Genome-Wide Detection and Characterization of Endogenous Retroviruses in *Bos taurus*[⊽]

Koldo Garcia-Etxebarria and Begoña Marina Jugo*

Genetika, Antropologia Fisikoa eta Animalien Fisiologia Saila, Zientzia eta Teknologia Fakultatea, Euskal Herriko Unibertsitatea, 644 Postakutxa, E-48080 Bilbao, Spain

Received 16 January 2010/Accepted 21 July 2010

Endogenous retroviruses (ERVs) are the proviral phase of exogenous retroviruses that become integrated into a host germ line. They can play an important role in the host genome. Bioinformatic tools have been used to detect ERVs in several vertebrates, primarily primates and rodents. Less information is available regarding ERVs in other mammalian groups, and the source of this information is basically experimental. We analyzed the genome of the cow (*Bos taurus*) using three different methods. A BLAST-based method detected 928 possible ERVs, LTR_STRUC detected 4,487 elements flanked by long terminal repeats (LTRs), and Retrotector detected 9,698 ERVs. The ERVs were not homogeneously distributed across chromosomes; the number of ERVs was positively correlated with chromosomal size and negatively correlated with chromosomal GC content. The bovine ERVs (BoERVs) were classified into 24 putative families, with 20 of them not previously described. One of these new families, BoERV1, was the most abundant family and appeared to be specific to ruminants. An analysis of representatives of ERV families from rodents, primates, and ruminants showed a phylogenetic relationship following their hosts' relationships. This study demonstrates the importance of using multiple methods when trying to identify new ERVs and shows that the number of bovine ERV families is not as limited as previously thought.

Endogenous retroviruses (ERVs) are the proviral phase of exogenous retroviruses that were once inserted into a host germ line and have remained integrated into the host genome for generations. ERVs have been detected in all mammals and a wide range of other vertebrates. Their typical structure is composed of a central part with the three major genes (*gag*, *pol*, and *env*) flanked by two long terminal repeats (LTRs) that were identical when the retrovirus entered the host germ line (4).

The biological significance of retrotransposons, including ERVs, ranges from their contributions to mutation, development, and disease to their roles in gene and genome evolution. In humans, mice, and sheep, for example, an env gene of retroviral origin, acquired independently in the different species, is involved in placenta morphogenesis (11). It has been suggested that ERVs could be possible contributors to or markers of disease in experimental animals and, in recent years, in human diseases, although their role as the etiological agent remains to be established (11). The expression of ERVs in humans has been linked to poor prognosis in breast cancer (9) and the malignant transformation of melanoma cells (36) and may play a role in multiple sclerosis (28). In addition, ERV-mediated recombination events have had profound effects on shaping hosts' genomes, and new ERV integrations introduce added variation to the host transcriptomes (11).

At present, there is no well-established or accepted standard for naming and classifying ERVs. For human ERVs (HERVs), tRNA complementary to the primer binding site (PBS) has traditionally been used for this purpose. This classification, however, is inaccurate, as proviruses from the same phylogenetic groups may display differences in the PBS, while otherwise unrelated proviruses may use the same tRNA as a primer. A more useful strategy for classifying ERVs is phylogenetic and related comparisons (11). ERVs are now sorted into three classes depending on the phylogenetic relationship with the exogenous retrovirus genus: class I ERVs are related to the genera *Gammaretrovirus* and *Epsilonretrovirus*; class II ERVs are related to the genera *Alpharetrovirus*, *Betaretrovirus*, *Deltaretrovirus*, and *Lentivirus*; and class III ERVs are related to the genus *Spumavirus* (8).

The availability of whole-genome sequences has made possible genome-wide analyses for the detection of ERVs using bioinformatic tools. In the last years RepeatMasker (A. F. A. Smit, R. Hubley, and P. Green, personal communication), a program designed to identify repetitive sequences using the Repbase database (13), has been widely used to generate an overview of repetitive elements in whole-genome sequences, among them the ERVs. For mammals, other programs have also been used in order to identify ERVs. BLAST-based searches were first used for humans (40, 41) and rodents (2). A program called LTR_STRUC (22) was applied to the chimpanzee genome in combination with a BLAST-based search (29). A new program specifically designed for ERV detection, called Retrotector, has recently been reported (38).

The mammalian order Cetartiodactyla has become a major focus of attention in comparative genomics because it comprises a phylogenetically distant clade of eutherian mammals related to primates, which diverged from a common ancestor ~85 million years ago (18). *Bos taurus* is one of the world's most important food animal species and is also among the

^{*} Corresponding author. Mailing address: Genetika, Antropologia Fisikoa eta Animalien Fisiologia Saila, Zientzia eta Teknologia Fakultatea, Euskal Herriko Unibertsitatea, 644 Postakutxa, E-48080 Bilbao, Spain. Phone: 34 94 6015518. Fax: 34 94 6013145. E-mail: begonamarina .jugo@ehu.es.

 $[\]overline{V}$ Published ahead of print on 4 August 2010.

Determine	Carrier	II+	GenBank	Quer	y sequence po	osition	Dista	nce between a	query sequence	s (bp)
Retrovirus	Genus	Host	accession no.	gag	pol	env	Start to gag	gag to pol	pol to env	env to end
Class I GALV MLV FeLV	Gammaretrovirus Gammaretrovirus Gammaretrovirus	<i>Hylobates</i> Muridae Felidae	NC_001885 NC_001501 NC_001940	1255–1888 1020–1653 1516–2149	2520–3276	6998–7555 6920–7457 7323–7860	1,255 1,020 1,516	846 867 822	3,508 3,644 3,593	533 875 588
Class II JSRV MMTV BLV HTLV EIAV HIV Visna virus	Betaretrovirus Betaretrovirus Deltaretrovirus Deltaretrovirus Lentivirus Lentivirus Lentivirus	Ovis aries Muridae Bos taurus Homo sapiens Equus caballus Homo sapiens Ovis aries	NC_001494 NC_001503 NC_001414 NC_001436 NC_001450 NC_001450 NC_001452	$\begin{array}{c} 1347 - 1667 \\ 1181 - 1774 \\ 790 - 1362 \\ 889 - 1455 \\ 945 - 1562 \\ 1035 - 1401 \\ 973 - 1566 \end{array}$		6502–7111 6786–7362 5760–6158 5765–6269 7235–7867 7727–8261 8176–8806	1,347 1,181 790 889 945 1,035 973	1,444 1,194 572 780 807 765 620	2,830 3,077 2,874 2,786 4,107 4,805 5,237	351 1,443 2,261 2,238 492 920 396
Class III HSRV BFV	Spumavirus Spumavirus	Homo sapiens Bos taurus	NC_001795 NC_001831	2111–2750 2016–2652	3487-4204	8788–9399 8903–9515	2,111 2,016	868 835	4,456 4,699	2,555 2,487
Mean \pm SD							$1,\!373\pm564$	868 ± 217	3,851 ± 799	$1{,}410\pm962$

TABLE 1. Previously characterized retroviruses from a variety of species used as a query

most biologically interesting due to the unique physiology of its digestive, reproductive, and immune systems. The unveiling of the cattle genome sequence in 2009 allowed the first comprehensive effort to catalogue the diversity of transposable elements in the cattle genome (5). Interspersed repeats cover 46.54% of the genome. Among these, non-LTR retrotransposon LINEs account for 23.29% of the genome, and SINEs account for 17.66% of it. LTR retrotransposons, which include ERVs, account for 3.20% of the genome (5).

The cattle genome has also been analyzed experimentally for ERV elements. A PCR-based approach (43) detected a number of bovine ERVs (BERVs [here BoERVs]), which were classified into four families, named β 3, γ 4, γ 7, and γ 9, on the basis of their similarity to ovine ERVs (OERVs). The structures and sequences of BERV- β 3 and the abundant BERV- γ 4 elements were also analyzed. Those studies suggested that the expansion of the ERV family was more limited in cows than it was in other artiodactyls such as pigs and sheep (42–44).

To detect ERVs in the cow genome, we used three different methods: BLAST-based searches using retroviral sequences, the LTR_STRUC program, and the Retrotector program. ERVs that were detected by at least two of the methods and whose reverse transcriptase (RT) region was longer than 500 nucleotides were used to define bovine ERV families. Finally, representatives from each bovine ERV family and from ERV families from other species were used to study the relationship between the ERVs of different species.

MATERIALS AND METHODS

Genomic sequence. We analyzed the Btau_3.1 version of a Hereford cow (*Bos taurus*) genome (×7.1 coverage). It was retrieved from the Baylor College of Medicine (http://www.hgsc.bcm.tmc.edu/projects/bovine/index.html).

Detection of BoERVs. Three different strategies for the detection of ERVs were applied and compared. The first strategy was based on the similarity of sequences. Segments of amino acid sequences from gag and env (TM region) genes from 12 well-annotated exogenous retroviruses were used as a search query (Table 1). In the case of the pol gene (RT region), one piece for each retroviral class was used (from Moloney murine leukemia virus [MLV], Mason-Pfizer monkey virus [MPMV], and bovine foamy virus). In each chromosome, individual gene segments were searched by using tBLASTn implemented in the NCBI-BLAST 2.2.14 program (1). The results of this search were parsed by

homemade scripts written in PHP. To build the possible ERV elements, the results for each gene were compared. Based on the 12 search query retroviruses, the distances between gene queries were calculated (Table 1). A region was considered to be a possible ERV if the distance between the *gag* and *pol* genes was 868 ± 217 bp, the distance between *pol* and *env* genes was $3,851 \pm 799$ bp, and the distance between the *gag* and *env* genes was $4,719 \pm 1,016$ bp. If at least two matches were within the limits of these distances, the region was defined as a possible ERV.

In the second strategy, LTR_STRUC 1.1 (22) was used to find LTR elements. LTR_STRUC scans the genomic sequence for the presence of similar regions of length typical for LTRs (LTR pairs) and within the expected size of full-length LTR ERVs. If putative LTRs are found, the program then searches for additional retrotransposon features, such as primer binding sites (PBSs), polypurine tracts (PPTs), and target site repeats (TSRs), and assigns a reliability score to the hit based on the presence or absence of each of these features. In this work the predefined parameters were used.

We also used Retrotector v.1.0 (38) to find possible ERV elements. Briefly, Retrotector recognizes consensus motifs and constructs putative ERV proteins ("puteins") from the different reading frames in the gene candidates. The program uses codon statistics, frequency of stop codons, and alignment to known retrovirus proteins to approximate an original open reading frame (ORF). The predefined parameters were used, and the cutoff score value was set at 250.

Nomenclature. We have used the name bovine endogenous retrovirus (BoERV) for the ERVs described in this work in order to avoid confusion with previous names (43).

Distribution of BoERVs. The distribution of the detected elements was tested chromosome by chromosome, as proposed previously by Villesen et al. (41). The expected number of elements (based on chromosomal mean density and length) was compared with the observed number by means of the χ^2 test and the G test (37), each with 1 degree of freedom: $\chi^2 = \Sigma$ [(observed – expected)²/expected] and G = 2 Σ observed × ln(observed/expected).

Correlation analyses between the number of elements detected by each method and chromosome length, GC content, and gene density were performed by using R language (32).

Classification of BoERVs. Relationships between the detected elements were determined by a phylogenetic analysis of the RT region. The 247 sequences with a high degree of similarity in BLAST-based searches (RT region of >500 nucleotides) were used along with the 12 exogenous retroviruses used as a search query, 8 detected cow ERVs (BoERVs) (43), and 12 sheep ERVs (17). Retroviruses used in previous research, such as BoEV (GenBank accession number X99924), HERV-E (accession number M10976), porcine ERV (PERV) (accession number AJ293656), MPMV (accession number NC_001550), intracisternal A particle (IAPM) (accession number M17551), ovine maedi visna virus (OMVV) (accession number NC_001511), feline foamy virus (FeFV) (accession number U78765), murine ERV-like (MuERV-L) (accession number Y12713), and *Drosophila* endogenous element ZAM (accession number AJ000387) were used as an outgroup. The sequences were aligned by using the MAFFT 5.861 program (14) (FFT-NS-1 option) and cleaned

with Gblocks 0.91 (6) (minimum length of block of 5, allowed gap position with half, minimum number of sequences for a flank position of 146, and maximum number of contiguous nonconserved positions of 10). Phylogenetic trees were built by using three different methods. One was the neighbor-joining (NJ) method implemented in MEGA 3.1 (19). It relies on *p* distance using the pairwise deletion option and 1,000 bootstrap replicates. A second method was the maximum likelihood (ML) method implemented in Phyml 2.4.4 (10). The model used in the analysis was the GTR+G model ($\alpha = 2.71$), as estimated by Modeltest 3.7 (31) with 1,000 bootstrap replicates. The third method was the Bayesian inference method implemented in MrBayes 3.1 (35). Four default-setting Metropolis-coupled Markov chain Monte Carlo methods were performed in two runs for 10⁶ generations with trees sampled every 100 generations. The analysis was set to use the GTR+I+G model. The first 2,500 trees were discarded in the burn-in, and a 50% majority-rule consensus tree was computed from the remaining trees.

The bovine ERV putative families that we detected were defined based on the support of phylogenetic trees. A cluster was considered a putative family when the clustering was significant in at least two of the phylogenetic methods (boot-strap values of >70 for neighbor joining and maximum likelihood and Bayesian posterior probability of >95 for Bayesian inference). In order to confirm the families with a solitary member, MegaBlast (45) searches were carried out by using the solitary sequences as a query.

To elucidate the insertion time of elements classified into families, we used the divergence of LTRs estimated by LTR_STRUC and Retrotector. The dates of ERV insertion can be estimated mainly by the LTR comparison and the individual divergence relative to a consensus sequence. However, there are many difficulties in obtaining an accurate consensus sequence, especially for short insertions, so, as was done by many other authors, the LTR comparison was chosen for estimates of insertion dates. These divergence figures were then corrected to account for the presence of multiple mutations at the same site, back mutations, and convergent substitutions by using the Kimura two-parameter model (16). Nine elements with highly divergent LTRs were not included in this analysis. The insertion time of each element was estimated by applying a substitution rate of 2.3 $\times 10^{-9}$ to 5.0 $\times 10^{-9}$ to the divergence (12).

The representative members of a BoERV family were chosen as the closest element to the consensus sequence of the family or, when consensus was not possible, to the element with fewer stop codons. To build the consensus sequence, the amino acid sequences of the members of each family were aligned by using ClustalW (39), and consensus was determined by using the cons program of the EMBOSS suite (33). The distances between individual sequences and the consensus were calculated by using MEGA 3.1 (number of differences and pairwise deletion options).

PCR amplification of BoERV1 family elements in sheep. In order to amplify BoERV1 in sheep, PCR was performed with the primers 5'-TGTGCTGAGAC AGAGGAAGC-3' (forward) and 5'-CCTATGGCCCTAGTCCCTTC-3' (reverse) in six samples of Latxa breed sheep. PCR conditions consisted of 5 min at 94°C followed by 25 cycles of a 55°C annealing step for 30 s, polymerization at 72°C for 30 s, denaturation at 94°C for 30 s, and one final cycle at 72°C for 7 min. Reaction conditions were as follows: 13.1 μ l of water, 2 μ l of buffer, 0.8 μ l of MgCl₂, 0.3 μ l of deoxynucleoside triphosphate (dNTP), 0.3 μ l of each primer, 0.2 μ l of TAQ, and 3 μ l of DNA.

Interspecies comparison. The relationships of the BoERV groups with other species' ERV groups were analyzed by using a phylogenetic tree. We used 16 representative sequences from human ERV families (40), 38 from chimpanzees (29), 27 from mice (2, 23), 7 from rats (2), 14 from sheep (17), 8 from pigs (27), and 24 from cows (our data). The sequences were aligned by using MAFFT (14) (linsi option). Positions with gaps in more than half of the sequences were eliminated. The tree was constructed by Bayesian inference implemented in MrBayes 3.1 (35) (10⁶ generations; RtREV matrix+G+I).

To estimate the insertion time of the human HERV-S71 element and the chimpanzee CERV3 element, we used the LTR divergence as described previously by Johnson and Coffin (12). The LTRs of each element were aligned with ClustalW, and the distance was calculated by using MEGA 3.1 (K2P substitution model). The insertion time was estimated by applying a substitution rate of 2.3×10^{-9} to 5.0×10^{-9} (12).

RESULTS

Detection and distribution of ERVs in the cow genome. In the cow genome, the BLAST-based search detected 928 ERVs, LTR_STRUC identified 4,487 elements flanked with LTRs, and Retrotector detected 9,698 possible ERVs (Table 2). Only

172 elements were detected by all three methods (Fig. 1). A total of 739 of the elements detected by the BLAST-based search were also detected by Retrotector. Retrotector identified 8,183 elements that were not detected by either the BLAST-based search or LTR_STRUC (Fig. 1). The elements detected by the three methods were those that were best preserved from a structural standpoint; e.g., in 81% of the elements detected, Retrotector detected motifs from the NC region of the *gag* gene, and in 83% of the elements, it detected motifs from the RT region of the *pol* gene. However, the elements detected by a single program and LTR_STRUC and Retrotector together did not provide evidence for these motifs in most cases (35% and 54%, respectively).

The gene compositions of the detected elements also differed from method to method (Table 3). The elements detected by Retrotector included the three major genes. Most of the BLAST-based search elements included two genes. Surprisingly, most of the elements detected by LTR_STRUC did not include the *pol* gene. Based on the Retrotector results, 78 regions encompassed either a *gag* or an *env* ORF longer than 500 codons or a *pol* ORF longer than 700 codons (which approached the size of intact viral proteins).

The distribution of the BoERVs was not concordant across the three methods (Table 2). In the BLAST-based search, significantly more elements were detected on chromosomes 18, 28, and X than would be expected for a homogeneous distribution, while significantly fewer elements were detected on chromosomes 14 and 20. The number of elements that LTR STRUC detected in chromosomes 1, 2, and 3 was significantly higher than would be expected for a homogeneous distribution, while significantly fewer elements were detected in chromosomes 17, 18, 19, and 22. Retrotector identified significantly more elements than expected in chromosomes 1, 2, 3, 6, 9, and X and fewer elements than expected in chromosomes 13, 18, 19, 22, 23, 24, and 25. In these analyses, the significance levels were similar by the χ^2 and G tests. Using the G test, the overall distribution of BoERVs in the entire genome was not significantly homogeneous for the three methods.

The number of ERVs detected was strongly and positively correlated with the chromosome size by all three methods (Spearman's $\rho = 0.7677$ and P < 0.001 by BLAST, $\rho = 0.9720$ and P < 0.001 by LTR_STRUC, and $\rho = 0.9735$ and P < 0.001 by Retrotector) and inversely correlated with the GC content ($\rho = -0.4968$ and P < 0.001 by BLAST, $\rho = -0.7066$ and P < 0.001 by LTR_STRUC, and $\rho = -0.6979$ and P < 0.001 by Retrotector). No significant correlation was observed between the number of ERVs detected and chromosomal gene density.

We conducted a chromosome-by-chromosome search for solo LTRs based on Retrotector results (data not shown). Based on this analysis, the average rate of solo LTRs/ERVs was 6.06.

Classification of BoERVs. A phylogenetic tree based on the well-conserved *pol* gene RT region of selected BoERVs (detected with at least two methods and having an RT region with >500 nucleotides) with other endogenous and exogenous retroviruses showed that most of the elements were related to class I or class II outgroup elements. Thus, they can be classified as such by homology. No class III-related elements were observed (Fig. 2). Based on this tree, the BoERV elements were classified into 24 families (BoERV1 to BoERV24) ac-

	Analvzed			BLAST					LTR_STRUC					Retrotector		
Chromosome	length (Mb)	No. of ERVs	Density (elements/Mb)	χ^2 test P value ^a	G test P value	Result	No. of ERVs	Density (elements/Mb)	χ^2 test P value ^a	G-test P value	Result	No. of ERVs	Density (elements/Mb)	χ^2 test P value ^a	G test P value	Result
_	147	58	0.39	0.7533	0.7548		310	2.11	0.0108*	0.0126*	→	727	4.95	6.3E-10***	2.44E-09***	→
2	126	43	0.34	0.4790	0.4718		263	2.09	0.0303*	0.0339^{*}	→-	554	4.40	0.0117*	0.0131*	→-
ယ	117	50	0.43	0.3849	0.3940		250	2.14	0.0128*	0.0152*	→ -	540	4.62	0.0002^{***}	0.0003^{***}	<i>→</i> -
4	11	46	0.41	0 5361	0 5419		201	1.81	0 8601	0 8 5 9 9	_	458	4 13	7275 0	0 3758	-
J.	119	42	0.35	0.6349	0.6310		214	1.80	0.7749	0.7743		497	4.18	0.2273	0.2312	
	117	4 4 7 0	0.32	0.0049	0.2150		221 112	2.06	0.0661	0.0714		CC7	1 75		2 00U 05***	>
0	1112	49	0.44	0.3040	0.3152		231	2.06	0.0661	0.0/14		332	4.75	1.8E-05	3.09E-05	
7	101	35	0.35	0.5874	0.5820		193	1.91	0.5544	0.5571		406	4.02	0.7640	0.7646	
8	104	33	0.32	0.2957	0.2824		199	1.91	0.5352	0.5380		428	4.12	0.4206	0.4235	
9	96	34	0.35	0.6858	0.6825		201	2.09	0.0541	0.0595		424	4.42	0.0222*	0.0247*	\rightarrow
10	96	30	0.31	0.2797	0.2653		194	2.02	0.1653	0.1720		370	3.85	0.5891	0.5878	-
11	102	32	0.31	0.2734	0.2594		187	1.83	0.9975	0.9975		375	3.68	0.1394	0.1348	
12	78	37	0.47	0.1648	0.1809		164	2.10	0.0738	0.0805		334	4.28	0.1484	0.1536	
13	84	22	0.26	0.0759	0.0604		133	1.58	0.0855	0.0785		283	3 37	0.0054**	0.0044**	
14	8	17	0.20	0.0087**	0.0041**	<u> </u>	152	1.83	0.9912	0.9912		$\frac{1}{324}$	3.90	0.7871	0.7866	~
15	76	24	0.32	0.3626	0.3488	•	154	2.03	0.2058	0.2133		326	4.29	0.1445	0.1498	
16	73	25	0.34	0.6059	0.6000		116	1.59	0.1181	0.1099		271	3.71	0.2773	0.2723	
17	71	23	0.32	0.4438	0.4324		108	1.52	0.0489*	0.0425*	←	268	3.77	0.4218	0.4182	
18	63	36	0.57	0.0120	0.0192^{*}	\rightarrow	91	1.44	0.0210*	0.0166*	<u> </u>	191	3.03	0.0001^{***}	9.00E-05***	~
19	64	27	0.42	0.5732	0.5798		83	1.30	0.0013^{**}	0.0007^{***}	←	174	2.72	4.1E-07***	8.28E-08***	←
20	69	16	0.23	0.0439^{*}	0.0301^{*}	~	126	1.83	0.9660	0.9660		254	3.68	0.2351	0.2296	
21	64	25	0.39	0.8792	0.8798		107	1.67	0.3348	0.3277		240	3.75	0.3887	0.3845	
22	60	15	0.25	0.1001	0.0798		70	1.17	0.0001^{***}	3.60E-05***	←	171	2.85	1.2E-05***	4.07E-06***	~
23	49	20	0.41	0.7384	0.7414		82	1.67	0.4049	0.3980		83	1.69	7.9E-16***	$1.06E - 19^{***}$	<u> </u>
24	61	16	0.26	0.1335	0.1124		111	1.82	0.9382	0.9382		199	3.26	0.0054^{**}	0.0041^{**}	<u>،</u>
25	43	20	0.47	0.3553	0.3718		67	1.56	0.1793	0.1682		118	2.74	5.2E-05***	$1.85E - 05^{***}$	<u>،</u>
26	48	22	0.46	0.3678	0.3830		74	1.54	0.1322	0.1218		168	3.50	0.1046	0.0978	
27	4	14	0.32	0.5079	0.4959		08	1.82	0.9418	0.9418		152	3.45	0.0881	0.0813	
28	41	27	0.66	0.0034^{**}	0.0079^{**}	→	66	1.61	0.2871	0.2771		165	4.02	0.8385	0.8389	
29	46	00	0 43	0 5357	0 5449	-	69	150	0 0000	0 0877		166	3 61	0 2247	0 2178	
X	100	70	0.70	1.03E-07***	1.74E-06***	→	191	1.91	0.5610	0.5636		500	5.00	1E-07***	2.96E-07***	→
	TOO	ò	0.70	1.0010 07	1.7412 000	_	171	1.71	0.010	0.0000		000	5.00		1.70L 07	_
Overall	2,448	928	0.38		3.683E-05***		4,487	1.83		3.40E-07***		869,6	3.96		1.36E-47***	

Vol. 84, 2010

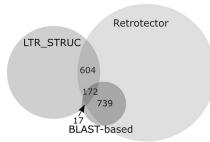


FIG. 1. Diagram representing the number of ERVs detected by each detection method (represented by the circle size) and common elements detected by two or more methods (included in the overlapping areas).

cording to the tree topology and the statistical support of the clustering relationships (Fig. 2 and Table 4). Overall, we defined 18 families related to class I ERVs and 6 related to class II ERVs. Some of these groups were also related to ovine ERVs. It proved possible to classify all 24 families into three groups based on the number of ERVs included (Table 4). BoERV1 (82 clustered elements), BoERV3 (66 elements), and BoERV18 (16 elements) in class I and BoERV24 (22 clustered elements) in class II were the most abundant families. A second group of families had between 4 and 10 elements. In the third group, eight elements were isolated and had no significant relationship with any other element.

Class I families (BoERV1 to BoERV18). Most of the BoERVs included in the phylogenetic analysis were related to class I elements and were classified into 18 families. The average length of class I-related families was between 8,209 bases (BoERV18 family) and 15,219 bases (BoERV16). The longest ERV was from the BoERV16 family (23,004 bases), due to a duplication of the *pol* and *env* retroviral genes. The shortest ERV was from BoERV12 (7,058). Apart from the examples described above, the duplication of *gag* and *env* genes was observed for other ERVs. To rule out the presence of other genes, a search was made for ORFs in the longest elements. No different *gag*, *pol*, or *env* ORFs were detected. Thus, it is very likely that the presence of these elements was due to assembly errors. The remaining ERVs showed a typical length of between 7 and 12 kb (4).

Among the most abundant families, the BoERV1 family was quite heterogeneous. In contrast, the elements of the BoERV3 and BoERV18 families were similarly homogeneous. The BoERV3 family was related to previously described cow BERV- γ 4 and sheep OERV- γ 4 elements, and BoERV18 was related to the ovine OERV- γ 9 element.

The representative element of the BoERV1 and BoERV3 families had the LPQG and YVDD motifs (Fig. 3). In the case of the BoERV18 family, the first motif was PPQG.

The remaining groups were represented by few or solitary elements. In the cases of BoERV2 and BoERV13, more related sequences were obtained in the MegaBlast analysis. They had not been included in the phylogenetic tree because they did not fulfill the conditions previously established.

The divergence level varied from family to family: the elements of BoERV8, BoERV9, and BoERV10 were less divergent than the elements of BoERV5, BoERV7, and BoERV15. BoERV7 was related to the previously detected cow BERV- γ 7 and ovine OERV- γ 7 elements, and BoERV16 was related to a previously detected BoERV (GenBank accession number X99924) and BERV- γ 9 bovine elements. In BoERV12 and BoERV15, the two motifs were conserved. In the representative element of BoERV9, the YVDD motif was present. However, the representative elements of the BoERV2, BoERV4, BoERV5, BoERV6, BoERV10, BoERV11, BoERV16, and BoERV17 groups did not keep at least one of the two characteristic motifs.

The representative BoERVs from families with one or two members had different conservation levels of their functional motifs. BoERV2 had a deletion in the YVDD motif, and BoERV4, BoERV13, BoERV14, and BoERV17 had an amino acid change. Neither of the motifs was conserved in BoERV6 and BoERV11.

BoERV1 could be the oldest of the class I-related families because it contained an ERV that was inserted between 126 and 58 million years ago (MYA) based on the LTR divergence. This family also had the most recent insertion activity, since the youngest member was inserted recently. In sheep samples, we were able to amplify a BoERV1 conserved sequence from the RT region with a length of 150 bases (data not shown). The BoERV7 family could be the youngest, inserted between 19 and 9 MYA. Due to the uncertainty of the age estimates of the ERV sequences, which were based on the comparison of the LTRs of the elements, these values are only a rough estimate of the insertion time. Although different evolution rates and a correction were applied, the effect of recombination or gene conversion events, leading to the homogenization of the 5' and 3' LTRs, must be taken into account, and in this sense, the divergence times calculated in our study may have been underestimated.

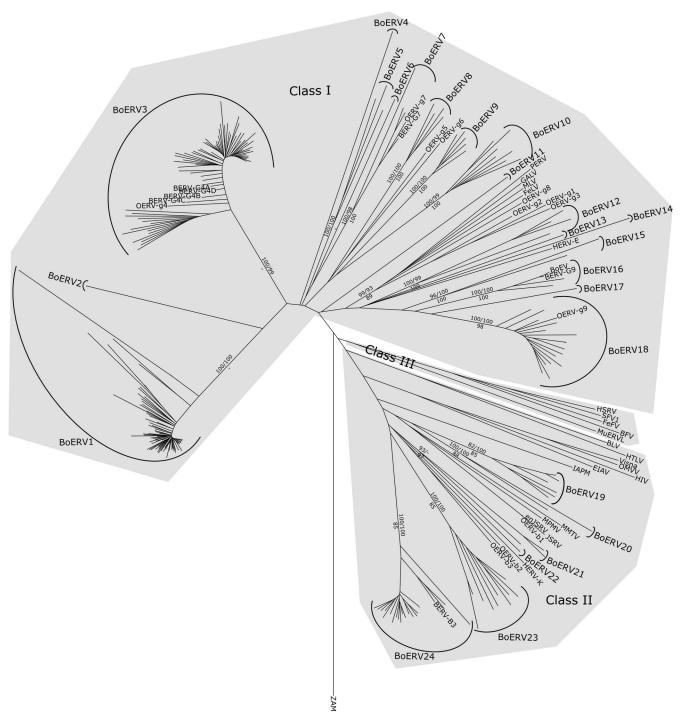
Class II families (BoERV19 to BoERV24). The class II elements were grouped into six families, ranging in average length from 8,881 bases to 11,077 bases. After the MegaBlast analysis, BoERV22 consisted of more than one element. Some of these elements had not been included in the phylogenetic tree because they did not fulfill the conditions previously established.

The longest ERV was from BoERV24 (25,030 bases), and the shortest was from BoERV21 (6,283 bases). As with the longest class I-related BoERV, the longest BoERV in the class II-related families contained all genes duplicated.

TABLE 3. Structure of ERV elements detected by each method

		No. (%) of dete	cted elements	а
Structure	Retrotector	LTR_STRUC	BLAST	Nonredundant elements
LTR-RT-LTR gag-pol-env gag-pol gag-env pol-env gag pol env	5,006 (51.61)7,466 (76.99)2,002 (20.64)91 (0.94)99 (1.02)18 (0.19)22 (0.23)0	383 (8.54)	NA 162 (17.46) 267 (28.77) 190 (20.47) 309 (33.30)	5,254 (38.57)
Pol presence Pol absence	9,589 (98.87) 109 (1.13)	383 (8.54) 4,104 (91.46)	738 (79.53) 190 (20.47)	
Total	9,698	4,487	928	13,622

^a NA, not applicable.



^{0.05}

FIG. 2. RT region-based phylogenetic tree of BoERVs. A total of 247 BoERVs detected by at least two methods and with the *pol* gene longer than 500 nucleotides from this work were included. Eight experimentally detected cow ERVs (43) and 12 sheep ERVs (17) were also included. Retroviruses used as queries and retroviruses used in previous phylogenetic studies, such as gibbon ape leukemia virus (GALV) (GenBank accession number NC_001585), MLV (accession number NC_001501), feline leukemia virus (FeLV) (accession number NC_001940), Jaagsiekte sheep retrovirus (JSRV) (accession number NC_001494), mouse mammary tumor virus (MMTV) (accession number NC_001503), bovine leukemia virus (BLV) (accession number NC_001414), human T-lymphotropic virus (HTLV) (accession number NC_001436), equine infectious anemia virus (ELAV) (accession number NC_001450), HIV (accession number NC_001803), Visna virus (accession number NC_001452), human spumaretrovirus (HSRV) (accession number NC_001795), bovine coronavirus (BCV) (accession number NC_001831), BoEV (accession number X99924), HERV-E (accession number M10976), PERV (accession number U78765), and MuERV-L (accession number Y12713), were used as an outgroup. Topology was based on the neighbor-joining method with a *p* distance of 1,000 bootstrap. The tree was rooted with the *Drosophila melanogaster* ZAM (accession number AJ00387) element. Above the branches, the NJ bootstrap values and ML bootstrap values are shown; below the branches, the Bayesian posterior probability is shown.

Family	No. of elements	Lei	ngth (bases)		Representative element ^a		ion time YA) ^b	Nomenclature ^c	Ovine ERV orthologue ^a
2	detected	Median	Range	PBS ^e	Chromosome:start-end (strand)	Oldest	Youngest		orthologue
Class I									
BoERV1	82	10,313	7,333-16,256	Pro	2:69350995-69361048 (-)	126-58	1-0		
BoERV2	1	10,688			4:94489122-94499810 (+)	ND			
BoERV3	66	9,776	7,437-15,164	His	7:55058524-55069351 (-)	106-49	7–3	BERV-γ4	OERV-γ4
BoERV4	1	9,917			24:39795592-39805509 (+)	59-27			
BoERV5	3	9,859	7,900-10,850		18:12944241–12952141 (+)	75-35	74–34		
BoERV6	1	10,869			3:10016186-10027055 (+)	22-10			
BoERV7	3	10,671	10,193-11,172	Tyr	19:36194623-36204816 (+)	19–9	2-1	BERV-γ7	OERV-γ7
BoERV8	4	10,077	8,583-11,079	-	5:43721672-43731333 (+)	81-37	52-24		
BoERV9	5	10,764	10,023-11,254	Phe	5:42333382-42344624 (+)	62-28	32-15		
BoERV10	7	10,456	7,509–12,994	Ser	3:97148459-97158193 (-)	38-18			
BoERV11	1	10,947			2:123894988-123905935 (+)	ND			
BoERV12	4	9,949	7,058-11,208	Ser	9:6107380-6115865 (-)	64-29	42-19		
BoERV13	1	11,087			14:26972828-26983915 (-)	ND			
BoERV14	1	9,859			23:27514449–27524308 (+)	66-30			
BoERV15	2	10,052	9,908-10,196	Pro	23:27654628-27664824 (+)	35-16			
BoERV16	3	15,219	11,234-23,004	Pro	13:77909792-77919633 (-)	40-18		BERV- _y 9	
BoERV17	1	11,106		Tyr	18:49897632-49908738 (+)	ND			
BoERV18	16	8,209	7,428–10,311	Pro	9:9401110-9409616 (+)	64–30	9–4		OERV-79
Class II									
BoERV19	7	10,354	8,670-13,275	His	4:33488089-33501364 (+)	27-13	0		
BoERV20	2	9,587	9,531-9,643	Met	10:23209873 - 23219516(-)	44-20	11-5		
BoERV21	3	8,881	6,283-10,345	Lys	2:19509989-19520334 (+)	34-16	16-7		
BoERV22	1	11,077	, , , ,	2	18:42572717-42583794 (-)	37-17			
BoERV23	10	10,085	7,876-11,307		7:6319253–6327129 ()	50-23	30-14		
BoERV24	22	10,703	8,801-25,030	Lys	4:69860298-69869823 (-)	56-26	11–5	BERV-β3	

^a The families were defined with BoERVs detected by at least two methods and with a *pol* gene longer than 500 nucleotides. In the case of families with one member only, one insertion time and the median of the length are shown.

^bND, not determined.

^c Nomenclature according to Xiao et al. (42-44).

^d According to Klymiuk et al. (17).

^e The PBS could not be predicted for some representative elements.

Within-group divergence varied. The elements of BoERV23 were more divergent than those of BoERV19. The BoERV24 group contained some tightly related elements and some highly divergent ones. The BoERV24 family was related to the previously detected bovine BERV- β 3 element. Moreover, we did not detect class II bovine ERVs related to endogenous Jaagsiekte sheep retroviruses (enJSRVs) in this genome version.

With one exception, the representative elements of the class II-related families conserved the LPQG and YMDD motifs. In the BoERV23 family, the LPQG motif was replaced by QPQG, and YMDD was replaced by YLDG.

The oldest class II-related family could be BoERV24, whose oldest ERV was inserted between 56 and 26 MYA. The newest family could be BoERV19 (between 27 and 13 MYA). In addition, a member of this family could be the youngest element, with a recent insertion.

Relationship of BoERVs with ERVs from other species. In the phylogenetic tree for the ERVs from cow, sheep, pig, human, chimpanzee, mouse, and rat, ERV elements were grouped into three classes. Within each class, the representative ERVs clustered following the relationships with their host genomes. There was a close relationship between ERVs from cow and sheep, which clustered together in four phylogenetic lineages (Fig. 4).

The representatives of the different classes were grouped into polytomic nodes. However, there were clear relationships between human/chimpanzee families (HERV-I/-ADP and CERV20/21/22/23/24/25), human/chimpanzee/bovine families (ERV9/HERV-W, CERV15/16/17/18/29, and BoERV1/2), chimpanzee/bovine families (CERV19 and BoERV11), and human/chimpanzee/mouse families (HERV-L, CERV42, and MuERV-L/Mmr20). There were also suggestions of bovine/ ovine relationships (BoERV15/16/17/18 and OERV-G9) and human/chimpanzee/porcine/bovine relationships (RRHERV-I/HERV-E, CERV4/5/6/7, PERV-g4, and BoERV12/13/14) (Fig. 4).

Surprisingly, one phylogenetic lineage contained elements from human, chimpanzee, mouse, pig, and sheep species but not from cows. This lineage was studied in depth, and the insertion time was estimated by using LTR divergence: the human HERV-S71 element was inserted between 19.5 and 8.9 MYA, and the chimpanzee CERV3 element was inserted between 33 and 15.8 MYA.

DISCUSSION

This study describes an attempt to systematically identify and characterize endogenous retroviruses in the cow genome. Although we used only located genomic information, leaving contigs untested, in this study we identified nearly 10,000 putative BoERVs that were distributed in a nonhomogeneous way across chromosomes. By comparing three different meth-

	10	00	20		<u> </u>	70		100	440	100	400	4.40
	10	20	30 40		60 I I I	70	80 90		110	120	130	140 l
BoERV1	QYPLKPEVKEGLKPI I E											
BoERV2	QYPLKLEA- QGFI PLI K	NLKEQDFFLI B	ECNSPONTSVLGV	OKQ- NGEWSI VODL	CLINEAVIPLHP	VPKPYALLSQI F	PEMADWFTVLDLK	DAFFCI PLAQSSOF	LFAFE DPSD	SATOLTWIT	LPQG FF	RDTPH
BoERV3	QYPLKREAREGLQPLI N											
BoERV4 BoERV5	- WERLKAGREGDDRG- D TRPPKVESRKGLESLVD											
BOERV5 BOERV6	NI LLREEVLI GSMPLI S											
BoERV7	QYPLKPEALEGLRPLVN											
BoERV8	TTVLSLPKTQQ_KPI VT											
BoERV9	QYSVTPEAHKGLKPI I D	HLLQAS-ILIF										
BoERV10 BoERV11	QPPLKREVKEAI KPM N							SLLFLHHSSOP				
BoERV12	- YPVP- GVHLGI WDHI (
BoERV13	QYLTP- EAQLGTQAHLD											
BoERV14	- FTGAFQGGPSGTI DGA											
BoERV15	QYQLPLEARI GI FPHI N											
BoERV16 BoERV17	FWPLEE-MPQGKEGEGS											
BoERV17	QYPLPI EAWAGI LPHI N											
BoERV19	QWPLTKEKLLAAKALI S											
BoERV20	AMAI NS- KVTGS- GFSY	GTTGSR- PYR/	AF PLNTPI FVI I	KKT-SGKWRLLODL	RAI NPTMEDMGAL	OPG - LPSPVAN	/PSQYSVI VTGLC	DCFFAI PLADQDYK	RFGFSLPSANF	Kapyrrfarm	LPQG MA	KNSPT
BoERV21	QWSLSHEKLQALEQLVE											
BoERV22 BoERV23	Q-PLKQ-KLEDLHELIN - WPLTKEKLQAAKELID											
BoERV23	QWPLTKEKLAALTQLVN											
	150	160	170 18		200		220 230		250	260		
BoFRV1		· . [<u>]</u>	<u></u>				1	.				
BoERV1 BoERV2	LFGQSLSRDLQNFNSS- LFGQALAKDLTAFWHQ	- EAVVLOYVDI - SL- KLSKM 1	DILLCAETEEAC-	- SRASEDFLNFLAG	 Cgykasrekaqlo Wilqniki - SSDN	XQQSVRYLGLIIS A QEVKCQGLKFS	. SegtrAigpe ShgtqAlgke	 RI KPI LNHPLPMTL RI KPI L- HPLPSAI	RQLRGFLG	- I TGYCRI W P - I TDYC- I CI P		
BoERV2 BoERV3	LFGQSLSRDLQNFNSS- LFGQALAKDLTAFWHQ- LFGQALGDLLDLDLGP-	- EAVVLOYVDI - SL- KLSKM-1 - NGKI LOYVDI	D LLCAETEEAC MAFSFAVTOKNS- DLLI CSPNEKSA-	- SRASEDFLNFLAG - VTKLPVI FLLCSF - QEHAI QVLNFLAE	 CGYKASREKAQLO WILQNI KI - SSDA RGYKVSCAKAQM		. Segtr Al Gpe Shgtq Al Gke Thgsr Rlssd	 RI KPI LNHPLPMTL RI KPI L- HPLPSAI RVQRI LQLPSPMTR	RQLRGFLG	- TGYCRI W P - I TDYC- I CI P - I TDYC- I CI P - LTGYCRI W P		
BoERV2 BoERV3 BoERV4	LFGQSLSRDLQNFNSS- LFGQALAKDLTAFWHQ- LFGQALGDLDLDLGP- LFG-ALAQDLQALTLEG	- Eavvloyvdi - Sl- Klskm M - Ngki Loyvdi - GTVI Lovdi	D LLCAETEEAC MAFSFAVTCKNS- DLLI CSPNEKSA- NLLI CYPSKI QAN	- SRASEDFLNFLAG - VTKLPVI FLLCSF - QEHAI QVLNFLAE - LHAI QVLNFLAG	 CGYKASREKAQLO WILQNIKI-SSDM RGYKVSCAKAQM WDYRVQVQ		. Segtr Al GPE Shgtq Al GRE Thgsr RLSSD TPGRH SLSVD	 Rikpilnhplpmtl Rikpil-hplpsai Rvqrilqlpspmtr Pieaifdpqtptth	RQLHGFLG	- I Tgycri Wip - I Tgycri Wip - I Tdyc- I Cip - Ltgycri Wip - Ltgtvgfr. P		
BoERV2 BoERV3 BoERV4 BoERV5	LFGQSLSRDLQNFNSS- LFGQALAKDLTAFWHQ LFGQALGDLLDLDLQP- LFG ALAQDLQALTLEC MFTQSLGKELRETQLQ	- Eavvl (avvd) - SL- Klskm M - Ngki Lgyvd) - GTVI Lgvd YTVSI LTMKEL	D LLCAETEEAC MAFSFAVTCKNS- DLLI CSPNEKSA- NLLI CYPSKI QAN LFYGM-I TH-HAG	- SRASEDFLNFLAG - VTKLPVI FLLCSP - QEHAI QVLNFLAE - LHAI QVLNFLAG LRSEPTEI LHFLGA	 CGYKASREKAQLO WILQNI KI - SSDN RGYKVSCAKAQM WDYRVQVQ - GYRLSQRKSTI H		. Segtr Al Gpe Shgtq Al Gke Thgsr Rlssd Tpgr Slsvd /ds Yld	 Ri kpi l.nhplpmtl Ri kpi l hplpsai Rvqri l.qlpspmtr Pi Eai Fdpqtptth Renkllqlsapktk	RQLRGFLG	- I Tgycri W P - I Tdyc- I Ci P - Ltgycri W P - Ltgycri W P - Ltgtvgfr- P - Ragfcri - FP		
BoERV2 BoERV3 BoERV4 BoERV5 BoERV6	LFGQSLSRDLQNFNSS- LFGQALAKDLTAFWHQ- LFGQALGDLLDLDLGP- LFG-ALAGDLQALTLEG MFTQSLGKELRETQLQ- LFS-ALAKDLSSLTLEG	- EAVVLOTVDI - SL- KLSKM I - NGKI LOTVDI - GTVI LOVDI YTVSI LTMKEL - ESCLLOTFEI	D LLCAETEEAC MAFSFAVTCKNS- DLI CSPNEKSA- NLI CYPSKI CAN LFYGM I TH- HAG DLI CSPDI LTS-	- SRASEDFLNFLAG - VTKLPVI FLLCSF - OEHAI QVLNFLAE - LHAI QVLNFLAG LRSEPTEI LHFLGA - DNNTI LVLKHLAE	 CGYKASREKACL (WILQNI KI - SSDI RGYKVSCAKACIM WDYRVQVQ - GYRLSCRKSTI H RGYWSPAKACI T		. Segtr Al Gpe Shgtq Al Gke Thgsr Rlssd Tpgr Slsvd /ds Yld Tpgkr Sl Apd	RI KPI LNHPLPMTL RI KPI L-HPLPSAI RVQRI LQLPSPMTR PI EAI FDPQTPTTH RENKLLQLSAPKTK RRKLI SDMAMPATK	RQLRGFLG RQLHGFLG KQLRAFLG KQLQAFLG	- I TGYCRI W P - I TDYC- I CI P - LTGYCRI W P - LTGYCRI W P - LTGTVGFR- P - RAGFCRI - FP - MINFCQI LDP		
BoERV2 BoERV3 BoERV4 BoERV5	LFGQSLSRDLQNFNSS- LFGQALAKDLTAFWHQ LFGQALGDLLDLDLQP- LFG ALAQDLQALTLEC MFTQSLGKELRETQLQ	- EAVVLOYVDI - SL- KLSKM I - NGKI LOYVDI - GTVI LOVDI YTVSI LTMKEL - SSOLLOYFEI - NSNLI QYVDI	LLCAETEEAC- MAFSFAVTCKNS- DLI CSPNEKSA- NLI CYPSKI QAN LYGM-I TH-HAG DLI CSPDI LTS DLI ASPDFTSS-	J J SRASEDFLNFLAG VTKLPVI FLLCSF OEHAI QVLNFLAE - LHAI QVLNFLAG RSEPTEI LHFLGA ONNTI LVLKHLAE QRDTVKTLNFLYE			. Segtr Al GPE Shgtq Al GPE Shgtq Al GRE Those Al SSD TPGRH SLSVD /DS YLD TPGKR SI APD /EGKR MLDPG	Ri KPI LNHPLPMTL Ri KPI L-HPLPSAI RVQRI LQLPSPMTR PI EAI FDPQTPTTH RRENKLLGLSAPKTK RRKLI SDMANPATK RKSLI LNTPYPQTK	RQLRGFLG RQLHGFLG KQLRAFLG KQLQAFLG KHLCAFLG	- I TGYCRI W P - I TDYC- I CI P - LTGYCRI W P - LTGYCRI W P - LTGTVGFR- P - RAGFCRI - FP - MINFCQI LDP - I TGFCRI W P		
BoERV2 BoERV3 BoERV4 BoERV5 BoERV6 BoERV7 BoERV8 BoERV9	LFGQLSDLTFFWD. LFGQLADLTFFWD. LFGQLQDLDLDLGP. LFGQLQDLQLDLDLGP. LFGALGQLQLLDLDLGP. LFGALGQLQLCDLGP. LFS.ALAKDLSSLTLEC LFGSLGFELRELTLK- FFGQALGX-DLGTDLT- FFGQALGX-DLGTDLT-	- EAVVLQYDD - SL- KLSKM N - NGKI LQYDD - GTVI LQYDD - GTVI LQYDD - SSLLQYFEI - NSNLI QYDD - LSTVLQYDD APSHLLQYI D	DILLCAETEEAC MAFSFAVTCINS ULI CSPNEKSA- ULI CYPSKI CAN FYGM I TH-HAG DLI CSPDI LTS- DLI ASPDFTS- DLI ASPDFTS- DLLCSPSFI HS- DLLCSPT- KLC	SRASEDFLNFLAG VTKLPVI FLLCSF QEHAI QVLNFLAG LRSEPTEI LHFLGA DNNTI LVLKHLAE QRHI I QLLNFLAG LQHTPKLLGALGS	UU.K. CGYKASREKALL WWLCNIKI - SCA KGYKVSCAKACIM WDYRVQVQ - GYRLSQRKSTIH KGYRVSPKKACI - GYRVSPKKACI - GYRVSPKKACI WGYRVSQSKACIA		. Egtr Al Gre HGSR ALGKE THGSR SLSVD /DSYLD /PGRH SLSVD /DSYLD PGRH SLAPD /EGKR MLDPC /EGKR YTTTD SHQQR TI LSD	RI KPI LNHPLPMTL RI KPI L-HPLPSAI RVORI LCLPSPMTR PI EAI FDPQTPTTH RENKLLGLSAPKTK RRKLI SDMANPATK RKSLI LNTPYPQTK RKPLI STLPI PTSK RI QALI NOPLPKTK	RQL RGFLG	- I TGYCRI WI P - I TGYCRI WI P - LTGYCRI WI P - LTGYCRR P - RAGFORI - FP - MINFCOI LDP - I TGFORI WI P - LAGYLSLWI P - LLNFFRI - I P		
BoERV2 BoERV3 BoERV4 BoERV5 BoERV6 BoERV7 BoERV8 BoERV9 BoERV10	LFGQALSDLDLDLG- LFGQALGDLDLDLG- LFGQALGDLDLDLG- LFGALGDLQLDLG- LFG-ALGOLQALTEG HFGSLGKELRETGLQ- HFGSLGKELRETGLG- HFGSLGKELRETLK- FFGQALSDLTSPDLT- FFGQALGK-DLGTLDL FFGQALGK-DLGTLDL	- EAVVLQYVDI - SL- KLSKM-I - NGKI LQYVDI - GTVI LCVDI YTVSI LTIMEL - ESCLLQYFEI - NSNLI QYVDI - LSTVLQYVDI - LSTVLQYI DI LPYNM-I NI DI	DILLCSPSTHC	SRASEDFLNFLAC VTKLPVI FLLCSF OEHAI QMLNFLAE - LHAI QMLNFLAE INSEPTEI LHFLQA DNNTI LVKHLAE ORDTVKTLNFLYE CQHI I QLLNFLAE LQHTFKLLGALGS LVHTAMLNALGA	CGYKASREKACL WLQNIKI-SSDM RGYKVSCAKACM WDYRVQ/Q GYRLSQRKSTIF GYRVSPKKACI KGYRVSPKKACI - GYRVSPTKVCLS WGYRVSCSKACI WGYRVSLSKTCI /		I I I. EGTR AI GPE HGSR RLSSD IPGRH SLSVD /DS SLSVD /PGKR SI APD /EGKR SI APD /EGKR SI APD /EGKR SI APD /EGKR SI APD /FIASK I I SAC	R KPI L.NHPLPMTL RVQRI L.QLPSPMTR RVQRI L.QLPSPMTR RPI EAI FDPOTPTTH RENKLLQLSAPKTKK RRKLI SDMANPATK RRKLI STLPI PTSK RRVLI STLPI PTSK RQ QALI NOTPRPOTK RL QALTOTPRPOTK	RQL RGFLG RQL HGFLG KQL RAFLG KHLCAFLG KHLCAFLG KQL RFLG TEI LSFLG RELLS LG RELLSLLG	- I TGYORI WI P - I TGYORI WI P - LTGYORI WI P - LTGYORI WI P - RAGFORI - FP - MINFCOL LDP - I TGFORI WI P - LAGFULSUWI P - LLNFFRI - I P - LLNFFRI WIP		
BoERV2 BoERV3 BoERV4 BoERV5 BoERV6 BoERV7 BoERV8 BoERV9 BoERV10 BoERV11	LFGQLASDLASDLAST LFGQLASDLAFFWD, LFGQLASDLAFFWD, LFGQLGDLDLDLGP, LFGALGQLQLDLGALTLEC MFTQSLGKELRETQLO, LFS-ALAKDLSSLTLEC IFGTSLGRELRELTLK- FFGQALGKDLOTLD FSAQPSKETYKL-T-AL IFVEAFLKDLKDLCLN-	- EAVVLCYVDC - SL- KLSKM I - NGKI LCYVDC - GTVI LCVD - GTVI LCVD - ULCYVDC - ULCYVDC - ULCYVDC - ULCYVDC - LSTVLCYVDC - LSTVLCYVDC - LCYVM I IN DC - EGI LLCYVC	DI LLCAETEEAC- VAFSFAVTCKNS- ULI CSPNEKSA- ULI CYPSKI CAN EYGM I TH- HAG DLI CSPDI LTS- DLI ASPDFTSS- ELLCSPSFI HS- DLLCSSSHNC- NLI ASKTKEAS-	SRASEDFLNFLAG SRASEDFLNFLAG OEHAI QALNFLAG INSEPTEI LICSP OEHAI QALNFLAG DINTI LVLKHLAG QRDTVKTLNFLAT CGHI I QLLNFLAT L GHTPKLLGALGS L VHTAMALNALGN DONTVRTLNFLAT	CGYKASREKAQL CGYKASREKAQL RGYKVSCAKAQM WDYRVQ/Q - GYRLSQRKSTI RGYWSPAKAQ KGYRVSPKKAQ VGYRVSPKKAQ WGYRVSLSKTQ RGYKVSLSKTQ RGYKVSLEKTQ	A QEVKOQQLKFS A QEVKOQQLKFS CETKVTYLGVQIT - QPSCLRTTMI LNI WDI L- ALED: S- QVTFLQLI LT LTQVKYLGFI I M SLPRVTHLRVI LL QTTVKYMQLLLT QPTVKYLGVELS	I I I. EGTR AI GPE SHGTQ ALGKE IPGRH SLSVD JDS YLD IPGKR SI APD GCKR M.DPC .PI KR YTTTD SHQCR TI LSD SHQCR NLLPD SKGQR NLLPD	R KPI L NH-PLPMIL R KPI L H-PLPSAI RVORI L CL PSPMIR RVORI L CL SPMIR REXKL L CL SPMIR RRXLI J DMMPATK RRXLI J DMMPATK RRXLI LNTPYPOTK RI CALI NCPLPTTK RL CALI NCPLPTTK RECAI ACMNPTTK	ROL ROFLG ROL HOFLG KOL RAFLG KOL RAFLG KOL QAFLG COLRI FLG COLRI FLG TEI LSFLG RELLSI LG RELLSI LG KOL ROFLG	- I TGYCRI WI P - I TDYC-I CI P - LTGYCRI WI P - LTGTVGFR-P - RAGFCRI - FP - MTNFCQI LDP - I TGFCRI WI P - LAGYLSLWI P - LLNFFRI - I P - LLNFFRI WIP - LLNFFRI WIP		
BoERV2 BoERV3 BoERV4 BoERV5 BoERV6 BoERV7 BoERV8 BoERV9 BoERV10 BoERV11 BoERV12	LFGQSLSRDLONFNSS- LFGQALGALDLARDLGAFUFGALAADLARDUG LFGQALGALDLDLGP- LFGALGADLQALTLEG MFTQSLGKELRETLG- LFGSLGRELRELTLK- FFGQALGSDLSSLTLEG FFGQALGSDLSSDLT- FFGQALGSDLSSDLT- FFGQALGSDLSSLTLGL FSGQESKETYKL-T-AL I FVEAFLKDLKDLQJN RFGGALGANLAAFPGGT	- EAWLOWDD - SL-KLSKM - NGH LOWDD - GTVI LCVD YTVSI LTIMEL - SSLLQYFEL - NSNLI QWDD - LSTVLQYVD APSHLLQYI DI LPYNM I NI DI - EGI LLQYVD FNCTLLQYMD	DILLCAETEERA WAFSFANTOKNS- DLII CSPNEKSA- LLI CYPSKI CAN FYGMI 1TH HAG DLII CSPDI LTS- DLII ASPDFTSS- DLILCSPT- KLC- DLLCSSSFINC- DLLCSSSFINC- NLI ASKTKEAS- DLLASPTC/NDC-	SRASEDFLNFLAC VTKLPVI ELLOSP OEHAI QVLNFLAC INSEPTEI LIFLCS INSEPTEI LIFFLCA DNNTI LVLKHLAE QRDTVKTLNFLV2 CQHI I QLLNFLAD LQHTFKLLGALGS LVHTAMVLNALGA DQNTVRTLNFLAD WRGTI ALLALLST	CGYKASREKACL CCGYKASREKACL WLON KI SSDN WDYRVQ/Q	A CEVKOURYLCLIIS A CEVKOZALKES FETKVTYLCVCIT OPSCLRTTM LN WDIL- ALEX S- OVTFLCLILT LTOMKYLCFIIN LTOMKYLCFIIN SUPRVTHLRVILL SGTVTYLCLSIS STTVMYMCLLIS GPTVKYLCFVIS	. EGTR AI GPE FIGTQ AL GRE FIGSR RLSSD (PGRH SLSVD (PGRR SLSVD (PGRR SLAPD (FGRR SLAPD (FGRR TI LSD (FAGR TI LSD (FAGR ALCHE SKGGR ALCHE	RI KPI LNHPLPMITL RI KPI L-HPLPSAI RVGRI LQLPSPMITR PI EAI FDPOTPTTH RENKLL GLSAPKTK RRKLL SDAMPATK RRKLL STLPI PTSK RRKLI STLPI PTSK RRKLI STLPI PTSK RRKLI STLPI PTSK RRKAI CSI PVPTTK RRAAI AGM/PTTK	RCL RGFLG RCL HGFLG KCL CAFLG KAL CAFLG CALRI FLG CALRI FLG TEI LSFLG TEI LSI LG RELLSI LG KEVCEFLG KEVCEFLG	I TOYCE I CI P I TOYCE I CI P LTGYCRI W P LTGTVGFR P RACFCRI - FP MTNFCOI LDP I TGFCRI W P LAGYLSLW P LLNFFRI - I P LLNFFRI W P KAEFCRI W P SAGCOW W P		
BoERV2 BoERV3 BoERV4 BoERV5 BoERV6 BoERV7 BoERV8 BoERV9 BoERV10 BoERV11	LFGQLASDLASDLAST LFGQLASDLAFFWD, LFGQLASDLAFFWD, LFGQLGDLDLDLGP, LFGALGQLQLDLGALTLEC MFTQSLGKELRETQLO, LFS-ALAKDLSSLTLEC IFGTSLGRELRELTLK- FFGQALGKDLOTLD FSAQPSKETYKL-T-AL IFVEAFLKDLKDLCLN-	- EAWL OVDO - EAWL OVDO - SL- KLSKM I - NGKI LGVDI - STV LCVD - SSL LTMEL - SSLL OVFEI - NSNLI OVDI - LSTVLOVDO - LSTVLOVDO - LSTVLOVDO LPVNM I NI DI - EGI LLOVIO J.SCALLOVID J.SCALLOVID	DILLCAPTERA MARSFANTO(NS- DLLICSPNEKSA- LLICSPNEKSA- LLICSPNEKSA- LLICSPTILTS- LLICSPTILTS- LLICSPTILS- LLICSSTINC LLICSSSINCO LLICSSSINCO LLICSSTORDO	SRASEDFLNFLAG VTKLPVI FLCSP OEHAI OXLNFLAG RSEPTEI LHFLCS ORDTVKTLNFLAG ORDTVKTLNFLAG OQHI I QLLNFLAG LQHTFKLLGALGS VRGTI ALLALLSLLME	CCSYKASREKACL CCSYKASREKACL CCSYKASREKACL RCSYKVSCAKACH WDYRVQ/Q GYRLSGRKSTI H RCSYWSPAKACL KCSYKVSKKACL KCSYKVSKKACL RCSYKVSKKACL L GYKVSKKKACL CSYKVSKKKACL CSYKVSKKKACL	ACTION AND A CONTRACT	II.A.GPE SEGTR AI GPE HGSR RLSSC IPGRH SLSVD US YLD PGRH SLSVD US YLD PGRR SLAPD SHOLR YTTTD SHOLR TI LSD SKGRR NLLPD SKGHR ALGHE	RI KPI LNHPLPMIL RI KPI L-HPLPSAI RVGRI LQ.PSPMIR PI EAI FDPOIPTTH RRVKLI G.SAPKTK RRVLI SDMANPATK RKPLI STLPI PTSK RKPLI STLPI PTSK RR QALI NCPRPOTK RREAI AGMNPTTK RRCAI CSI PVPNTK RKQAI CSI PVPNTK	RCLRGFLG	I TGYCR W P I TGYCR W P L TGYCR W P L TGYCR W P L TGYCR OR - FP RAGFOR - FP MTNFCOLLDP I TGFCR W P L LGYTR - I P LLNFFR W P KAEFCR W P SAGCOW W P		
BoERV2 BoERV3 BoERV4 BoERV5 BoERV5 BoERV7 BoERV7 BoERV10 BoERV11 BoERV12 BoERV13 BoERV13	LFGQALGALDLAKDLTAFWHO LFGQALGALDLAKDLTAFWHO LFGQALGALDLAGLUG LFG-ALAGALQALTLEG HFG-ALAGALQALTLEG HFG-ALAGALQALTLEG HFG-ALAGALQALTLEG FFG-ALAGALGALGAL FFGQALAGALGALGAL FFGQALAGALGALGAL FFGQALGALGALGALGALGALGALGALGALGALGALGA FGGALGALGALGALGALGALGALGALGALGALGALGALGAL	- EAVVLQVDD - EXVLQVDD - GTVLCVDD - GTVLCVDD - SCLLQYFEL - NSNLQVDD - LSTVLQVDD - LSTVLQVDD - LSTVLQVDD - LSTVLQVDD - EGLLQVVD - EGLLQVVD - SCQLLQVVD - SCQLLQVVD - SCQLLQVVD - SCQLLQVVD - SCQLLQVVD - SCQLLQVVD - SCQLLQVVD	D LLCAETEERC MAFSFAVTO(NS- DLLI CSPNEKSA- LLI CSPNEKSA- LLI CSPNEKSA DLLI CSPNI LTS- DLLI CSPDI LTS- DLLI CSPDI TSS- DLLI CSPSFI HS- DLLI CSPSFI HS- DLLI CSPSFI HS- DLLI CSPSFI HS- DLLI CSPSFI HS- DLLI CSPTOMOC- DLLI ASKTKEAS- DLLI ASKTKEAS- DLLI ASKTRACO	SRASEDFLNFLAG VTKLPVI FLCSF CEHAI QXLNFLAG LRSEPTEI LHFLQA DNNTL LVLKHLAG QRDTVKTLNFLXF QRDTVKTLNFLXF QRDTVKTLNFLXF QRDTVKTLNFLXF QRDTVLLALLST LGHTFKLLQALGS LVHTAMLNALQL DQRTFKLLQLLEG WRGTIALLSLLM WEGTKALLSLM	I	A GEVKCOLLIS A GEVKCOLLKFS (ETK/TYLGVC) (ETK/TYLGVC) (ETK/TYLGVC) (ETK/TYLG) (ETK/TY	EGR AI GPE SEGR AI GPE FHGR AL GPE FHGR RLSSC /DS SLSVD /DS SLSVD /DS SLAPC /DS SLAPC /DS SLAPC /DS TLLSD /HOR TLLSD /HOR ALG-E /HEG ALG-E	RI KPI LINHPLPMIL RI KPI L-HPLPSAI RVGRI LQ.PSPMIR RVGRI LQ.SPMIR RENKLL GLSAPKTK RRKLI SDMMPATK RRKLI SJDMPPTIK RRKLI SJDPPTIK RRAL SJDPPTIK RRAL AGMNPTIK RRAL AGMNPTIK RRAL AGMNPTIK RRAL AGMNPTIK RRAL RAL RPPKT-	RCLRGFLG	I TGYCRI W P I TDYC I CI P L TGYCRI W P L TGYVGR. P RAGFCRI - FP MTNFCCI LDP I TGFCRI W P L LAGTXLSLW P LLNFFRI - I P LLNFFRI WP KAEFCRI W P SAGCOW W P - SAGCOW W P		
BoERV2 BoERV3 BoERV4 BoERV5 BoERV6 BoERV7 BoERV10 BoERV11 BoERV12 BoERV13 BoERV14 BoERV14	LFGQALGXLTAFWHOL LFGQALGXLTAFWHOL LFGQALGXLTAFWHOL LFGQALGXLDLDLDG- LFG ALAQDLQALTLEC MFTQSLGKELRETQLQ LFS ALAKDLSSLTLEC IFGTSLGRELRETLKL FFGQALGK-DLQTDL FFGQALGK-DLQTDL FFGQALGK-DLQTDL FFGQALGK-DLQTDL FFGQALGK-DLQTDL FFGQALGKDQCQ IFGEALSDLAKFPQGT IFGEALASDLDSFQPG FPSLNQAG IEARAGU	- EAWLQVDD - SL-KLSKM - NGKI LQVDD - GTVI LCVD YTVSI LTIMEL - SESLLQYEE - NSNL QVDD - LSTVLQVDD - LSTVLQVD - LSTVLQVD - EGI LLQVD - EGI LLQVD - EGI LLQVD - EGI LLQVD - SCALLQVND NSCALQVND - SCALQVND - SCALQVND	DILLCAPTERA WAFSFANTGKNS- DLLI CSPNEKSA- LLI CSPNEKSA- CLI CSPNEKSA- LLI CSPNI LTS- LLI CSPDI LTS- LLI CSPT LTS- LLI CSPT-KLC- LLI CSPT-KLC- LLI CSST-KLC- LLI CSST-KLC- LLI CSST-KLC- LLI CSST-KLC- LLI CSTTOACC- LLI CSTTOACC- CST	SRASEDFLNFLAG SRASEDFLNFLAG CEHAI QALNFLAG SEPTEI LHFLQS CEHAI QALNFLAG CRSEPTEI LHFLQA ONDTI LVLKHLAG QRDTVKTLNFLAG QCHI QLLNFLAG LQHTPKLLGALGS UVHTAMLNALG DQNTVRTLNFLAG UVHTAMLALSLLM VRGTKALLEPLME ASQCGRTGLQAR	CSYKASREKACI CSYKASREKACI NDYRVQVQ	ACTION AC	EGR ALGE FIG2 - ALGE FIG2 - ALGE FIG2 - ALGE FIG2 - ALGE SC SLSVC CS YLC FIG2 - SLSVC CS YLC FIG2 - SLSVC FIG2 - SLSVC FIG2 NLPC SC ALGE SC	RKPIL-HPLPMIL RKPIL-HPLPSAI RVCRILGLPSPMIR RVCRILGLSPMIR RRKLILGSAPKTK RRKLIJSDMANPATK RKSLILNTPYPOTK RKSLILNTPYPOTK RKSLILNTPYPOTK RRKLAICSIPVPNTK RRKAI AGVNYTTK RRKAI CSIPVPNTK RRKAICSIPVPNTK RRKAILCSIPVPNTK RRKAILR P-PKT- KVCTCLDLKDVFG	RCLRGFLG RCLHGFLG KOLRAFLG KOLQAFLG CALQAFLG CALQAFLG CALQAFLG CALQAFLG CALQAFLG CALQAFLG RCLSLLG KOLKFFLG CALKFFLG CALKFFLG RCVREF-G RCVREF-G RCVREF-G RCVREFCG RCVR	I TGYCR W P I TGYCR W P L TGYCR W P L TGYCR W P L TGYCR W P I TGYCR W P I TGYCR W P L LAGYLSLW P LLNFFR I W SAGCOW W P SAGCOW W P ATGFCR W P		
BoERV2 BoERV3 BoERV5 BoERV5 BoERV7 BoERV7 BoERV70 BoERV10 BoERV11 BoERV12 BoERV14 BoERV15 BoERV15 BoERV17	LFGQALGALDLAKDLTAFVHO- LFGQALGALDLAKDLTAFVHO- LFGQALGALDLDLGP- LFGALGADLQALTLEG MFTQSLGKELREITLK- FFGQALGSLTSTLEG IFGTSLGRELREITLK- FFGQALGSDLTSPDLT- FFGQALGX-DLGTLD FSGQESKETYKL-T-AL IFVGALKALLAKPPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT	- EAWLOYDD - SL- KLSKM I - NGKI LGYDD 3 - GTVI LGYD YTVSI LTIKEL 3 - SGULQYEL - NSNLI GYDD - LSTVLGYDD - LSTVLGYDD LLPYNM I NI D - EG LLQYYD XSCALLGYND XSCALLGYND SFRWLLQYDD FFGWLLQYDD FFGWLLQYDD	D LLCAETEEAC WRFSFAVTO(NS- D.LI CSPNEKSA- LLI CSPSKI QAN LI CSPSKI QAN LI CSPSKI QAN LI SSPDITSS- ELLCSPSFI HS- D.LLCSPSFI HS- D.LLCSSSSHNC- D.LLASTTQACC- D.LLASTTQACC- D.LLAASTTQACC- D.LLAASTTCACC- D.LLAASTTCACC- D.LLAASTCACC- D.LLAASTCACC- D.LLAAENKEHC- S.VEQC/QLEVAN	SRASEDFLNFLAG VTKLPVI FLLCSP CEHAI QXLNFLAG LRSEPTEI LHFLG DNNTI LVLKHLAE COTIVKTLNFLXE COTIVKTLNFLXE COTIVKTLALGALGS LVHTAM/LNALGAL DONTVRTLNFLAE LOTIFKLLSLLW WRGTI ALLALLST LEGTKALLSLLM VEGTKALLEPLME PASQKGRRTQLQAF	I	A GEVKOLGLIS A GEVKOLGLKFS FETK/TYLOVA TETK/TYLOVA TETK/TYLOVA S OVTFLGLIS S OVTFLGLIS S OVTFLGLIS S TVNYMGLLL GTIV/TYLGSIS STTVNYMGLLL GTIV/TYLGFUS S TVNYMGLLCFVI GTEVKYLGFVLS KREVRYLGFVLF KEEVRYLGFVLF	EGRR AI GPE SHETC AI GPE SHETC AI GPE THOSR RLSC JDS SLSVD JDS SLSVD JDS SLSVD JDS SLSVD JDS SLSVD JDS SLSVD JPACR TLLSD SHOCR TLLSD SHOCR TLLSD SKGRR LLDOS SLLCL R A	RI KPI LNHPLPMILL RI KPI L-HPLPSAI RVGR LGLPSPMIR RVGR LGLSPMIR RRVGL SDMMPATK RRVGLI SDMMPATK RRVGLI SDMMPATK RRVGLI SDMPATK RRVGLI SDMPATK RRVGLI SDMPATK RRVGLI SDMPATK RRVGLI SDMPATK RRVGLI SI PVPNTK RRVGLI SI PVPNTK RRVGLI GI PVPNTK RRVGLI CI PVPKT KVGTCI DLKDVFG SMPLSVMOG LPV	RQL RGFLG RGL HGFLG KQL RAFLG KQL QAFLG KHL CAFLG CAL CAFLG CAL CAFLG CAL CAFLG CAL CAFLG RGL SLLG RGL SLLG RGL SLLG RGL SLLG RGL SLLG RGL RFLG RGUREFLG RGU	I TGYCRI W P I TDYC-I CI P L TGYCRI W P L TGYVGR-P RAGFCRI - FP MTNFCCI LDP I TGFCRI W P L AGYLSLW P LLNFFRI WP KAEFCRI W P LLNFFRI WP KAEFCRI W P ATGFCRI W P ATGFCRI W P		
BoERV2 BoERV3 BoERV4 BoERV5 BoERV6 BoERV7 BoERV10 BoERV11 BoERV12 BoERV13 BoERV14 BoERV14	LFGQALGALDLARD LFGQALARDLTAFVHO LFGQALGALDLAG LFGQALGALDLDLGP- LFGALGGELRETOLO LFGSLGGELRETOLO LFGSLGGELRETOLO LFGSLGGELRETILK- FFGQALGX-DLGTDL FFGQALGX-DLGTDL FFGQALGX-DLGTDL LFSGDLARALARFPGGT LFSGDLARALARFPGGT LFSGDLARALARFPGGT IFGGALTTELLSFP-ST IFGGALGSDLDSFOPGG PPSLNQAG I EARAG L IFGCALSDLDSFOPGG PFGLALSDLNSFHPE FGGALSDLSFOPGG	- EAWLQVDD - EAWLQVDD - SL-KLSKM - NGKI LQVDD - GTVI LQVD - GTVI LQVD - GTVI LQVD - SSLI QVD - LSTVI QVDD - LSTVI QVDD - LSTVI QVDD - LSTVI QVDD - EGI LQVVG LQVGU LQVDD LSCALLQVVI AKCO LQVVE - FRVILQVD - PH NRLKAGS (FGWLLQVDD	D LLCAETEERC MAFSFANTQ(NS- DLLI CSPNEKSA- LLI CSPNEKSA- LLI CSPNEKSA- DLLI CSPDT LTS- DLLI CSPT ITH-HAG DLLI CSPTFIS- DLLICSPT-FIS- DLLICSPT-FIS- DLLICSPT-FIC- DLLICSSSHMC- DLLICSSSHMC- DLLICASTTQACC- DLLICASTTQACC- DLLICAETKERC- SLICAETKERC-	SRASEDFLNFLAG VTKLPVI FLCSF CEHAI QMLNFLAG CREVENI QMLNFLAG CREVENI LHFLQA CREVENI LHFLQA CREVENI LHFLQA CREVENI LHFLQA CONTINUTLNFLAG CONTINTLALALST UNTTAMLNALQA VRGTI ALLALLST LEGTKALLSLLME VRGTIALLEPLME PASQKGRRTQLQAF VRGTIALLEPLME VRGTIALLEPLME		ACCONTRACTOR ACCONT ACCONTRACTOR ACCONT ACCONTRACTOR ACCONT ACCONTRACTOR ACCONT ACCONTRACTOR ACCONTRACTOR ACCONTRACTOR ACCONTRACTON	EGR AI GPE SEGR AI GPE HGSR AL GRE HGSR RLSSC /DS RLSSC /DS RLSSC /DS YLD PGR SLAPC /DS YLD SHOR SLAPC SHOR TI LSC /SGR AL GPE HGEG AL GRE LLDOS SLL CLR A (RGTR LLOSS	R KPI LNHPLPMIL R KPI L-HPLPSAI RVCRI LQ.PSPMIR RVCRI LQ.SAPKTK RRXLI LSDMANPATK RRXLI LSDMANPATK RXSLI LNTPPOTK RXSLI LNTPPOTK RXAI CSI PPOTK RXAI CSI PPOTK NGI PLQY RXAI CSI PPOTK KVCTOL DLKDVFC SAMPLSVMG COPT RXAVFTQL DLKDVFC SAMPLSVMG CM SAMPLSVMG CM SAMPLSVMG CM CM CSAMPLSVMG CM CM CM CM CM CM CM CM CM CM	RCL RGFLG	I TGYCR W P I TDYC I C P L TGYCR W P L TGYVGR P RAGCRI - FP MTNFCCI LDP I TGYCR W P L LAGYLSLW P LLNFFR W KAEFCRI W P SAGCOW W P SAGCOW W P SMCFCW W P DP- MCFFC DMP- MCFFC		
BoERV2 BoERV3 BoERV4 BoERV5 BoERV7 BoERV7 BoERV10 BoERV10 BoERV12 BoERV12 BoERV13 BoERV14 BoERV16 BoERV16 BoERV18	LFGQALGALDLAKDLTAFVHO- LFGQALGALDLAKDLTAFVHO- LFGQALGALDLDLGP- LFGALGADLQALTLEG MFTQSLGKELREITLK- FFGQALGSLTSTLEG IFGTSLGRELREITLK- FFGQALGSDLTSPDLT- FFGQALGX-DLGTLD FSGQESKETYKL-T-AL IFVGALKALLAKPPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT IFGGALGAALAAFPGCT	- EAWLQWDQ - EAWLQWDQ - SL-KLSKM - NGKI LQWDQ - GTVI LCVD - TIMKEL - SSLLQYFEL - NSNLI QWDQ - NSNLI QWDQ - NSNLI QWDQ - EGI LLQWQ - EGI LLQWQ - EGI LLQWQ - SCALLQWDQ - SCALQWQ - SCALQWQ - PH NRLKQAG FGQWLQWDQ - FQQWLLQWDQ - FQQWLQWDQ - FQQWLQWDQ - FQQWLQWDQ - FQQWLQWDQ	DILLCAPTERA WAFSFANTG(NS- NLI) CSPNEKSA- LLI) CSPNEKSA- LLI) CSPNEKSA- LLI) CSPNI LTS- LLI) CSPNI LTS- CSPNI	SRASEDFLNFLAG SRASEDFLNFLAG OTKLPVI FLCSP OEHAI OMLNFLAG INSEPTEI LHFLGA ORDTVKTLNFLAG ORDTVKTLNFLAG ORDTVKTLNFLAG UNTFMMLNALGA UNTFMMLNALGA UNTFMMLSLLM WEGTKALLGLLEG WEGTKALLGLLED WEGTKALLGLLM OTVLODLTGALTC	CSYKASREKACI CSYKASREKACI NUCANI KI - SSDI NUCANI KI - SSDI NUCANI KI - SSDI NUCANI - SSDI NUCANI - SSDI NUCANI - SCH ROYKYSCA KOYKYSKKACI NUCANI - SST ROYKYSKKACI NUCANI - SST ROYKYSKKACI NUCANI - SST ROYKYSKKACI NUCANI - SST NUCANI - SST	ACKING III ACKING IIII ACKING IIII ACKING IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	L	IRI KPI L-IN-PLPMIL RI KPI L-IN-PLPSAI RVCRI LGLPSPMIR RVCRI LGLSSPMIR RRVKLI LGLSSPMIR RRKLI JDMANPATK RRKLI JDMANPATK RRKLI SDMANPATK RRKLI SDMANPATK RRKAI AGMNPTTK RRKAI CSI PVPNIK RRKAI CSI PVPNIK RRKAI CSI PVPNIK RRVENI CSI PVPNIK RRVENI LRI P-PKT- RKOVFNCI CCPT K SAMPLSVMCG LPW I LMALAS- NKONCO	RCLRGFLG RCLHGFLG KCLQAFLG KCLQAFLG KCLQAFLG CLQAFLG CLQAFLG CLQAFLG CLQAFLG CLQAFLG CLQAFLG RCLSLLG RCLKFLG	I TGYCR W P I TGYCR W P L TGYCR W P L TGYCR W P L TGYCR W P I TGYCR W P I TGYCR W P LAGYLSLW P LAGYLSLW P LAGYLSLW P LAGYLSLW P SAGCOW W P SAGCOW W P SAGCOW W P ATGFCR W P CMP- MGFFC HOD-KG D I WW RPYLK		
BoERV2 BoERV3 BoERV4 BoERV5 BoERV7 BoERV7 BoERV10 BoERV10 BoERV12 BoERV12 BoERV13 BoERV14 BoERV15 BoERV16 BoERV17 BoERV18 BoERV19 BoERV21	LFGQLADLDLOLG- LFGQLADLTAFWHO- LFGQLGDLDLDLG- LFGALGGLERETOLO- LFG-ALAODLADLDLG- LFG-ALAODLADLTEC MTOSLGGELRETOLO- LFGSLGFELRETTLK- FFGQALASDLTSPDLT- FFGQALASDLTSPDLT- FFGQALASDLTSPDLT- FFGQALASDLASDLGLAN RFGEALSALAAFPGET LFGGLASDLSFPF- ST IFGEALADLSFCPGE PPSLNQAG I EARAGI LFGQLASDLNSFHPEE LOCKFVNAAI EDI RAKY LCONFVOALONARTKY LCONFVOALONARTKY LCONFVOALONARTKY	- EAWLQYOD - EAWLQYOL - SL- KLSKM I - NGKI LQYOD - SL- KLSKM I YTVSI LITIKEL - SSLLQYFE - NSNL QYOD - LSTVLQYOD - LSTVLQYOD - LSTVLQYOD - LSTVLQYOD - EGLLQYVQ - SCALLQYND - SCALLQYND - SCALLQYND - PH NRLQYOD - PH NRLQYOD - PHAYI LYND - FQCWLQYOD - PHAYI LYND - FQCWLQYOD - PHAYI LYND - FQCWLQYOD - PHAYI LYND - FQCWLQYOD - FQCWLQY	DILLASPTENDO LLISSPIERSA- LLICSPIERSA- LLICSPIERSA- LLICSPIERSA- LLICSPIERSA- LLICSPIERSA- LLICSPIERS- LLICSSSINCO- LLICSSSINCO- LLIASPTONOCO- LLIASPTONOCO- LLIASPTONOCO- LLIASPTONOCO- LLIASPTONOCO- LLIASPTONOCO- LLIAAENKEHC- S. VECOVOLEYAN D.LLIAAENKEHC- D.LLIAAENKEHC- D.LLIAAENKEHC- D.LLIAAENKEHC- D.LLIAAENKEHC- D.LLIAAENKEHC- D.LLIAAENKEHC- D.LLIAAENKEHC- D.LLIAAENKEHC- D.LLIAAENKEHC- D.LLIAAENKEHC- D.LLIAAENKEHC- D.LLIAAENKEHC- D.LLIAAENKEHC-	SRASEDFLNFLAG SRASEDFLNFLAG - LHAI QALNFLAG - SEPTEI LHFLGA CBHAI QALNFLAG - SEPTEI LHFLGA ONTVI LVLKHLAG - CADTVKTLNFLAG - CADTVKTLNFLAG - CADTVKTLNFLAG - DONTVRTLNFLAG - DONTVRTLNFLAG - DONTVRTLNFLAG - DONTVRTLNFLAG - DONTVRTLLGLEG - WEGTKALLSLIME - WEGTKALLSLIME - WEGTKALLSLIME - OTVLODLTQALTE - COTVLODLTQALTE - COTVLODLTQALTE - COTVLODLTQALTE	Cosykasrekaci Cosykasrekaci NUCANI KI - SSDI WUCANI KI - SSDI WUTKVQVQ Cyrll, SGRKSTI H RGYWSPAKACI KGYRVSPKKACI GYRVSPTKVCL GYRVSSKKACI GYRVSSKKACI GYRVSVKKACI GYRVSVKKACI GYRVSVKKACI GYRVSVKKACI GYRVSVKKACI GYRVSVKKACI GYRVSVKKACI GYRVSVKKACI GXGRVSKKKACI CAGSTVSKKKKACI CAGSTVSKKKKACI CAGSTVSKKKKACI CAGSTVSKKKKACI CAGSTVSKKKKACI CAGSTVSKKKKKACI CAGSTVSKKKKACI CAGSTVSKKKKACI CAGSTVSKKKKACI CAGSTVSKKKKACI CAGSTVSKKKKACI CAGSTVSKKKKACI CAGSTVSKKKKACI CAGSTVSKKKKACI CAGSTVSKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK	ACCONTRACTOR ACCONT ACCONTRACTOR ACCONT ACCONTRACTOR ACCONT ACCONTRACTOR ACCONT ACCONTRACTOR ACCONTRACTOR ACCONTRACTOR ACCONTRACTON	L, J, JPE BEGR AI GPE FHG2 AL GRE FHG2 AL GRE FHG2 RLSS/C DS YLC FPGR SLSVC DS YLC FPGR SLAPC FHG2 TLSS FHG2	IN I	RCLRGFLG RCLHGFLG KCLRAFLG KCLQAFLG CALQAFLG CALQAFLG CALQAFLG CALQAFLG CALQAFLG CALQAFLG RCLSLLG RCLSLG RCLKFLG RCKEF	I TGYCR W P I TDYC I C P L TGYCR W P L TGYCR W P L TGYCFR P RAGCCR - FP MTNFCCI LDW L LNFFR I W LLNFFR W W KAEFCR W P SAGCOW W P ATGFCR W P ATGFCR W P ATGFCR W P ATGFCR W P DMP-MCFFC DMP		
BoERV2 BoERV2 BoERV3 BoERV4 BoERV6 BoERV7 BoERV10 BoERV11 BoERV12 BoERV13 BoERV16 BoERV17 BoERV18 BoERV17 BoERV18 BoERV20 BoERV20	LFGQALGALDLQCLQF LFGQALGALDLAGLGF LFGQALGALDLQG- LFGQALGALDLQG- LFGALGADLQAUTLEG MFTQSLGKELRETCLQ- LFGSLGKELRETCLQ- LFGSLGKELRETCLK- FFGQALGSDLTSPDLT- FFGQALGSDLTSPDLT- FFGQALGK-DLGTLDL FSGQLGADLGKETYKL-T-AL IFVGALKALAGPPGGT IFGGALTSDLSFPGG JFGGALGADLAKFPGQT IFGGALTSDLNSFHPEE LCQKFVMAW ED RAYG LCQNFVDQALGNARTKY LCQNFVDQALGNARTKY LCQNFVDQALGNARTKY		D LL CAETEEAC WRFSFAVTO(NS- D LL CSPNEKSA- LL CSPSKI CAN FYGM I TH- HAG D LL CSPSKI CAN FYGM I TH- HAG D LL CSPST HS- LL CSPST HS- D LL CSPST - KL C- D LL CSSSSHNC- N LL ASKTKEAS- D LLASTTOACC- D LLAASTKEAC- D LLAASTKEAC- D LLAASTKEAC- D LLAAETKEKC- D LLAAETKEKC- D LLAAETKEKC- D LLAAETKEKC- D LLAAETKEKC- D LLAAETKEKC- D LLAAETKEKC- D LLAASTAKDC- D LLAASTKACD- D LLAASTKACD- D LLAASTKEAC- D LLAASTKEKC- D LLAASTKEKC- D LLAASTKEAC- D LLAASTKEKC- D LLAASTKERC- D LLASSAKEL-	SRASEDFLNFLAG VTKLPVI FLLCSP CEHAI QALNFLAG LRSEPTEI LHFLGA DNNTI LVLKHLAE COTIVICIUNFLAG COTIVICIUNFLAG COTIVICIUNFLAG COTIVICIUNFLAG COTIVICIUNFLAG COTIVICIUNFLAG COTIVICIUNFLAG COTIVICIUNFLAG VRGTIALLALLST VRGTIALLALLST VRGTIALLALLSL VRGTKALLEPLME PASQKGRRTQLQAF VRGTKALLGLIEG VEGTKALLEPLME VRGTKALLGLIEG VRGTKALLGLIEG COTIVICIUNTQATT COTIVICIUNCA		A GEVKOLGLIS A GEVKOLGLKFS ETKVTVLOVGI ETKVTVLOVGI - OPSCLRTIM LNI VØLL-ALEV TOVKYLGFIN LTOVKYLGFIN STTVNYMELLLT GETVKYLGFUS STTVNYMELLLT GETVKYLGFUS XRGWKYLGFVI KEEVRYLGFVLF KEEVRYLGFVLF KEEVRYLGFVLG MPPI TVLGRVI NPFSYLGRVL MPFIYLGRVI NPFSYLGRVLG	EGRR AI GPE SHETQ AI GPE SHETQ AI GPE THOSR RLSSC JDS SLSVD JDS SLSVD JDS SLSVD JDS SLSVD JDS SLSVD JDS	RI KPI LNHPLPMILL RI KPI L-HPLPSAI RVGR LGLPSPMIR RVGR LGLPSPMIR RENKLLG.SAPKIR RRVLI SDAMPATK RRVLI SDA	RQL RGFLG RGL HGFLG KQL RAFLG KQL QAFLG CKL QAFLG CKL QAFLG CKL CAFLG CKL RAFLG RELSLG RCMREF-G RCMREF	I TGYCR W P I TDYC I C P L TGYCR W P L TGYCR W P L TGYCFR P RAGCCR - FP MTNFCCI LDW L LNFFR I W LLNFFR W P LLNFFR W P KAEFCR W P ATGFCR W P ATGFCR W P ATGFCR W P ATGFCR W P DMP-MCFFC DMP		
BoERV2 BoERV3 BoERV4 BoERV5 BoERV7 BoERV7 BoERV10 BoERV10 BoERV12 BoERV12 BoERV13 BoERV14 BoERV15 BoERV16 BoERV17 BoERV18 BoERV19 BoERV21	LFGQLADLDLOLG- LFGQLADLTAFWHO- LFGQLGDLDLDLG- LFGALGGLERETOLO- LFG-ALAODLADLDLG- LFG-ALAODLADLTEC MTOSLGGELRETOLO- LFGSLGFELRETTLK- FFGQALASDLTSPDLT- FFGQALASDLTSPDLT- FFGQALASDLTSPDLT- FFGQALASDLASDLGLAN RFGEALSALAAFPGET LFGGLASDLSFPF- ST IFGEALADLSFCPGE PPSLNQAG I EARAGI LFGQLASDLNSFHPEE LOCKFVNAAI EDI RAKY LCONFVOALONARTKY LCONFVOALONARTKY LCONFVOALONARTKY	- EAWLQVDD - EAWLQVDD - SL- KLSKM - NGKI LQVDD - GTVI LQVD YTVSI LITIKEL - SSLLQVDE - SSLLQVDE - NSNL QVD - LSTVLQVDD - LSTVLQVDD - LSTVLQVD LQVD - EG LLQVXQ - LQVNLQVD - EG LLQVXQ - EG LLQVXQ - EG LLQVXQ - EG LLQVXQ - FROTLLQVD - FROTLQVD - FROTLQVD - FROTLQVD - FROTLQVD - FROTLQVD - FROTLQVD - FOR - FOR	D LLCAETEERC D LLCAETEERC LLI CSPNEKSA- LLI CSPNEKSA- LLI CSPNEKSA- LLI CSPNITS- LLI CSPNITS- LLCSSSFINC- LLCSSSFINC- LLCSSSFINC- LLCSSSFINC- LLCSSSFINC- LLCSSSFINC- LLCSSSFINC- LLCSSSFINC- LLCSSSFINC- LLCSSSFINC- LLAAENKDC- LLA	SRASEDFLNFLAG VTKLPVI FLCSF CEHAI QALNFLAG CREVENI QALNFLAG CREVENI LHFLQA CREVENI LHFLQA CREVENI LHFLQA CREVENI LHFLQA COTI I QLLNFLAU COTI I QLLNFLAU COTI I QLLNFLAU VRGTI ALLALLST LEGTIAALLSLIME VEGTIALLEPIM PASQKGRRTQLQA VRGTI ALLALLST VRGTI ALLALLST VRGTI ALLALLST VRGTI ALLALLST VRGTI ALLALLST VRGTI ALLALLST COTI SELI FALES CHI FLNTEI KLKE CEYFLKLKEQLEC	Corycasrekaci (Corycasrekaci (Corycasrekaci (Corycasrekaci (Corycasrekaci (Correspondent) Correspondent (Correspondent)	ACCONTRACTOR	L	R KPI LNHPLPMIL R KPI L-HPLPSAI RVCRI LQ. PSPMIR RVCRI LQ. SPMIR RRVCLI G.SAPKTK RRXLI LSDNANPATK RRXLI SDLPPTPT RRXLI LSDLPPTSK R QALI NCPLPKT RRXLI CSI PVPTTK RRQAI CSI PVPTTK NGI PLQY RRQAI CSI PVPTTK NGI PLQY RRQAI CSI PVPTTK NGI PLQY RRQAI CSI PVPTTK NGI PLQY LWLAS NKDNCO HAPLKLRKDH_VTL LMALAS NKDNCO HAPLKLRKDH_VTL PQKVQVCRDSI KI L	RCL RGFLG	I TGYCR W P I TDYC I C P LTGYCR W P LTGYCFR. P RAGCRI - FP MTNFCCI LDP LLNFFR VWP LACYLSLW P LLNFFR WP KAEFCRI W P SAGCOW W P SAGCOW W P SMCFCV W P DP- NCFFC DWP- NCFFC DWP - NCFFC SMCFCV W P SMCFCV W P SMCV W P SMCFCV W P SMCFCV W P SMCFCV W P SMCV W P		

FIG. 3. Partial amino acid sequence of the RT region of the representative ERVs from the 24 putative families. The positions of the functional motifs LPQG and YV/MDD are boxed.

ods for ERV detection, we found that each method yields different and, in some cases, discordant information.

The BLAST-based search detected the fewest elements (928 elements), most of which were also detected by Retrotector. As the criteria used in the BLAST-based search were quite strict, the elements that it detected could be considered to be highly conserved ERVs.

LTR_STRUC detected 4,487 elements. It identified more elements without the RT region than did BLAST and Retrotector. It also detected many elements that were not identified by BLAST and Retrotector. Because LTR_STRUC is designed to find elements flanked by LTRs, it may be able to detect elements with a noncanonical structure (22).

Retrotector detected the most possible BoERVs (9,698) and had the most overlapping detections. In most of the elements detected by Retrotector, all three main genes were identified. It is thus clear that it is more efficient than BLAST-based detection and able to detect elements that are not as highly conserved (38).

Comparison of different genomes is problematic because various methods have been used to detect ERVs. In previous studies of human (20), mouse (26), rat (7), dog (21), cat (30), and cow (5), RepeatMasker and Repbase were used to detect repetitive elements. However, as stated previously by Sperber et al. (38), results from RepeatMasker and Retrotector cannot be directly compared because the RepeatMasker output is difficult to organize into proviruses. In addition, Retrotector rarely detects elements less than 1,000 bp long, and Repeat-Masker can detect much shorter repeats and single LTRs. Moreover, the secondary integration of proviruses into each other, a feature of old elements, can also be a problem (38).

In a previous study of the cow genome, 142,096 ERVs were detected with PALS/PILER (5), while we identified 928 with BLAST, 4,487 with LTR_STRUC, and 9,698 with Retrotector. The genome coverage of the elements detected by the different programs was also discordant: 1.75% of the genome by PALS/ PILER, 0.36% by BLAST, 1.77% by LTR STRUC, and 4.29% by Retrotector. These data suggest that the coverage is similar or greater with fewer elements. Thus, the abundance of short elements by methods such as RepeatMasker and PALS/PILER make cross-species comparisons difficult. In addition, the classification of the elements detected by the different programs adds complexity to the comparison: RepeatMasker uses the Repbase annotation (13; Smit et al., personal communication), and Retrotector uses its own motif database (38). Thus, we found that RepeatMasker and Retrotector did not routinely sort the same element into the same class. For example, among ERVs classified as class I by the Retrotector method, 64.72% were classified as ERV1 and 35.28% were classified as ERVL by RepeatMasker.



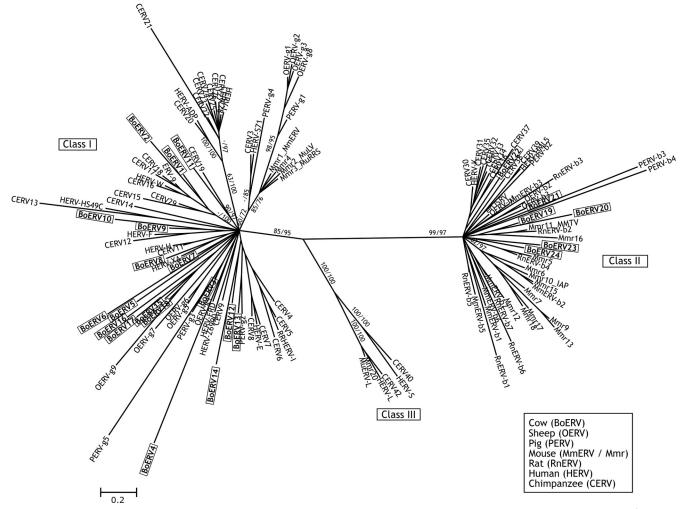


FIG. 4. RT region-based unrooted phylogenetic tree of ERVs from different species. Topology is based on Bayesian inference (10^6 generations). In the branches on the left, the maximum likelihood bootstrap value is shown; on the right, the Bayesian posterior probability is shown. Sixteen representative sequences from human ERV families (40), 38 from chimpanzees (29), 27 from mice (2, 23), 7 from rats (2), 14 from sheep (17), 8 from pigs (27), and 24 from cows (our data) are included. Representatives of BoERV families are boxed.

One explanation for the different distributions of ERVs across bovine chromosomes could lie in the target elements employed by the methods used to identify ERVs. In the analysis of the chromosomal distribution of the elements detected, the various methods showed different chromosomes that did not follow any homogeneous distribution. The nature of the elements detected by each method could be a good reason for this discrepancy.

Across chromosomes, the BLAST-based and Retrotector methods identified significantly more ERVs in the X chromosome than would be expected from a homogeneous distribution. A similar excess of ERVs has been observed for the human X chromosome (41).

The number of ERVs detected was positively correlated with chromosome length (P < 0.001 for all three methods) and negatively correlated with the GC content of the chromosome (P < 0.001 for all three methods). No correlation was observed between the number of ERVs detected and gene or pseudo-gene density. In humans, the number of class I and class III ERVs—but not the number of class II ERVs—has been neg-

atively correlated with GC content (24). The insertion preferences of ERVs in the cow genome should be analyzed in greater detail to gain a better understanding of the preferences of bovine ERVs.

Phylogenetic analysis based on the RT region of a number of selected elements was used to cluster these elements into 24 putative families, which we called BoERV families. Previously, 4 retroviral families were detected (43), which are included in the 24 families that we identified. Although it wad previously suggested that the BERV- γ 4 family, referred to here as BoERV3, was the most abundant (43), we found that BoERV1 was actually the most abundant. This family had not previously been identified in any mammal. One possible explanation for this is that the members of this family have some nucleotide differences in the region where hybridization took place with the primers used for pig, sheep, and cow (43). We used PCR to amplify a 150-base sequence in sheep, so it is possible that BoERV1 could be a ruminant-specific ERV family.

The comparison of ERV family numbers was limited to four species with defined families (human, chimpanzee, mouse, and

rat). In cows, the number of families (24 putative families) was higher than that for mouse (20 families) (23) and lower than those for chimpanzee (42 families) (29) and human (31 families) (15). In the case of rodents, where information is available only for class II elements in two species, the number of families in cow (six families) was similar to those in rat and mouse (seven families) (2). To the best of our knowledge, no information is currently available on dog and cat ERV families.

We did not detect any class III-related ERVs. Although this could be an artifact due to the distance from the reference sequences used for the BLAST-based search and the limits of class III element detection by Retrotector (38), it is more likely because the presence of class III ERVs in the cow genome is limited. In fact, although a number of sequences related to class III were amplified previously by Bénit et al. (3), the amplification signal was weak, and these sequences were quite short.

The relationship between representatives of the ERV families from different species is interesting. In general, the lineages of the different ERV groups are divided following the species phylogeny, with humans and chimpanzees on one side and cows, sheep, and pigs on the other side. Representative elements of the scarce murine class I families were included in our analysis, but their relationship with representative elements of the ERV families of other species remains obscure. Even so, representative elements of the human/chimpanzee groups and, to a lesser extent, mouse/rat and pig/sheep groups tend to follow the pattern of previously reported comparisons of each pair (2, 17, 29). Following this pattern, the representative bovine elements cluster with the representative sheep elements, as obtained by experiments (17) with most of the lineages. In some cow breeds, ovine enJSRV-related env, orf-x, and LTR sequences have been detected (25). However, bovine ERVs closely related to enJSRVs were not detected in the version of the genome used in our study. This genome sequence belongs to a Hereford animal, while Morozov et al. analyzed animals from Simmental and Limousine breeds. For humans, it was suggested that a combination of genetic and environmental factors could contribute to determining the prevalence of enJSRV-related sequences in different populations (34). Thus, it is possible that different breeds of cow could also have different prevalences of enJSRV-related sequences.

Related to the relationship of ERV families of different species, in one lineage, representatives of human, chimpanzee, pig, and sheep groups were present, while cattle elements were absent. To account for this absence, we estimated the insertion time of the elements in this lineage. As there is no genomic information available for pigs and sheep, estimates were available only for human (19.5 to 8.9 MYA) and chimpanzee (33 to 15.8 MYA) elements. These insertion times were later than the divergence of ruminants and primates. Based on the weak support of the tree topology, a single infection is unlikely. In this lineage, two independent infections by a similar virus could have been detected, and in the case of ruminants, it is possible that cows lost this element at some point.

The absence of some ERV families in cows, compared with sheep and pigs, has prompted some authors to suggest that cows have a limited number of ERV families (43). Taking into account that the numbers of ERV families described were 31 for humans (15), 42 for chimpanzees (29), 20 for mice (23), and 24 putative families for cows (this study), BoERVs may not be as scarce as previously stated. Moreover, we detected one family, BoERV1, that had not been detected previously but that appears to be present at least in ruminants.

As described above, we did not detect any class III elements. It was suggested previously that in primates and mice (18), ERVs related to this class have been subjected to one or two bursts of copy number. If so, it is possible that the difference in the number of ERV families with primates and mice could be based on this burst of class III-related ERVs. Finally, the whole picture could be also confused by the intense selective breeding processes that have accompanied the domestication of cows (4).

In conclusion, we identified several thousand ERVs in the genome of *Bos taurus* by three different methods. The number detected depended on the technique used, ranging from a low of 928 using a BLAST-based method to 9,698 using Retrotector. When attempting to detect new ERVs, the use of different methods is advisable. ERVs did not appear to be randomly scattered across the chromosome but were more abundant on some, especially the X chromosome, than on others. Among the 24 detected families, 20 were newly described ERV families. The most abundant BoERV1 family is described for the first time. Finally, representatives of ERV families from rodents, primates, and ruminants showed a phylogenetic relationship following their hosts' relationships.

This is indeed the first genome-wide approach for the detection and characterization of bovine endogenous retroviruses. Further in-depth analyses are thus needed to uncover the whole picture of these genomic elements in cattle.

ACKNOWLEDGMENTS

K.G.-E. was a recipient of a UPV/EHU grant (Vice-Rectorate of Basque and Plurilingualism). This work has been partially funded by UPV/EHU by means of projects EHU06/107 and GIU07/62 (B.M.J.).

We thank anonymous reviewers for highly useful comments and improvements to the manuscript. We thank Maialen Sistiaga for helping in the experimental work.

REFERENCES

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.
- Baillie, G., L. van de Lagemaat, C. Baust, and D. Mager. 2004. Multiple groups of endogenous betaretroviruses in mice, rats, and other mammals. J. Virol. 78:5784–5798.
- Bénit, L., J. Lallemand, J. Casella, H. Philippe, and T. Heidmann. 1999. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. J. Virol. 73:3301–3308.
- Boeke, J., and J. Stoye. 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements, p. 343–435. *In J. Coffin, S. Hughes, and* H. Varmos (ed.), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Bovine Genome Sequencing and Analysis Consortium, C. Elsik, R. Tellam, K. Worley, et al. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 324:522–528.
- Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17:540–552.
- Gibbs, R. A., G. M. Weinstock, M. L. Metzger, D. M. Muzny, E. J. Sodergren, et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493–521.
- Gifford, R., P. Kabat, J. Martin, C. Lynch, and M. Tristem. 2005. Evolution and distribution of class II-related endogenous retroviruses. J. Virol. 79: 6478–6486.
- Golan, M., A. Hizi, J. H. Resau, N. Yaal-Hahoshen, H. Reichman, I. Keydar, and I. Tsarfaty. 2008. Human endogenous retrovirus (HERV-K) reverse transcriptase as a breast cancer prognostic marker. Neoplasia 10:521–523.
- Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

- Jern, P., and J. Coffin. 2008. Effects of retroviruses on host genome function. Annu. Rev. Genet. 42:709–732.
- Johnson, W., and J. Coffin. 1999. Constructing primate phylogenies front ancient retrovirus sequences. Proc. Natl. Acad. Sci. U. S. A. 96:10254–10260.
- Jurka, J. 2000. Repbase update—a database and an electronic journal of repetitive elements. Trends Genet. 16:418–420.
- Katoh, K., K. Kuma, H. Toh, and T. Miyata. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33:511–518.
- Katzourakis, A., A. Rambaut, and O. Pybus. 2005. The evolutionary dynamics of endogenous retroviruses. Trends Microbiol. 13:463–468.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. J. Mol. Evol. 16:111–120.
- Klymiuk, N., M. Muller, G. Brem, and B. Aigner. 2003. Characterization of endogenous retroviruses in sheep. J. Virol. 77:11268–11273.
- Kumar, S., and S. Hedges. 1998. A molecular timescale for vertebrate evolution. Nature 392:917–920.
- Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5:150–163.
- Lander, S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.
- 21. Lindblad-Toh, K., C. Wade, T. Mikkelsen, E. Karlsson, D. Jaffe, M. Kamal, M. Clamp, J. Chang, E. Kulbokas, M. Zody, E. Mauceli, X. Xie, M. Breen, R. Wayne, E. Ostrander, C. Ponting, F. Galibert, D. Smith, P. deJong, F. Kirkness, P. Alvarez, T. Biagi, W. Brockman, J. Butler, C. Chin, A. Cook, J. Cuff, M. Daly, D. DeCaprio, S. Gnerre, M. Grabherr, M. Kellis, M. Kleber, C. Bardeleben, L. Goodstadt, A. Heger, C. Hitte, L. Kim, K. Koepfli, H. Parker, J. Pollinger, S. Searle, N. Sutter, R. Thomas, C. Webber, E. Lander, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438:803-819.
- McCarthy, E., and J. McDonald. 2003. LTR_STRUC: a novel search and identification program for LTR retrotransposons. Bioinformatics 19:362– 367.
- McCarthy, E. M., and J. F. McDonald. 2004. Long terminal repeat retrotransposons of Mus musculus. Genome Biol. 5:R14.
- Medstrand, P., L. van de Lagemaat, and D. Mager. 2002. Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res. 12:1483–1495.
- Morozov, V., A. Morozov, and S. Lagaye. 2007. Endogenous JSRV like proviruses in domestic cattle: analysis of sequences and transcripts. Virology 367:59–70.
- Mouse Genome Sequencing Consortium, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562.
- Patience, C., W. Switzer, Y. Takeuchi, D. Griffiths, M. Goward, W. Heneine, J. Stoye, and R. Weiss. 2001. Multiple groups of novel retroviral genomes in pigs and related species. J. Virol. 75:2771–2775.
- Perron, H., C. Bernard, J. B. Bertrand, A. B. Lang, I. Popa, K. Sanhadji, and J. Portoukalian. 2009. Endogenous retroviral genes, herpesviruses and gender in multiple sclerosis. J. Neurol. Sci. 286:65–72.
- 29. Polavarapu, N., N. J. Bowen, and J. F. McDonald. 2006. Identification,

characterization and comparative genomics of chimpanzee endogenous retroviruses. Genome Biol. 7:R51.

- 30. Pontius, J. U., J. C. Mullikin, D. R. Smith, the Agencourt Sequencing Team, K. Lindblad-Toh, S. Gnerre, M. Clamp, J. Chang, R. Stephens, B. Neelam, N. Volfovsky, A. A. Schaffer, R. Agarwala, K. Narfstrom, W. J. Murphy, U. Giger, A. L. Roca, A. Antunes, M. Menotti-Raymond, N. Yuhki, J. Pecon-Slattery, W. E. Johnson, G. Bourque, G. Tesler, the NISC Comparative Sequencing Program, and S. J. O'Brien. 2007. Initial sequence and comparative analysis of the cat genome. Genome Res. 17:1675–1689.
- Posada, D., and K. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.
- 32. R Development Core Team. 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16:276–277.
- Rocca, S., M. Sanna, A. Leoni, A. Cossu, A. Lissia, F. Tanda, M. Satta, and G. Palmieri. 2008. Presence of Jaagsiekte sheep retrovirus in tissue sections from human bronchioloalveolar carcinoma depends on patients' geographical origin. Hum. Pathol. 39:303–304.
- Ronquist, F., and J. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.
- 36. Serafino, A., E. Balestrieri, P. Pierimarchi, C. Matteucci, G. Moroni, E. Oricchio, G. Rasi, A. Mastino, C. Spadafora, E. Garaci, and P. S. Vallebona. 2009. The activation of human endogenous retrovirus K (HERV-K) is implicated in melanoma cell malignant transformation. Exp. Cell Res. 315:849–862.
- 37. Sokal, R., and F. Rohls. 1969. Biometry: the principles and practise of statistics in biological research. W. H. Freeman and Co., New York, NY.
- Sperber, G., T. Airola, P. Jern, and J. Blomberg. 2007. Automated recognition of retroviral sequences in genomic data—RetroTector. Nucleic Acids Res. 35:4964–4976.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
- Tristem, M. 2000. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the Human Genome Mapping Project database. J. Virol. 74:3715–3730.
- Villesen, P., L. Aagaard, C. Wiuf, and F. S. Pedersen. 2004. Identification of endogenous retroviral reading frames in the human genome. Retrovirology 1:32.
- Xiao, R., J. Kim, H. Choi, K. Park, H. Lee, and C. Park. 2008. Characterization of the bovine endogenous retrovirus beta 3 genome. Mol. Cells 25:142–147.
- Xiao, R., K. Park, H. Lee, J. Kim, and C. Park. 2008. Identification and classification of endogenous retroviruses in cattle. J. Virol. 82:582–587.
- 44. Xiao, R., K. Park, Y. Oh, J. Kim, and C. Park. 2008. Structural characterization of the genome of BERV gamma 4, the most abundant endogenous retrovirus family in cattle. Mol. Cells 26:404–408.
- Zhang, Z., S. Schwartz, L. Wagner, and W. Miller. 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 7:203–214.