fluorescent microscope equipped with a DS-QiMc monochromatic camera (Nikon, Tokyo, Japan) and appropriate filters. Images were false-colored (red for the probe and cyan for DAPI), overlaid and the contrast adjusted with NIS-Elements BR3.1 software (Nikon) and Photoshop (Adobe, Mountain View, CA, USA) using only cropping and functions affecting the whole image equally.

Table 1. Primer sequences, PCR products, and probe names used for amplification and in situ hybridizations.

Probe Names	Name of Primers [Sequence (5′–3′)]	Expected Product Size (bp)	Annealing Temp.
CL18C5_ERV1	F = ATCTTGGCTGAGCGATGCG R = GGGCTCTTGTCTAACACTCGG	246	62
CL20C5_ERV1	F = TGTGTTGCCATGACCACTCC R = TGCCAGCATTCTTGACTCC	574	62
CL23C4_ERV1	F = CAAGGAATTTGGAGTGGTGGG R = TCGGTGGTCCTGTTGTAGCC	195	62
CL25_ERV1	F = TGTCATCTGGTCACTGCTGC R = AGGGAGTTTGCAGGATGTGG	402	62
22mer_ERV1.A	F = CACTCTTTTGCCCAATCCGG R = CAGCTACTTTTCGAGCTGCC	545	60
32mer_ERV1.RE	F = GGTTTTAGATGGGACCGGGC R = TCTTCCTGCCATTGGAAGGC	564	62
32mer_ERV1.T3	F = TGCTTCTTTTCAACGCACCC R = CTTGATGGAGCCAGGTACCC	541	64
CL14C75_ERV2	F = GGTGATTACATCATCTTCTGGCC R = AGCTTGCCTAACAGGTTCCC	505	62
OuttopCL_ERV2	F = AAAGGTCACGAGGATGAGGC R = AGGACAAAGGTGCAGTGGG	555	60
CL37_ERV2	F = TGTCTTTTCCTCTCTCGGC R = CATGCTTATGTCTGGGCTGC	488	62
CL67_ERV3	F = ATTCAATCTCCTAATATTCCCACCC R = GTTAGTAGTCAAGCTTTTGTCTGGC	315	58
CL27C1_ERV1+ERV3	F = GCAGGTCGGTGTATCTTCCC R = GGGAAGTTGCAAGAGTGGGG	619	62
32mer_ERV1+ERV3	F = CTTGCAAGAGTGGGGAAAGC R = GCAGGTCGGTGTATCTTCCC	615	62
32mer_ERV1+CRC	F = TACAGAGCAAAGGGGATGGG R = TGGTTGTTTCTTTCCACCATTCC	468	60

2.4. Assembly of the Complete Genome of the Endogenous Jaagsiekte Sheep Retrovirus (enJSRV)

The paired raw reads of the five samples of genomic DNA of Karadi and Hamdani were assembled to the complete genome of the Inner Mongolian Strain of the Endogenous Betaretroviruses Jaagsiekte Sheep Retrovirus (enJSRV; GenBank accession DQ838493) [23] to generate consensus sequences for five samples named HamJ1, HamJ2, HamM, KarJ and KarM (Supplementary Materials, Figure S4). The complete genome of enJSRV from Iraqi sheep breeds are available under the GenBank accession numbers of the NCBI MF175067, MF175068, MF175069, MF175070 and MF175071.

2.5. Data Analysis and Phylogenetic Relationships

The five complete endogenous beta retroviruses (enJSRV) genomes of Iraqi sheep breeds were aligned with published enJSRV genomes (NCBI) from various sheep breeds worldwide. The phylogenetic status was identified by constructing Geneious Tree Builder within the Geneious Prime software. The parameters for the phylogenetic tree were set on Tamura-Nei as a genetic distance model, and Neighbor-Joining as a build method with 500 as a number of replicates. The complete genome of ovine enzootic nasal tumor virus (GU292314) was used as the out-group. Copy numbers, genomic proportions and coverage of ERV sequence (probe fragment) or the complete enJSRV genome were estimated by mapping the whole genome raw reads to the consensus (Table 2).

Table 2. Copy numbers of various Endogenous Retroviruses (ERVs) related fragments used for in situ hybridization.

Probes of Endogenous Retroviruses ERVs	PCR Product bp	HamJ1_Male Genome			KarJ_Female Genome		
		Assembled Reads	Copies of Probe	Genomic Proportion %	Assembled Reads	Copies of Probe	Genomic Proportion %
CL18C5_ERV1	246	5973	3642	0.0115	7525	4588	0.0124
CL20C5_ERV1	574	3124	816	0.0060	3676	961	0.0061
CL23C4_ERV1	195	2157	1659	0.0041	2000	1538	0.0033
CL25_ERV1	402	6900	2575	0.0133	4364	1628	0.0072
22mer_ERV1.A	545	3452	950	0.0066	3008	828	0.0050
32mer_ERV1.RE	564	3753	998	0.0072	3426	911	0.0057
32mer_ERV1.T3	541	1709	474	0.0033	2554	708	0.0042
CL14C75_ERV2	505	5294	1572	0.0102	6943	2062	0.0115
OuttopCL_ERV2	555	1299	351	0.0025	1177	318	0.0019
CL37_ERV2	488	2113	649	0.0041	2338	719	0.0039
CL67_ERV3	315	1000	476	0.0019	0	0	0
CL27C1_ERV1+ERV3	619	5907	1431	0.0113	5595	1356	0.0092
32mer_ERV1+ERV3	615	5500	1341	0.0106	5300	1293	0.0087
32mer_ERV1+CRC	468	5700	1827	0.0110	5760	1846	0.0095

3. Results

3.1. Identification of Endogenous Retroviruses Related Sequences

RepeatExplorer was used to identify and classify LTR retrotransposons by domain sequence homology and order provided by the program. Several clusters containing ERV sequences were distributed over the RepeatExplorer clusters, each with different genomic proportions (Table 3). By comparing the clustered sequences to Repbase databases, all classes of endogenous retroviruses, related to LTR retrotransposons, including ERV1, ERV2 and ERV3 (Bao et al., 2015) were identified. The second approach to identify ERV-related sequences was to use assembled motifs from the most abundant k-mers. Different classes of ERV repeats (ERV1, ERV2 and ERV3) were found by BLAST searches of the top 100 k-mer contigs against repeat databases. Copy numbers and genomic proportion of each probe representing different classes of ERVs related DNA repetitive elements used in this study were estimated following mapping the raw reads against consensus of ERV sequences (Table 2). The total genomic proportions of all classes of endogenous retroviruses about 0.55% was estimated based on the genomic abundance of ERV related clusters from the RepeatExplorer outcome.

Table 3. Graph-based clusters with similarity to ERV from RepeatExplorer [19,20] analysis and Repbase database comparisons.








Clusters	Total Length	Number of Reads	Genome Proportion [%]	Repbase Database Similarities	Graph Layout
CL14C75_ERV2	309,579	2057	0.229	ERV2	
CL18C5_ERV1	32,673	217	0.024	ERV1	
CL20C5_ERV1	28,460	189	0.021	ERV1	

Table 3. Cont.

Clusters	Total Length	Number of Reads	Genome Proportion [%]	Rebase Database Similarities	Graph Layout
CL23C4_ERV1	21,077	140	0.016	ERV1	
CL27C1_ERV3+ERV1	17,746	118	0.013	ERV3 and ERV1	
CL25_ERV1	18,501	123	0.0147	ERV1	
CL37_ERV2	9926	66	0.0079	ERV2	

3.2. Identification and Quantification of ERV Repeats in Ancestral and *Bos Taurus* ERV Sequences

Raw reads of sheep were also mapped to concatenated ERVs of 105 ancestral sequences (total length; 145 kbp) and 90 sequences of *Bos taurus* (total length; 173 kbp) from the Repbase database (Supplementary Materials, Figures S1 and S2). Some 0.02% of the sheep reads (10,518 read pairs) were assembled to the ancestral ERV sequences. BLAST analysis against the Repbase database indicated that these ancestral reads were in type of ERV3 sequences with high (70–88%) similarity. For the concatenated ERVs from *Bos taurus*, 0.45% of the sheep reads (238,182 reads) were mapped.

3.3. Abundance and Genomic Organization ERV-Related Repetitive Elements

Amplified PCR products representing selected ERVs were labelled and hybridized to male sheep metaphase chromosomes (2n = 54 with three pairs of submetacentric autosomes, 23 autosomal acrocentric pairs, and the X and Y sex chromosomes). Signals varied from dispersed over all or most chromosomes to localized at centromeres or interstitial positions (Figures 1–5, Table 4).

3.3.1. ERV1

ERV1 sequences represented in the probe of CL18C5_ERV1 produced strong signals on the centromeres of all 23 autosomal acrocentric chromosome pairs. One pair of submetacentric autosomes had strong centromeric signals, while the other two pairs had weak signals while both X and Y chromosomes had weak centromeric signals (Figure 1A). Probe CL20C5_ERV1 showed specific signals at the centromeres of acrocentric autosomes, weaker signals on submetacentric autosomes, and no signals were seen on the largest submetacentrics and sex chromosomes (Figure 1B). Signals of probe CL23C4_ERV1 were present on about half the centromeres of acrocentric autosomes, with a more dispersed signals on some but not all chromosomes or arms in the submetacentric chromosomes. There were weak signals on the Y chromosome and stronger signals on the X chromosome (Figure 1C). Probe CL25_ERV1 showed variable signals, and slightly dispersed over all chromosomes including the sex chromosomes, with signals close to the centromeres of a few acrocentric chromosomes (Figure 1D). ERV1 sequences from k-mer analysis (22mers and 32mers GT100) also showed centromeric to dispersed patterns on sheep chromosomes: probe 22mer_ERV1.A (Figure 2A,B) labelled centromeres of all acrocentric autosomes with

small dots over the sex chromosomes X and Y, and weak signals or very small dots at centromeric and subtelomeric regions of submetacentrics. Probe 32mer_ERV1.RE from 32mers GT100 was rather dispersed with a concentration at centromeres, some gaps on submetacentric chromosomes and signals at centromeres of X and Y chromosomes (Figure 2C). Probe 32mer_ERV1.T3 from 32mers GT100 showed centromeric and dispersed dots over all chromosomes, mostly rather uniform, including the Y chromosome, while hybridization was more centromeric on the X chromosome (Figure 2D).

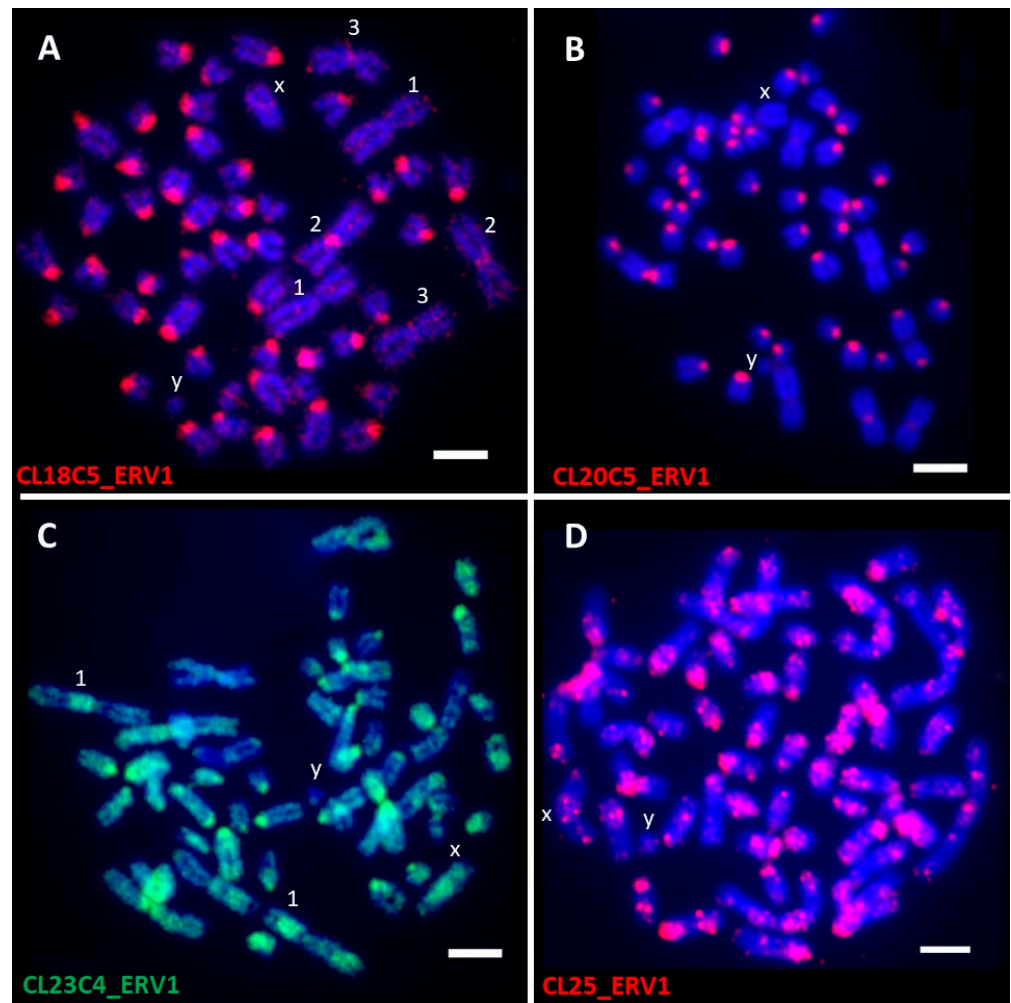


Figure 1. In situ hybridization of ERV1 probes (CL18C5_ERV1, CL20C5_ERV1, CL23C4_ERV1; CL25_ERV1) from RepeatExplorer clusters to the male sheep metaphase chromosomes ($2n = 54$) (fluorescing blue with DAPI). (A). Probe CL18C5_ERV1 (magenta) produced strong signals on the centromeres of acrocentric chromosomes. One pair of submetacentric chromosomes (No. 1) has strong centromeric signals, while the other two pairs (No. 2 and 3) have weaker signals. Both X and Y chromosomes have weak but noticeable signals. (B). Probe CL20C5_ERV1 (magenta) shows specific signals to the centromeres of acrocentric and weaker signals to the submetacentric chromosomes. No signals were seen on the largest submetacentric and the sex chromosomes. (C). Signals of probe CL23C4_ERV1 (green) is present on a few centromeres of about half of the acrocentrics; additionally, intercalary and variably dispersed signal is visible on some but not all chromosomes or arms of the submetacentric and the X chromosomes. There was a weak signal on the Y chromosome. (D). Probe CL25_ERV1 showed variable dotted signals dispersed over all chromosomes including the sex chromosomes. Signals were close to the centromeric domains of few acrocentrics. Numbers 1–3 refers to the three pairs of metacentric chromosomes, X and Y to the sex chromosomes. Scale bar equals 5 μm .

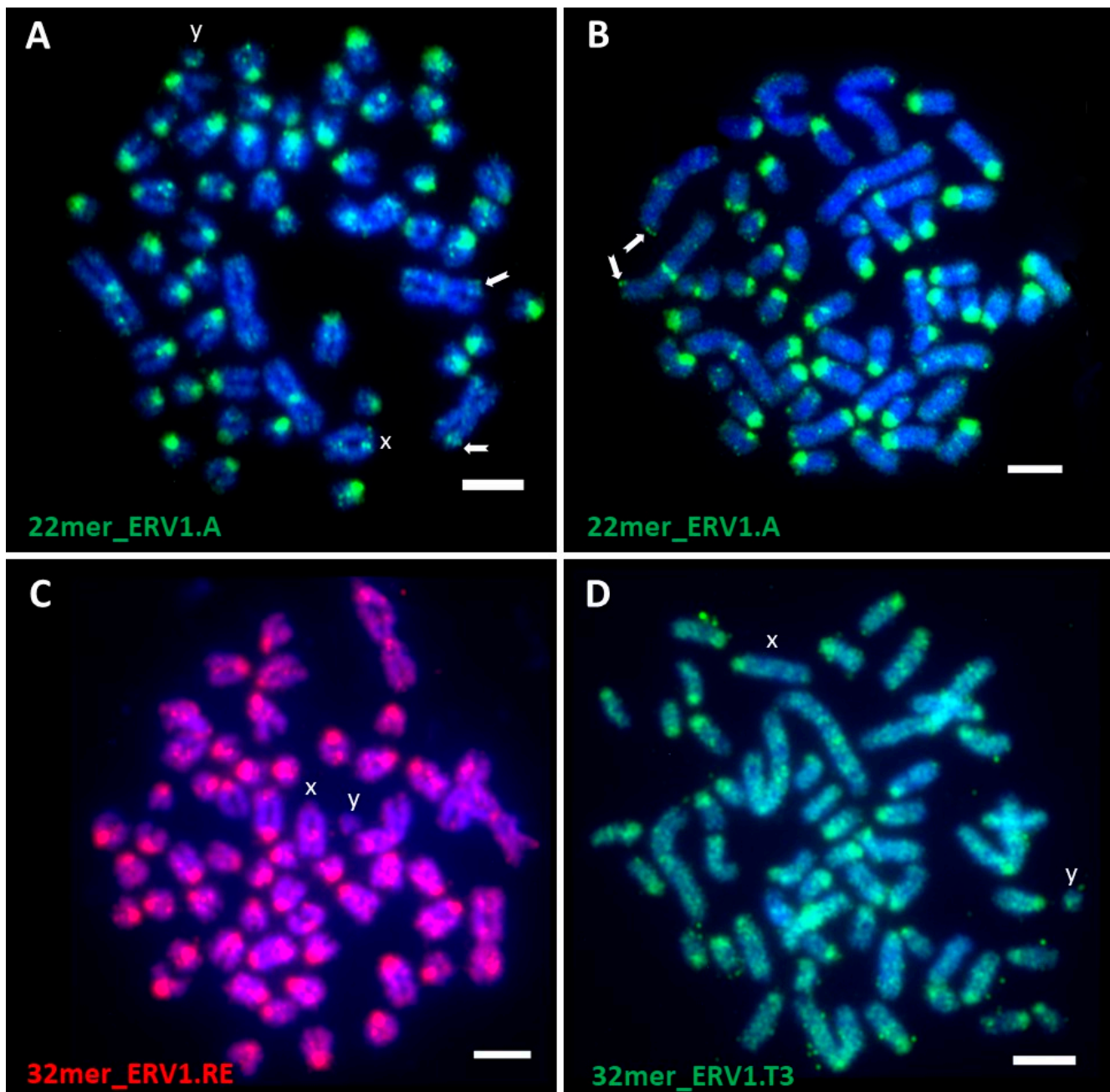


Figure 2. In situ hybridization of ERV1 probes (22 and 32 mer) from k-mers analysis to the male sheep metaphase chromosomes ($2n = 54$) (fluorescing blue with DAPI). (A,B). Probe 22mer_ERV1.A (green) labelled the centromeric regions of all acrocentric chromosomes. It showed small dots over the sex chromosomes X and Y (marked). Weak signals and very small dots were seen at centromeric and subtelomeric regions of submetacentrics (arrows). (C). Probe 32mer_ERV1.RE (magenta) was concentrated at centromeres with some additional dispersed signal. Notable signals were also seen on the X and Y chromosomes. (D). Probe 32mer_ERV1.T3 (green) showed centromeric and dispersed signal over all chromosomes Scale bar equals 5 μm .

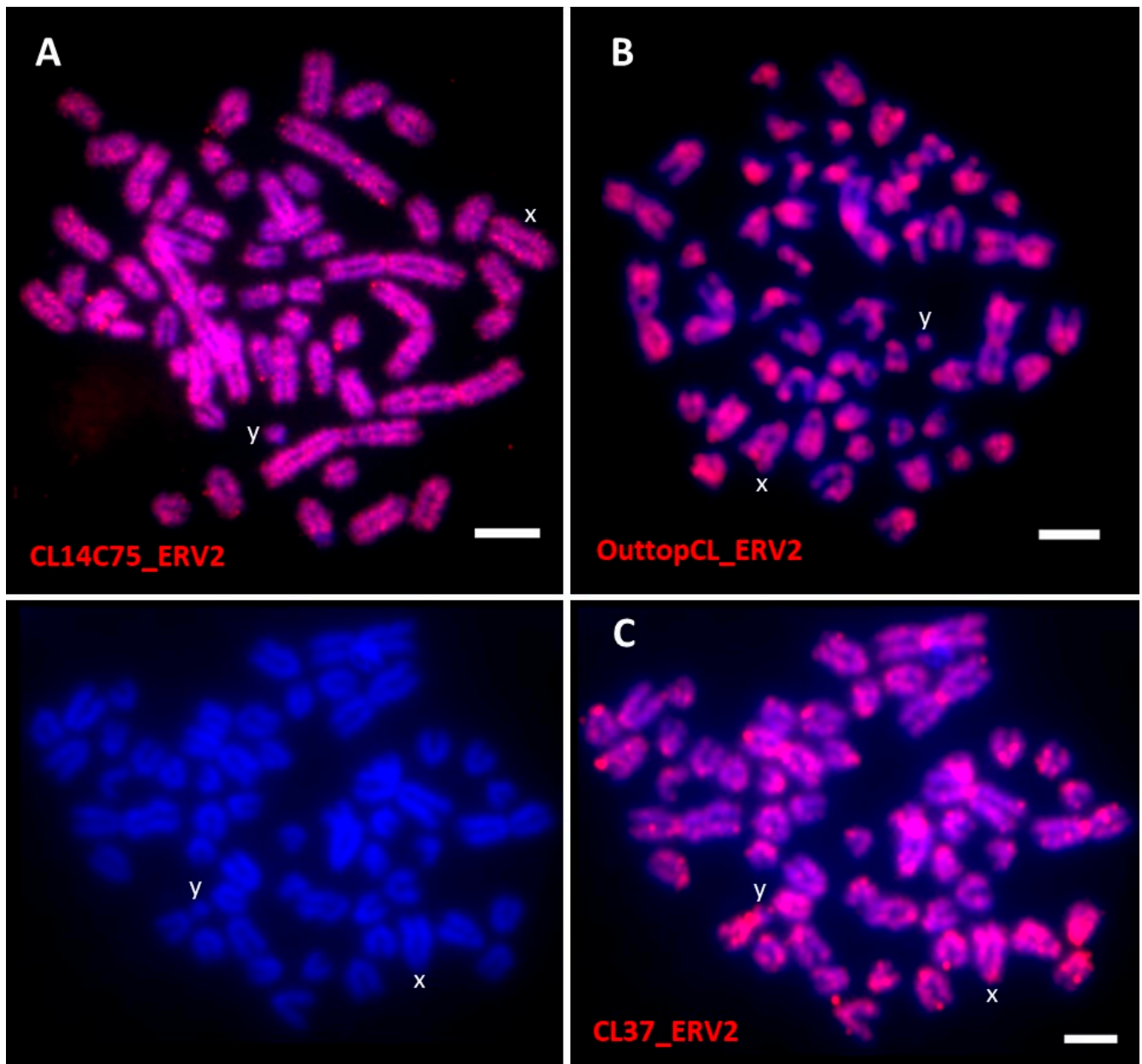


Figure 3. In situ hybridization of ERV2 probes (CL14C75_ERV2, OuttopCL_ERV2; CL37_ERV2) from RepeatExplorer clusters to the male sheep metaphase chromosomes (2n = 54) (fluorescing blue with DAPI). (A). Signals of probe CL14C75_ERV2 (Magenta) were broadly dispersed on all chromosomes with a banding like pattern. The probe also hybridized to the sex chromosomes X and Y (marked). (B). Probe OuttopCL_ERV2 (magenta) showed signals at both centromeric and telomeric domains of all chromosomes including the X and Y chromosomes. (C). DAPI image on the left and probe CL37_ERV2 (magenta on the right). Probe signals are distributed over most centromeric regions and dispersed along the arms of all chromosomes. Scale bar equals 5 μ m.

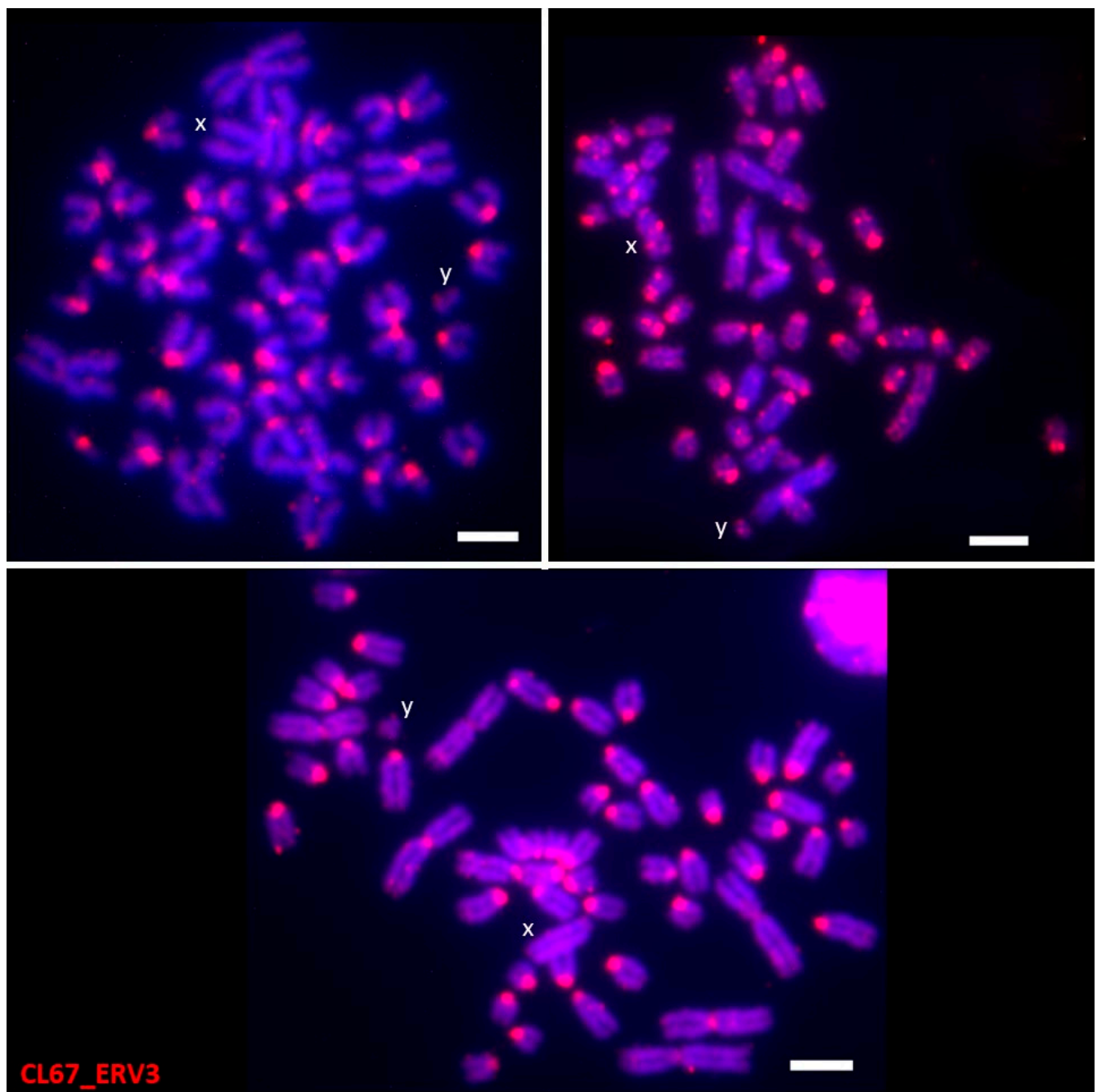


Figure 4. In situ hybridization of ERV3 probes (CL67_ERV3) from RepeatExplorer clusters to the male sheep metaphase chromosomes ($2n = 54$) (fluorescing blue with DAPI). Probe CL67_ERV3 (magenta) showed much more abundant signals at centromeres of all acrocentrics, and one submetacentric pair, while weaker signals were seen on the other two pairs of submetacentrics. A few weak bands or double dots or slight signals were seen throughout chromosome. Signals were also present on sex chromosomes X and Y (marked). Scale bar equals 5 μm .

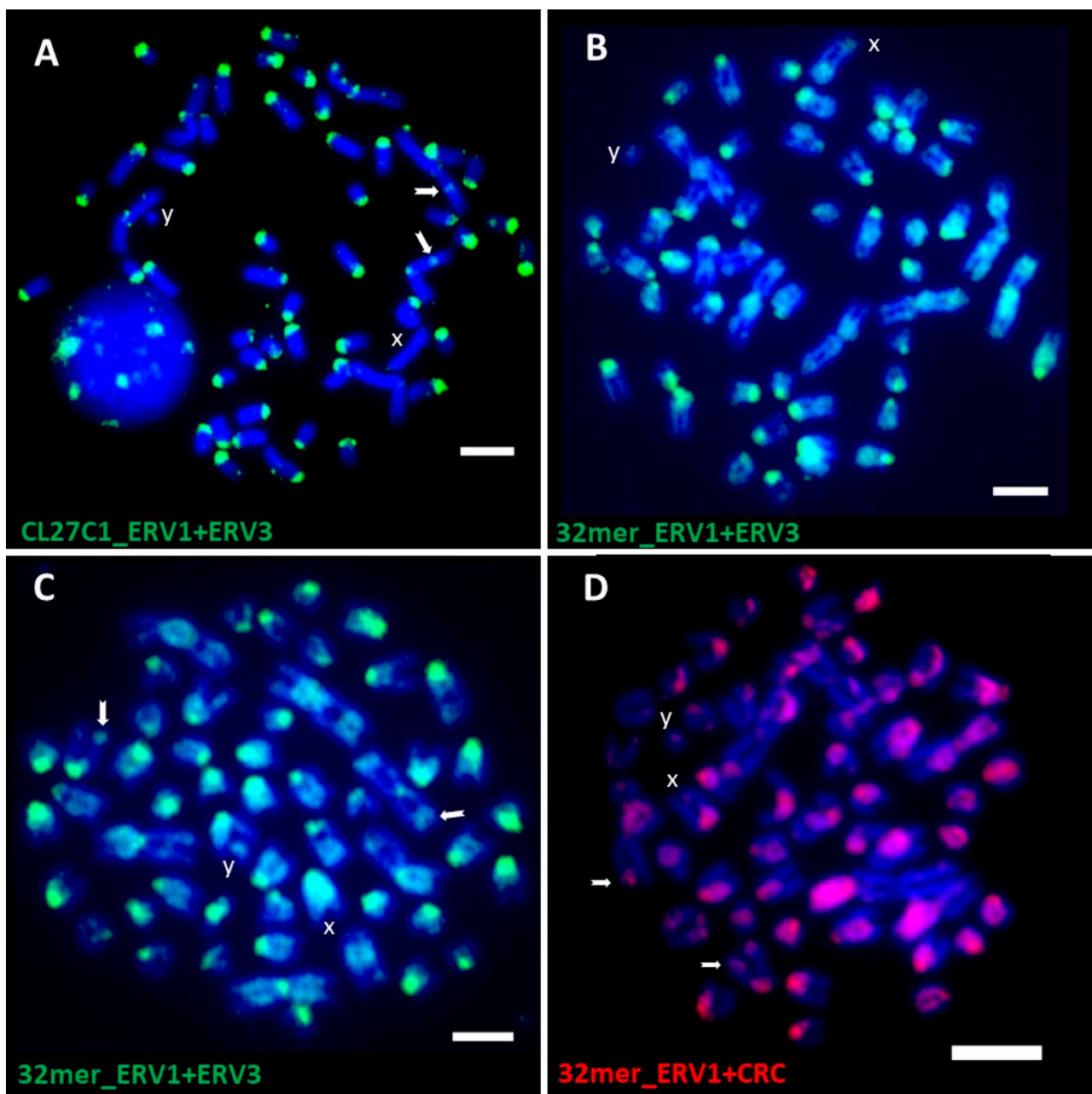


Figure 5. In situ hybridization of combined ERV1+ERV3 probes (CL27C1_ERV1+ERV3; 32mer_ERV1+ERV3) and ERV1-satellite-like sequences probe (32mer_ERV1+CRC) from RepeatExplorer clusters and k-mers analysis to the male sheep metaphase chromosomes ($2n = 54$) (fluorescing blue with DAPI). (A). Probe CL27C1_ERV1+ERV3 (green) strongly hybridized to the centromeres of all acrocentrics, while weak signals can be seen at centromeric and some intercalary positions (arrows) of some of submetacentrics. Signals were not seen on sex chromosomes X and Y. (B,C). Probe 32mer_ERV1+ERV3 (green) two metaphases showing centromeric signals with some bands or broader sites along arms of chromosomes. Telomeric signals on some acrocentrics and submetacentrics were present (arrows (D)). Probe 32mer_ERV1+CRC (magenta) showed different signals, some were present on centromeric locations, while one arm of a submetacentric chromosome pair has extended signal. There were small signals on the sex chromosomes (X and Y, marked) and the telomeric regions of acrocentric and submetacentrics (arrows). Scale bar equals 5 μm .

Table 4. Chromosomal characterization of all classes of ERVs.

ERV Probes	Chromosomes	Localization	Figures
CL18C5_ERV1	All chromosomes except XY and the largest pair of submetacentrics	Centromere	1
CL20C5_ERV1	All chromosomes except XY and the largest pair of submetacentrics	Centromere	
CL23C4_ERV1	All chromosomes	Centromere to disperse	
CL25_ERV1	All chromosomes	Centromere- and dispersed-like dots	
22mer_ERV1.A	All chromosomes	Centromere of all acrocentrics, XY, pair of submetacentrics; Telomere of other 2 pairs of submetacentric	2
32mer_ERV1.RE	All chromosomes	Centromere to disperse	
32mer_ERV1.T3	All chromosomes	Centromere to disperse	
CL14C75_ERV2	All chromosomes	Dispersed	3
OuttopCL_ERV2	All chromosomes	Centromere to disperse	
CL37_ERV2	All chromosomes	Centromere to disperse	
CL67_ERV3	All chromosomes	Centromere- and dispersed-like dots	4
CL27C1_ERV1+ERV3	All chromosomes except XY	Centromere and subtelomere	5
32mer_ERV1+ERV3	All chromosomes	Centromere to disperse and few telomere	
32mer_ERV1+CRC	Most chromosomes	Centromere and few telomere	

3.3.2. ERV2

The chromosomal location of ERV2 showed different patterns to ERV1 (Figure 3). Signals of probe CL14C75_ERV2 were broadly dispersed on both autosomes and sex chromosomes, while some DAPI gaps were seen on some chromosomes (Figure 3A). The OuttopCL_ERV2 probe showed signals at both centromeric and telomeric domains of acrocentric chromosomes, the submetacentrics had banding-like signals and sex chromosomes had strong signals (Figure 3B). Probe CL37_ERV2 showed signals scattered over centromeric regions of some acrocentric and submetacentrics, and slight dot-like signals were seen on all chromosomes including X and Y (Figure 3C).

3.3.3. ERV3

ERV3 sequences represented in the probe of CL67_ERV3 formed strong signals at centromeres of all acrocentric chromosomes, and one submetacentric pair, while the other two pairs of submetacentrics and the sex chromosomes had weaker signals (Figure 4). Consensus sequences of some clusters and k-mer contigs included similarity to both ERV1 and ERV3. The CL27C1_ERV1+ERV3 probe strongly hybridized to the centromeres of all acrocentrics, while centromeric and subtelomeric regions of some submetacentrics had weak signals. Sex chromosomes had no signals (Figure 5A). Similarly, probe 32mer_ERV1+ERV3 showed centromeric signals with some bands or broader sites along arms of chromosomes. Telomeric signals on some acrocentrics and submetacentrics were seen. Signals were also incorporated in X and Y chromosomes (Figure 5B,C).

3.3.4. Combined ERV1 and Satellite like Sequences

Combined sequences of ERV1 and satellite-like repeats was found in the results of both k-mer analysis and RepeatExplorer, suggesting some interspersed of satellite and ERV sequences (Supplementary Materials, Figure S3). Probe 32mer_ERV1+CRC hybridized with the centromeric and the telomeric regions of some acrocentric and submetacentric chromosomes. Some dots were seen on sex chromosomes (Figure 5D). Probe CRC (with half of the consensus of 32mer_ERV1 and half CRC) showed 60% similarity to the centromeric repetitive DNA from Cervidae species such as *Muntiacus muntjak vaginalis* (AY064466-AY064469).

3.4. The Complete Genome of the Endogenous Jaagsiekte Sheep Retrovirus (enJSRV)

The complete consensus endogenous beta retroviruses enJSRV genome of three Hamdani and two Karadi sheep breed were assembled. Each was 7941 bp long and estimated to be present in 71 to 124 copies (0.0087% to 0.0118% genomic proportion) (Table 5). The enJSRV sequences, gene features (predicted proteins, with start and stop codons) and other characteristics such as position, sizes and strand distribution of all 4 protein-coding genes (gag, pro, pol and env genes) and LTR repeats are available in GenBank of the NCBI under accession numbers MF175067, MF175068, MF175069, MF175070 and MF175071. The phylogenetic analysis indicated that the consensus sequences of enJSRV from the large fat-tailed sheep breeds from the Iraqi Kurdistan region was placed with the recognized *Ovis aries* clade including EF680302 and DQ838493 Jaagsiekte sheep retrovirus sampled from geographically different locations (Figure 6).

Table 5. Copy numbers and genomic proportion of complete genomes of the endogenous beta retroviruses enJSRV integrated in the main sheep breeds of Iraqi Kurdistan region.

enJSRV Breeds (GenBank ID)	Complete enJSRV Genome bp	Assembled Reads	Total Reads of Each Genome (Coverage X)	Genomic Proportion %	Copies of enJSRV (Assembled Reads*150/7941)
HamJ1 (MF175067)	7941	5390	52,048,068 (2.6)	0.0104	101.81
HamJ2 (MF175068)	7941	6606	56,220,882 (2.81)	0.0118	124.78
HamM (MF175069)	7941	3809	43,596,654 (2.18)	0.0087	71.95
KaJ (MF175070)	7941	5386	60,605,648 (3.03)	0.0089	101.74
KarM (MF175071)	7941	4846	44,933,034 (2.25)	0.0108	91.54

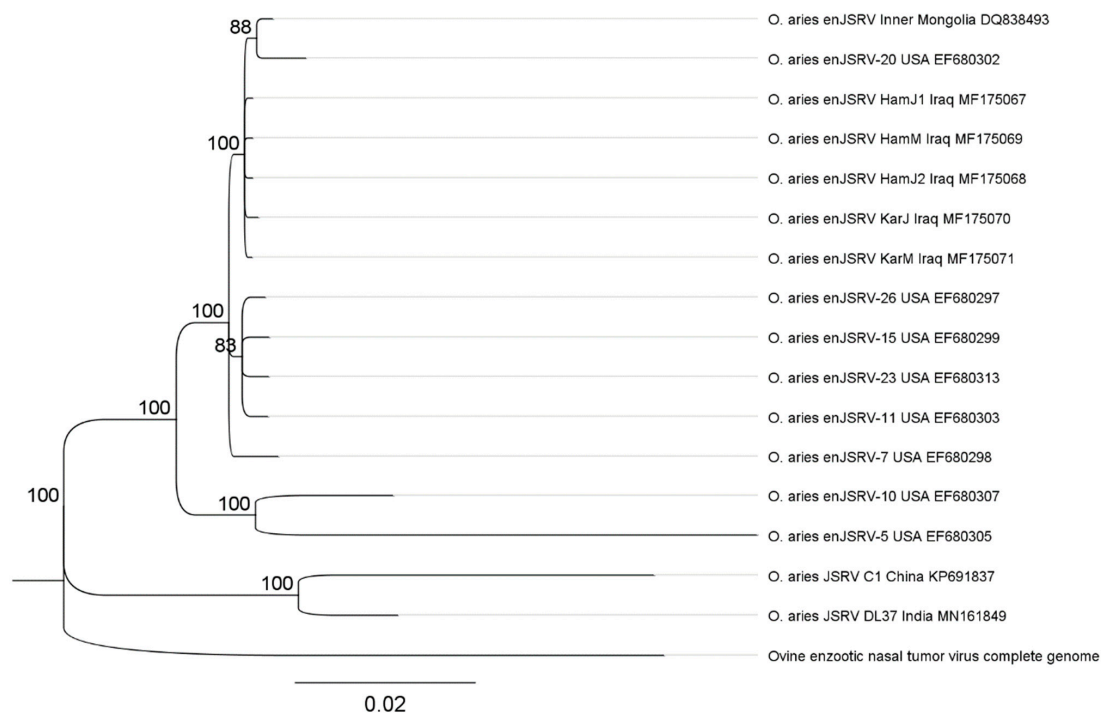


Figure 6. Phylogenetic relationship showing position of complete genomes of endogenous beta retroviruses (enJSRV) of Iraqi sheep breeds (Hamdani; HamJ1, HamJ2, HamM and Karadi; KarM, KarJ) in relation to other strains of enJSRV from other sheep breeds worldwide. Nodes are labelled with consensus support%. Scale bar indicate the length of the branches of the tree (Scale Range= 0.02 and Line Weight = 1).

4. Discussion

4.1. Genomic Distribution and Chromosomal Organization of ERVs

Major classes of endogenous retroviruses in the whole genome raw reads of sheep genome were identified using graph-based approaches (RepeatExplorer) [19,20] and analysis of abundant k-mers. Fluorescent in situ hybridization (FISH) showed probes from individual ERV sequence families have characteristic but different chromosomal distribution patterns and abundance over the sheep autosomal, acrocentric, metacentric and sex chromosomes. The recognized subfamilies of ERV1, ERV2 and ERV3 showed abundant centromeric to dispersed distribution over sheep chromosomes (Figures 1–5 and Table 4). In humans, novel sequences of endogenous retrovirus HERV-K (HML-2) provirus named (K111) have been discovered in centromeric regions of several chromosomes in multiple copies [24,25]. The authors proposed that the retroviral sequences in human genomes have been templates for regular cross-over events during the recombination process leading to amplification of repetitive DNA sequences at the centromere. Likewise, in the current study, we have found strong hybridization of ERV sequences in the centromeric regions of sheep chromosomes. Thus, ERVs are either amplified or accumulate at centromeres, with transposition or retrotransposition and mutation generating to complex family [26]. It is notable that the submetacentric autosomes (derived by fusion of acrocentric chromosomes in the ancestral Bovidae karyotype) show contrasting distributions for some ERVs, as has been found for the major centromeric satellite sequences [27]. In kangaroo genomes, amplification of endogenous retrovirus occurred in a lineage-specific fashion which is limited to the centromeres of chromosomes [28]. It will be interesting to investigate whether ERVs (in the centromeric and pericentromeric domains depleted in genes) [29], have any involvement in the karyotypic rearrangement and the evolutionary fusion of the ancestral acrocentric chromosomes, or are carried with the more abundant satellite DNA sequences as has been suggested in bovine chromosome fusions [30].

The genomic proportion of ERVs in the sheep genome was estimated as 0.55% in the raw reads from RepeatExplorer, very similar to the proportion aligned to reference bovid ERV sequences from Repbase database [31,32] (0.02% and 0.45% of sheep raw reads have a high similarity to the ancestral and *B. taurus* ERV sequences respectively). In three species of Bos; cattle (*B. taurus*), zebu (*B. taurus* spp. *indicus*), and yaks (*Bos grunniens*), about 30 ERV types were identified and these ERVs were also detected with different percentages in eight species of animals using different types of data (assembly; contigs and reads) [33]. In human, approximately 8% of the genome constitutes of retroviral origin sequences, considered a result of continuous infections of the germ line of the host lineage by ancient viruses over millions of years of evolution [2,34], much higher than the data here suggests are found in sheep.

The biological significance of ERVs has been argued, but ERVs are now considered to have a variety of beneficial roles in their host genome contributing to genome plasticity [35–37]. In plants, the endogenous pararetrovirus sequences incorporated in the genome, first discovered in banana by in situ hybridization [38], are now thought to protect the host via RNAi mechanisms [39]. Correspondingly, in sheep, protective mutation has been described in a copy of the enJSRV which emerged shortly after integration of JSRV in the 6q13 locus before diversification of *Ovis* and after divergence of *Capra* and *Ovis* [40]. The genomic amplifications and fixation of such protective mutation in *Ovis* species may be due to the natural selection before domestication of *O. aries* from *Ovis orientalis*. Thus, the integrated ERVs in the genome of Iraqi sheep may have a protective role in reducing retroviral infection. Furthermore, ovine enJSRVs have contributed to the protection of the uterus from viral infection and regulators of placental function and morphogenesis. For example, mRNA of enJSRVs env was found expressed in the trophoblast giant binucleate cells (BNC), glandular epithelium (GE), the endometrial luminal epithelium (LE) and multinucleated syncytia of the placenta suggesting that their expression played important physiological roles in conceptus implantation and placentation [41,42].

4.2. Complete Genome of Endogenous Betaretroviruses (enJSRV) and Their Abundance in Iraqi Sheep Breeds

Various endogenous retroviruses from different genera have been characterized from a variety of mammalian species [43]. Endogenous retroviruses can be categorized through comparison of sequences in a phylogeny [35]. Based on the enJSRV genome, the five assembled Iraqi endogenous beta retroviruses (enJSRV) genomes were grouped together and placed on sister branches with other enJSRV proviruses in sheep within an enJSRV clade (Figure 6). Although the five Iraqi enJSRV genome clustered together, several nucleotide differences were found (Supplementary Materials, Figure S5). In the present study, we assembled complete genome of enJSRV from HTS data, while others cloned the complete beta retroviruses from sheep sampled from Inner Mongolia [23]. Our study indicates that enJSRV sequences from the Iraqi sheep are most similar to those in sheep breeds from China and USA, signifying that there are potential homogenization and fixation of the enJSRV sequences in sheep genome despite the presence of polymorphic sites between them. enJSRV sequences are thought to have entered the host genome within the last 3 million years, before and during speciation within the genus *Ovis*, and are characterized by a transdominant phenotype able to block late replication steps of related exogenous retroviruses [15].

About 27 enJSRV proviruses were isolated and characterized in the genomes of different species of genus *Ovis* within the Caprinae subfamily including (*O. aries*, *Ovis ammon*, *Ovis canadensis* and *Ovis dalli*) [15]. Furthermore, copies of enJSRVs and their integration sites in domestic and wild species of the sheep lineage were detected by amplification the env-LTR region by PCR, and 103 enJSRV sequences were produced across 10 individuals and enJSRV integrations were found on 11 of the 28 sheep chromosomes [44] (see FISH results). Similarly, we found that about 294 sequence fragments with different lengths (37 bp–7899 bp; total length 136 kbp) from the nuclear chromosome assemblies of *O. aries* Oar_v4.0 databases have high similarity to the genomic sequences of enJSRV. Correspondingly, in the cow genome about 928, 4487, 9698 ERVs related sequences were detected using three different methods, BLAST-based searches, LTR_STRUC and RetroTector, respectively [43]. The current study is the first time assembling the complete genome of enJSRV of Iraqi sheep breeds from the center of domestication and diversity species and it also the first work investigating copy numbers 72 to 125 copies and genomic proportions 0.0087% to 0.0118% of complete genome of endogenous beta retroviruses (enJSRV) using whole genome high-throughput sequencing (HTS) data of sheep breeds. The retroviral *pro-pol* sequences of two retroviral families (B/D-type) and (C-type) and the copy numbers (5–100 copies) of B type ERVs in several sheep breeds using Southern blot analysis were analyzed [45]. The ERV sequences identified here fall within the range of diversity previously reported, but form a distinct group (Figure 6), suggesting that there is either homogenization of the sequences (including potentially through gene conversion), or loss and replacement with new copies. However, unlike retroelements, it is notable that the copy number of enJSRV is similar across all sheep breeds studied. The results here show that the sheep breeds near the center of diversity and domestication have similar abundances of enJSRV beta retroviruses and more generally endogenous retroviruses, ERVs to other sheep.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/dna2010007/s1>. Table S1, DNA samples and sample locations used for high-throughput sequencing (HTS) data. F; female and M; male, Figure S1. Mapping of high-throughput sequencing (HTS) data of sheep to ancestral sequences of ERVs, Figure S2. Mapping of high-throughput sequencing (HTS) data of sheep to *Bos taurus* sequences of ERVs, Figure S3. Consensus of CL15C14 (4191bp) of RepeatExplorer including combined ERV1+CRC sequences, three copies of 32merC16_Sat_CRC satellite like sequences and ERVs, Figure S4. Assembly of high-throughput sequencing (HTS) data to reference complete genome of enJSRV DQ838493, Figure S5. Distribution of 50 SNPs along the complete genome of enJSRV of five samples of the Iraqi sheep breeds Hamdani and Karadi.

Author Contributions: Conceptualization, S.I.M., T.S. and J.S.H.-H.; methodology, S.I.M., T.S. and J.S.H.-H.; software, validation, and formal analysis, S.I.M. and J.S.H.-H.; investigation, S.I.M., T.S.; resources and data curation, S.I.M. and J.S.H.-H.; writing—original draft preparation, S.I.M.; writing—review and editing, S.I.M., T.S. and J.S.H.-H.; visualization, S.I.M.; supervision and project administration, T.S. and J.S.H.-H.; funding acquisition, S.I.M. and J.S.H.-H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a PhD scholarship (Human Capacity Development Program; HCDP) from the Kurdistan Region Government-Iraq to S.I.M.

Institutional Review Board Statement: The animal study protocol was carried out with the approval of the scientific committee (Official Meeting No.7 on 26/01/2015) of the Department of Animal Production, College of Agriculture, University of Duhok, Iraq. Fresh blood for chromosome preparation was obtained post-mortem from a UK licensed abattoir (number UK4327), Leicester, UK.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data pertaining to the present study have been included in Table and/or Figure form in the manuscript and authors are pleased to share analyzed/raw data upon reasonable request.

Acknowledgments: The authors thank the staff of JM Morris, Lutterworth, UK for provision of sheep blood.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Biscotti, M.A.; Olmo, E.; Heslop-Harrison, J.S. (Pat) Repetitive DNA in Eukaryotic Genomes. *Chromosom. Res.* **2015**, *23*, 415–420. [[CrossRef](#)] [[PubMed](#)]
2. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; Fitzhugh, W.; et al. Initial Sequencing and Analysis of the Human Genome. *Nature* **2001**, *409*, 860–921. [[CrossRef](#)] [[PubMed](#)]
3. Mikkelsen, T.S.; Hillier, L.W.; Eichler, E.E.; Zody, M.C.; Jaffe, D.B.; Yang, S.P.; Enard, W.; Hellmann, I.; Lindblad-Toh, K.; Altheide, T.K.; et al. Initial Sequence of the Chimpanzee Genome and Comparison with the Human Genome. *Nature* **2005**, *437*, 69–87. [[CrossRef](#)]
4. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A Unified Classification System for Eukaryotic Transposable Elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [[CrossRef](#)] [[PubMed](#)]
5. Mager, D.L.; Stoye, J.P. Mammalian Endogenous Retroviruses. *Mob. DNA III* **2015**, 1079–1100. [[CrossRef](#)]
6. Feschotte, C.; Gilbert, C. Endogenous Viruses: Insights into Viral Evolution and Impact on Host Biology. *Nat. Rev. Genet.* **2012**, *13*, 283–296. [[CrossRef](#)] [[PubMed](#)]
7. Bannert, N.; Kurth, R. The Evolutionary Dynamics of Human Endogenous Retroviral Families. *Annu. Rev. Genomics Hum. Genet.* **2006**, *7*, 149–173. [[CrossRef](#)] [[PubMed](#)]
8. Weiss, R.A. The Discovery of Endogenous Retroviruses. *Retrovirology* **2006**, *3*, 1–11. [[CrossRef](#)]
9. Kojima, K.K. Human Transposable Elements in Repbase: Genomic Footprints from Fish to Humans. *Mob. DNA* **2018**, *9*, 1–14. [[CrossRef](#)] [[PubMed](#)]
10. Zhuo, X.; Feschotte, C. Cross-Species Transmission and Differential Fate of an Endogenous Retrovirus in Three Mammal Lineages. *PLoS Pathog.* **2015**, *11*, e1005279. [[CrossRef](#)]
11. Garcia-Montojo, M.; Doucet-O'Hare, T.; Henderson, L.; Nath, A. Human Endogenous Retrovirus-K (HML-2): A Comprehensive Review. *Crit. Rev. Microbiol.* **2018**, *44*, 715–738. [[CrossRef](#)] [[PubMed](#)]
12. Katoh, I.; Kurata, S.I. Association of Endogenous Retroviruses and Long Terminal Repeats with Human Disorders. *Front. Oncol.* **2013**, *3*, 1–8. [[CrossRef](#)] [[PubMed](#)]
13. Balestrieri, E.; Pica, F.; Matteucci, C.; Zenobi, R.; Sorrentino, R.; Argaw-Denboba, A.; Cipriani, C.; Bucci, I.; Sinibaldi-Vallebona, P. Transcriptional Activity of Human Endogenous Retroviruses in Human Peripheral Blood Mononuclear Cells. *Biomed Res. Int.* **2015**, *2015*, 164529. [[CrossRef](#)] [[PubMed](#)]
14. Grandi, N.; Tramontano, E. Type W Human Endogenous Retrovirus (HERV-W) Integrations and Their Mobilization by L1 Machinery: Contribution to the Human Transcriptome and Impact on the Host Physiopathology. *Viruses* **2017**, *9*, 162. [[CrossRef](#)] [[PubMed](#)]
15. Arnaud, F.; Caporale, M.; Varela, M.; Biek, R.; Chessa, B.; Alberti, A.; Golder, M.; Mura, M.; Zhang, Y.P.; Yu, L.; et al. A Paradigm for Virus-Host Coevolution: Sequential Counter-Adaptations between Endogenous and Exogenous Retroviruses. *PLoS Pathog.* **2007**, *3*, 1716–1729. [[CrossRef](#)] [[PubMed](#)]
16. Palmarini, M.; Mura, M.; Spencer, T.E. Endogenous Betaretroviruses of Sheep: Teaching New Lessons in Retroviral Interference and Adaptation. *J. Gen. Virol.* **2004**, *85*, 1–13. [[CrossRef](#)]

17. Murcia, P.R.; Arnaud, F.; Palmarini, M. The Transdominant Endogenous Retrovirus EnJS56A1 Associates with and Blocks Intracellular Trafficking of Jaagsiekte Sheep Retrovirus Gag. *J. Virol.* **2007**, *81*, 1762–1772. [\[CrossRef\]](#)
18. Qi, J.W.; Xu, M.J.; Liu, S.Y.; Zhang, Y.F.; Liu, Y.; Zhang, Y.K.; Cao, G.F. Identification of Sheep Endogenous Beta-Retroviruses with Uterus-Specific Expression in the Pregnant Mongolian Ewe. *J. Integr. Agric.* **2013**, *12*, 884–891. [\[CrossRef\]](#)
19. Novák, P.; Neumann, P.; Macas, J. Graph-Based Clustering and Characterization of Repetitive Sequences in next-Generation Sequencing Data. *BMC Bioinformatics* **2010**, *11*, 378. [\[CrossRef\]](#)
20. Novák, P.; Neumann, P.; Pech, J.; Steinhaisl, J.; Macas, J. RepeatExplorer: A Galaxy-Based Web Server for Genome-Wide Characterization of Eukaryotic Repetitive Elements from next-Generation Sequence Reads. *Bioinformatics* **2013**, *29*, 792–793. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Marçais, G.; Kingsford, C. A Fast, Lock-Free Approach for Efficient Parallel Counting of Occurrences of k-Mers. *Bioinformatics* **2011**, *27*, 764–770. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data. *Bioinformatics* **2012**, *28*, 1647–1649. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Wang, Y.; Liu, S.Y.; Li, J.Y.; Han, M.; Wang, Z.L. Cloning and Sequence Analysis of Genome from the Inner Mongolia Strain of the Endogenous Betaretroviruses (EnJSRV). *Virol. Sin.* **2008**, *23*, 15–24. [\[CrossRef\]](#)
24. Contreras-Galindo, R.; Kaplan, M.H.; He, S.; Contreras-Galindo, A.C.; Gonzalez-Hernandez, M.J.; Kappes, F.; Dube, D.; Chan, S.M.; Robinson, D.; Meng, F.; et al. HIV Infection Reveals Widespread Expansion of Novel Centromeric Human Endogenous Retroviruses. *Genome Res.* **2013**, *23*, 1505–1513. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Zahn, J.; Kaplan, M.H.; Fischer, S.; Dai, M.; Meng, F.; Saha, A.K.; Cervantes, P.; Chan, S.M.; Dube, D.; Omenn, G.S.; et al. Expansion of a Novel Endogenous Retrovirus throughout the Pericentromeres of Modern Humans. *Genome Biol.* **2015**, *16*, 1–24. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Prudhomme, S.; Bonnaud, B.; Mallet, F. Endogenous Retroviruses and Animal Reproduction. *Cytogenet. Genome Res.* **2005**, *110*, 353–364. [\[CrossRef\]](#)
27. Chaves, R.; Guedes-Pinto, H.; Heslop-Harrison, J.S.; Schwarzacher, T. The Species and Chromosomal Distribution of the Centromeric α -Satellite I Sequence from Sheep in the Tribe Caprini and Other Bovidae. *Cytogenet. Genome Res.* **2000**, *91*, 62–66. [\[CrossRef\]](#)
28. Ferreri, G.C.; Brown, J.D.; Oberfell, C.; Jue, N.; Finn, C.E.; O'Neill, M.J.; O'Neill, R.J. Recent Amplification of the Kangaroo Endogenous Retrovirus, KERV, Limited to the Centromere. *J. Virol.* **2011**, *85*, 4761–4771. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Lomiento, M.; Jiang, Z.; D'Addabbo, P.; Eichler, E.E.; Rocchi, M. Evolutionary-New Centromeres Preferentially Emerge within Gene Deserts. *Genome Biol.* **2008**, *9*, R173. [\[CrossRef\]](#)
30. Escudeiro, A.; Ferreira, D.; Mendes-da-Silva, A.; Heslop-Harrison, J.S.; Adegas, F.; Chaves, R. Bovine Satellite DNAs—a History of the Evolution of Complexity and Its Impact in the Bovidae Family. *Eur. Zool. J.* **2019**, *86*, 20–37. [\[CrossRef\]](#)
31. Jurka, J.; Kapitonov, V.V.; Pavlicek, A.; Klonowski, P.; Kohany, O.; Walichiewicz, J. Repbase Update, a Database of Eukaryotic Repetitive Elements. *Cytogenet. Genome Res.* **2005**, *110*, 462–467. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Bao, W.; Kojima, K.K.; Kohany, O. Repbase Update, a Database of Repetitive Elements in Eukaryotic Genomes. *Mob. DNA* **2015**, *6*, 4–9. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Garcia-Etxebarria, K.; Jugo, B.M. Evolutionary History of Bovine Endogenous Retroviruses in the Bovidae Family. *BMC Evol. Biol.* **2013**, *13*, 256. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Pačes, J.; Pavlicek, A.; Pačes, V. HERVd: Database of Human Endogenous Retroviruses. *Nucleic Acids Res.* **2002**, *30*, 205–206. [\[CrossRef\]](#)
35. Jern, P.; Coffin, J.M. Effects of Retroviruses on Host Genome Function. *Annu. Rev. Genet.* **2008**, *42*, 709–732. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Varela, M.; Spencer, T.E.; Palmarini, M.; Arnaud, F. Friendly Viruses: The Special Relationship between Endogenous Retroviruses and Their Host. *Ann. N. Y. Acad. Sci.* **2009**, *1178*, 157–172. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Kurth, R.; Bannert, N. Beneficial and Detrimental Effects of Human Endogenous Retroviruses. *Int. J. Cancer* **2010**, *126*, 306–314. [\[CrossRef\]](#)
38. Harper, G.; Osuji, J.O.; Heslop-Harrison, J.S.; Hull, R. Integration of Banana Streak Badnavirus into the Musa Genome: Molecular and Cytogenetic Evidence. *Virology* **1999**, *255*, 207–213. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Noreen, F.; Akbergenov, R.; Hohn, T.; Richert-Pöggeler, K.R. Distinct Expression of Endogenous Petunia Vein Clearing Virus and the DNA Transposon DTph1 in Two Petunia Hybrid Lines Is Correlated with Differences in Histone Modification and siRNA Production. *Plant J.* **2007**, *50*, 219–229. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Cumer, T.; Pompanon, F.; Boyer, F. Old Origin of a Protective Endogenous Retrovirus (EnJSRV) in the Ovis Genus. *Heredity* **2019**, *122*, 187–194. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Dunlap, K.A.; Palmarini, M.; Adelson, D.L.; Spencer, T.E. Sheep Endogenous Betaretroviruses (EnJSRVs) and the Hyaluronidase 2 (HYAL2) Receptor in the Ovine Uterus and Conceptus. *Biol. Reprod.* **2005**, *73*, 271–279. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Dunlap, K.A.; Palmarini, M.; Spencer, T.E. Ovine Endogenous Betaretroviruses (EnJSRVs) and Placental Morphogenesis. *Placenta* **2006**, *27*, 135–140. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Garcia-Etxebarria, K.; Jugo, B.M. Genome-Wide Detection and Characterization of Endogenous Retroviruses in Bos Taurus. *J. Virol.* **2010**, *84*, 10852–10862. [\[CrossRef\]](#) [\[PubMed\]](#)

-
44. Sistiaga-Poveda, M.; Jugo, B.M. Evolutionary Dynamics of Endogenous Jaagsiekte Sheep Retroviruses Proliferation in the Domestic Sheep, Mouflon and Pyrenean Chamois. *Heredity (Edinb)*. **2014**, *112*, 571–578. [[CrossRef](#)]
 45. Nikolai, K.; Mathias, M.; Gottfried, B.; Bernhard, A. Characterization of Endogenous Retroviruses in Sheep. *J. Virol.* **2003**, *77*, 11268–11273. [[CrossRef](#)]