Unraveling viral identity: Avoiding the trap of endogenous sequences for viral surveillance of small ruminant oncogenic retroviruses

Riocreux-Verney Benjamin[1], Verneret Marie[1,2], Dolmazon Christine[1], Shirin Ashraf[3], Stella A. Atim[3,4,5], Navratil Vincent [6,7], Leroux Caroline[1 ¶,*], Turpin Jocelyn[1 ¶,*]

[1] IVPC UMR754, INRAE, Universite Claude Bernard Lyon 1, EPHE, Université PSL, France.

[2] Universite Claude Bernard Lyon 1, LBBE, UMR 5558, CNRS, VetAgro Sup, Villeurbanne, France.

[3] MRC-University of Glasgow Centre for Virus Research, (CVR), Glasgow, UK.

[4] MRC/UVRI and LSHTM Uganda Research Unit, Entebbe, Uganda.

[5] Ministry of Agriculture, Animal Industry and Fisheries, Entebbe, Uganda

[6] PRABI, Pôle Rhône-Alpes Bioinformatics Center, Universite Claude Bernard Lyon 1, Villeurbanne, France.

[7] European Virus Bioinformatics Center, Leutragraben 1, Jena, Germany.

*Correspondence: caroline.leroux@inrae.fr and jocelyn.turpin@inrae.fr

¶These authors contributed equally to this work and share last authorship.

Keywords

Virome, endogenous retrovirus, vectors, small ruminants, JSRV, ENTV

Abstract

Small ruminants (sheep and goats) are one of the few mammals in which an exogenous retrovirus (XRV) and closely related endogenous retroviral elements (ERV) coexist within the same host genome. The betaretroviruses Jaagsiekte sheep retrovirus (JSRV) and enzootic nasal tumor virus (ENTV) cause pulmonary and nasal adenocarcinomas, respectively, and share extensive sequence similarity with their endogenous counterparts. Consequently, molecular surveillance must rely on assays that can unequivocally distinguish true exogenous infection from ERV-derived templates; failure to do so compromises diagnosis, phylogenetic inference, and epidemiological conclusions.

We retrieved all complete JSRV, ENTV-1/2, and related ERV genomes deposited in public repositories and performed a comprehensive alignment. Only a limited number of genomic segments were capable of distinguishing exogenous from endogenous sequences. We refer to these as discriminating regions (DRs). Phylogenies built using DRs revealed that several entries annotated as XRV are, in fact, ERV-derived or chimeric artefacts generated by short-amplicon reconstruction.

A systematic literature review of over 100 articles identified 286 distinct primers and probes used for the XRV amplification. *In-silico* mapping of each oligonucleotide onto the full alignment showed that only 28 % reliably differentiate XRV from ERV. We experimentally validated the predictive power of this approach for 17 primer/probe sets, confirming that non-discriminating assays produce false-positive signals from endogenous templates.

The misannotation of ERV sequences as exogenous viruses has resulting in the population of databases with dubious entries, fostering erroneous hypotheses such as

46 vector-borne transmission of JSRV and ENTV. To address this issue, we propose a
47 concise set of criteria for assay design, validation, and database annotation
48 emphasizing DR targeting, specificity testing against endogenous templates, and
49 transparent reporting.

50 Although this framework was developed for small ruminants, it is readily applicable to
51 any host–virus system in which exogenous viruses coexist with endogenous viral
52 elements. This will strengthen viral surveillance, phylogenetics, and the One Health
53 initiatives.

54 Introduction

55 Cancers induced by the exogenous retroviruses (XRVs), Jaagsiekte Sheep RetroVirus
56 (JSRV) and Enzootic Nasal Tumor Virus (ENTV), have been documented in all regions
57 where sheep and goats are reared, except in Australia, New Zealand, the Falkland
58 Islands and Iceland, where JSRV has been eradicated [1–3]. Multiple copies of
59 endogenous retrovirus (ERV) family II.5, previously named enJSRV and enENTV
60 [4,5], are present in goats (ERV II.5 Ch for *Capra hircus*) and sheep (ERV II.5 Oa for
61 *Ovis aries*), with an average of 223 and 98 loci respectively [6]. The presence of whole
62 copies with little divergence and with preserved open reading frames (*gag, pro, pol*
63 and *env*), suggests the recent activity of ERV II.5 Ch in goats [6]. Despite their close
64 genetic relationship with domestic sheep and goats XRVs, ERVs are not implicated in
65 the development of respiratory cancers in domestic small ruminants. Therefore, the
66 exogenous retroviruses JSRV and ENTV remain the only known aetiological agents
67 of respiratory cancers in domestic small ruminants.

68 Developing robust virus detection tools is an essential for managing the infectious risk
69 posed by these incurable, fatal and transmissible cancers that threaten small-ruminant

70 farming. Serological detection of antibodies is a tool widely used in both veterinary and
71 human medicine. Apart from the fact that it only allows indirect detection of infection
72 through the demonstration of antibody production, it requires a detectable and stable
73 humoral response. In a context of multiple ERV copies impairing the development of
74 a XRV-specific humoral response, antibody detection in JSRV or ENTV infected
75 animals is not a reliable tool to establish the infectious status [7–10].

76 Therefore infection can only be detected *ante or post mortem* using molecular
77 approaches, *i.e.* PCR and/or RT-PCR [7,11]. The genetic diversity of JSRV and ENTV
78 is only partially known, and before our recent study reporting on 29 JSRV and 24 ENTV
79 partial and full proviral sequences [12], only a few dozen complete sequences were
80 available in the databases, i.e. nine complete sequences of JSRV isolated in Africa,
81 North America, Europe and Asia [13–17], and 50 complete or near-full length ENTV
82 genomes isolated in Canada and Europe for ENTV-1 [18,19] and in Europe and Asia
83 for ENTV-2 [20–26], including sequences not associated with accessible publications
84 or unpublished.

85 The widespread use of high-throughput sequencing tools has made pathogen
86 sequences widely available. In theory, detecting viral or proviral genomes using
87 targeted molecular approaches is straightforward, given the multiple bioinformatic
88 tools available for designing primers and probes. However, when dealing with the
89 coexistence of genetically closely related ERV family II.5 and XRVs, it is crucial to
90 ensure the strict XRV specificity of the primers or probes. To optimize the molecular
91 XRV detection tools in regards to published studies, we reviewed multiple reports,
92 focusing on the molecular detection and sequencing tools used and their relevance in
93 the context of the coexistence of XRVs with ERV II.5 *loci* [6]. A recent study of all

94 available sequences in databases allowed us to identified sequences wrongly
95 identified as exogenous JSRV or ENTV despite the absence of the YXXM motif in the
96 region of the *env* gene encoding the intracytoplasmic tail (CT), suggesting their
97 endogenous origin [12]. The YXXM motif is an essential oncogenic determinant, and
98 a genetic marker of exogenous sequences, that is strictly absent from all the ERV II.5
99 Oa and Ch sequences identified up to date [6,12,18,27]. On the other hand, some
100 studies have relied on RNAs to specifically detect XRVs [20,23,28–31]. This approach
101 assumes that ERVs are not expressed in tissues, which is risky as we will show in this
102 study.

103 In a large retrospective study based on published and unpublished sequences
104 included in the public archives, we questioned the truly exogenous nature of the
105 detected and sequenced viral genomes, after highlighting errors in the annotation of
106 sequences unduly identified as XRV. To this end, we defined the regions of the JSRV
107 and ENTV genomes specific to XRVs. We then reanalysed the available sequences
108 and published amplification tools to determine their specificity, verified the XRV
109 specificity of primers reported in the literature, and proposed a systematic approach
110 with rigorous controls to ensure the exogenous nature of the sequences. We apply our
111 methodology to case studies reporting on the detection of JSRV or ENTV in ticks and
112 mosquitos [32–37] using metagenomics or virome approaches and their potential
113 transmission by those vectors.

114 Our study raises questions about the robustness of the virus discovery in the context
115 of XRVs/ERVs coexistence in small ruminants and the overall challenge in
116 metagenomic and transcriptomic studies in the presence of ERVs and more globally
117 endogenous viral elements (EVEs) [38].

118 ## Materials and Methods

119 ### Determination of discriminating regions between XRVs et ERVs

120 Complete nucleic sequences from 30 JSRV, 14 ENTV-1 and 35 ENTV-2 sequences
121 available in GenBank (S1 Table, sequences identified with an asterisk *) were aligned
122 with the ERV II.5 consensus for *Ovis aries* for JSRV and ENTV-1, and for *Capra hircus*
123 for ENTV-2 [6]. Nucleotide alignments were performed with the 'MAFFT v7.490'
124 software [39] and processed with the 'R v4.4.1' suite to picture the variability of the
125 identity scores obtained from the comparison between XRV sequences and the ERV
126 consensus and to calculate the percentage of identity for each position in the
127 alignment, averaged over a 100 bp window with a 10 bp step. The analysis was
128 performed individually for each sequence of the alignment, and the mean percentage
129 of identity was plotted. A 95% confidence interval was estimated using a bootstrap
130 approach with the 'boot v1.3-31' suite [40], debiased and corrected for asymmetric
131 data. The limited number of samples used and the potential lack of independence
132 between sequences originated the same origins (herd, geographical location, etc.) are
133 potential biases that we have identified. Four discriminant regions (referred further as
134 DR1, DR2, DR3, DR4), were identified with a percentage of identity lower than 80%.
135 DR2 sequence variability was evaluated by aligning (MAFFT Alignment v7.490 [39])
136 copies containing the *gag* region, *i.e.* 28 ERV II.5 copies annotated from the *Ovis aries*
137 assemby (ARS-UI_Ramb_v2.0 -GCA_016772045.1) and 155 ERV II.5 copies from the
138 *Capra hircus* assembly (ARS1.2 -GCA_001704415.2), and 27 ovine and 10 caprine
139 ERV sequences recovered from GenBank [6] (S1 Table). The percentage of each
140 nucleotide among all copies was calculated at each position.

141 ### *in silico* and experimental evaluation of XRV primers specificity

142 The localization prediction of 286 published PCR primers and probes (S2 Table)
143 reported for the detection of JSRV et ENTVs, was determined by mapping them onto
144 the respective alignments for each virus. Publications not written in English were
145 automatically translated with DeepL. A maximum of 4 mismatches from the XRV
146 sequences was allowed, and the position of each primer was extracted from the
147 alignment retaining the multi-mapped positions. Primers including additional
148 sequences, *e.g.* restriction sites, were not included. Primers only mapping on XRV
149 were defined as XRV specific whereas primers mapping on a least one copy of ERV,
150 were defined as non-specific to exogenous sequences. A selection of seventeen
151 primer pairs has been tested on various DNA templates, covering different parts of the
152 viral genomes. Fourteen of them were tested by PCR on mammalian DNAs; they
153 targeted ENTV-2 for primer pairs # 01-10 or JSRV for primers pairs 12-15 (S3 Table).
154 Their ability to detect only XRV was tested using genomic DNAs from JSRV-induced
155 lung tumours of clades I (FR2054) and II (FR989), and ENTV-2-induced nasal tumours
156 (FR3824) [12] (Table 1). To evaluate their potential for cross hybridization with ERV
157 sequences, PCRs were performed using DNAs from ovine IDO5 cells (Institut Mérieux,
158 Lyon, France), caprine TiGEF cells [41] and human HeLa cells [42] (Table 1). Primer
159 pairs 10 (ENTV-2 U3/gag) and 15 (JSRV U3/gag) have been recently published by
160 our team and shown to be highly specific for JSRV and ENTV-2, with no cross
161 hybridization with ERV sequences [12]. The amplicons obtained with pairs 01 and 13
162 for each condition (JSRV, ENTV and cell references), with pairs 03 to 08 and 14 for
163 the IDO5 cell reference and with pair 04 for HeLa cells were subjected to Sanger
164 sequencing (TubeSeq Supreme Eurofins) and aligned against XRV and ERV
165 reference proviral sequences. A BLAT search [43] on the human genome (hg38) was

166 performed to identify the nature of the amplicon sequence obtained with pair 04 from
167 human HeLa cells.

168 A similar approach was taken for the primer pairs published to target ENTV (primer
169 pair #11) and JSRV (primer pair 16 and 17) in arthropods [35,36] by adding DNA
170 extracted from an ENTV-1-induced nasal tumour (FR4403) [12] and mosquitos (*Aedes*
171 *aegypti and Culex Quique/asciatus*) that have fed on human blood. All the amplicons
172 were subjected to Sanger sequencing (TubeSeq Supreme Eurofins).

173 **Table 1. Biological samples: origin and sequences.**

Name	Biological description	Species origin	Genbank accession #	Reference
FR4403	French ENTV-1	*Ovis aries*	PP669280	(Riocreux-Verney et al., 2024)
FR3824	French ENTV-2	*Capra hircus*	PP669281	(Riocreux-Verney et al., 2024)
FR2054	French JSRV Clade I	*Ovis aries*	PP707050	(Riocreux-Verney et al., 2024)
FR989	French JSRV Clade II	*Ovis aries*	PP707047	(Riocreux-Verney et al., 2024)
IDO5	Ovine dermal fibroblasts	*Ovis aries*	NA	NA
TiGEF	T-immortalized Goat Embryonic Fibroblasts	*Capra hircus*	NA	(Da Silva Teixeira et al., 1997)
HeLa	Immortalized human cell line	*Homo Sapiens*	NA	(Scherer et al., 1953)
Aa	Mosquito	*Aedes aegypti*	NA	NA
Cq/a	Mosquito	*Culex quinque/asciatus*	NA	NA

174

175 **ERV II.5 expression in sheep tissues**

176 Data of mRNA-seq from 60 tissues (Table 2, Ovine FAANG project [44,45]) of the
177 Benz2616 ewe, used to assemble the ARS-UI_Ramb_v2.0 ovine genome
178 (GCA_016772045. 1, [46]), and of lung RNA-seq of $JSRV_{JS21}$-experimentally infected
179 lambs and the associated uninfected controls [47] were collected from the ENA
180 (European Nucleotide Archive) database. ERV II.5 family - specific k-mers of 21 base
181 pairs were defined in the DR1 and DR3 regions in order to cover the largest number
182 of copies of ERVs annotated in the ARS-UI_Ramb_v2.0 genome (Table 3 and S1 Fig)
183 [6]. These k-mers were used to identify and count the reads corresponding to ERV

184 sequences in the selected RNAseq data using the BBduk analysis tool (Bbtools
185 v38.95, https://quay.io/repository/biocontainers/bbmap) by allowing two mismatches.
186 As a control, k-mers specific to JSRV were defined in DR3 and DR4 and searched in
187 the same RNAseq data (Table 3 and S1 Fig), as previously described [12]. The reads
188 containing the kmers were extracted from the raw data (FastQC file), low-quality
189 sequences and adapters were removed with Trimmomatic [48]. The sequences were
190 then mapped (Bwa-mem, [49]) on the *Ovis aries* family II.5 consensus sequence [6].
191 The alignment was visualised with IGV (Integrative Genomics Viewer, [50]).

192 **Table 2. Tissues used for the ERV expression from RNA-seq data of the sheep**
193 **Benz2616. Tissues were grouped by physiological systems.**

System	Tissue	ENA accession #
Cardiac	Right ventricle	ERR3728426
	Atrioventricular valve	ERR3728429
	Right atrium	ERR3665709
	Left ventricle	SRR7612070
	Vena cava	ERR3728430
Immunitary	Spleen	ERR3728431
	Mesenteric lymph node	ERR3665711
	Prescapular lymph node	ERR3665712
	Mandibular lymph node	SRR7543059
	Alveolar macrophage	SRR6651987
Gastro-intestinal	Ileum	ERR3728427
	Ileum Peyer Patches	ERR3665710
	Cecum	ERR3665706
	Jejunum	SRR7543057
	Duodenum	ERR3650374
	Rectum	ERR3728432
	Rectum	ERR3665713
	Descending colon	SRR7543055
	Spiral colon	ERR3665714
	Esophagus	ERR3665707
	Abomasum	ERR3665705
	Abomasum pylorus	ERR3650373
	Reticulum	ERR3650375
	Rumen atrium	ERR3650376
	Rumen ventral	ERR3650377
	Omasum	SRR7543060
	Gall bladder	SRR7543056
	Liver	SRR7612072
Skin	Skin hairless	ERR3665715
	Skin of back	SRR7543062

System	Tissues	ENA accession #
Neurological	Retina	ERR3728428
	Spinal cord	ERR3728434
	Pons	ERR3728433
	Thalamus	ERR3728435
	Hippocampus	ERR3665708
	Pituitary gland	ERR3650379
	Cerebellum	SRR6651979
	Frontal cortex	SRR7612068
Renal	Adrenal medulla	SRR7543054
	Adrenal cortex	SRR6651976
	Renal medulla	SRR7543058
	Renal cortex	SRR7612071
Respiratory and buccal	Tongue	SRR7612076
	Soft palate	ERR3650378
	Epiglottis	SRR6651978
	Palatine tonsil	SRR7612075
	Lung	SRR6651986
Urinary	Urinary bladder	ERR3728424
	Urethra	ERR3665716
	Ureter	SRR7543063
Endocrine	Mammary gland	ERR3728425
Genital	Uterus	ERR3665717
	Oviduct	SRR7543061
	Ovary	SRR6651991
	Vagina	SRR7612077
Muscular	Biceps femoris	SRR6651988
	Semimembranosus	SRR6651990
	Longissimus dorsi	SRR7612073
	Supraspinatus	SRR7612074
	Diaphragm	SRR7612069

194

195 **Table 3. k-mer sequences and localization for the identification of tissue**
196 **expression of ERVs of family II.5 or the exogenous virus.**

Target	Discriminating region (DR)	Name	5' - 3 ' sequence
ERV	DR1	kERV-DR1	TATGCTGTTCCCGAGGGGGTT
	DR3	kERV-DR3	GAGAGTTTTAATACATAAAAA
JSRV	DR3	kJSRV-DR3	GAGAGTTGAAATGCTACATAT
	DR4	kJSRV-DR4	TCTGATTATGTAAGAATCCGG

197

198 **Alignment and phylogenetic analyses**

199 The alignments of *pol*, DR1-DR2, DR3-DR4, complete genomes and sequenced
200 amplicons were carried out using MAFFT Alignment v7.490 [39] and Geneious Prime
201 software (v 2025.0.2). Maximum Likelihood (ML) phylogenetic trees in the *pol*, DR1-
202 DR2, DR3-DR4 regions were generated using IQ-TREE v1.5.3 [51] with 10,000
203 ultrafast bootstrap and rooting on the BT ERV (Bos taurus ERV) family II.4 consensus
204 sequence [6]. ML-type phylogenetic trees (PhyML 3.3.20180621, Jukes Cantor
205 substitution model) of the *gag* (pair #01, 826nt) and *env* (pair #13, 864 nt) sequences
206 were created using the ENTV1-FR4403 sequence (GenBank #PP669280, [12]) for
207 rooting, with 1000 iterations (bootstraps) and edited through the iTOL software.

208 Publicly available sequences used in the study are reported in S1 Table. ERV copies
209 from the family II.5, annotated in the *Capra hircus* assembly ARS1.2
210 (GCA_001704415.2) and the *Ovis aries* assembly ARS-UI_Ramb_v2.0
211 (GCA_016772045.1), were retrieved from [6]. The partial proviral sequence
212 OR991120 described in [35] and the partial *env* sequence KT266728 [36] were also
213 analyzed.

214

215 ## Results

216 **The sections of the genome that differ between ERVs and XRVs are limited**

217 To determine the diversity of XRV sequences, we compiled a dataset comprising 79
218 complete sequences of JSRV, ENTV-1, ENTV-2 and ERV II.5 consensus sequences.
219 Analysis of the variation, in the percentage of identity of the XRV genomes against the
220 ERV II.5 Oa consensus for the JSRV and ENTV-1, which infect sheep, or the ERV II.5
221 Ch consensus for the ENTV-2, which infects goats, clearly showed that regions with
222 less than 80% identity between the XRVs and the ERVs are restricted to four
223 discriminating regions, DR1 to DR4. The other regions of the genome do not allow for
224 discrimination between XRVs and ERVs (Fig1). The confidence interval for the ENTV-
225 2 percentage identity is higher than those calculated for JSRV or ENTV-1. In the non-
226 discriminating *pro* and *pol* regions of ENTV-2, the diversity is greater than that
227 observed in the equivalent regions of JSRV and to a lower extent to ENTV-1, notably
228 due to sequences that are genetically close to the goat ERV consensus (S2 Fig, *e.g.*
229 sequences KU258870.1, MK210250.1, ON843769.1, MK164396.1). The DR2 region
230 is less marked in ENTV-1 compared to JSRV and ENTV-2, with a percentage of
231 identity higher than 80%. In addition to the nucleotide polymorphism, the DR1, 3 and
232 4 regions, displayed specific insertion/deletion (Indels) patterns that can discriminate
233 between exogenous and endogenous sequences (S3 Fig). The insertion in DR2 is
234 only present in ERV II.5 Ch and its length varies between ERV copies (S3B Fig).

235 Of the 286 PCR primers/probes listed in the 109 published studies, only 28% (79 out
236 286) were qualified as discriminating between XRVs and ERVs by our *in silico* analysis
237 (S2 Table). Notably, among the 286 primers, 74 (26%) are in the *pro* or *pol* genes,
238 which are regions that are highly similar between XRV and ERV sequences. Similarly,

239 less than one third of the primers in R / U5 region are specific. Therefore, in order to
240 minimize the risk of amplifying ERVs instead of XRV, they must be combined with XRV
241 specific primers.

242 **Phylogenetic reconstructions reveal the challenge of correctly classifying XRV**

243 Phylogenetic trees, were reconstructed in the *pol* region (Fig 2), which weakly
244 discriminates between XRVs and ERVs, as well as in the DR1-DR2 (Fig 2B) and the
245 DR3-DR4 (Fig 2C) segments from 32 JSRV, 14 ENTV-1, 45 ENTV-2, 27 ERV II.5 Oa
246 and 10 ERV II.5 Ch sequences were extracted from GenBank and supplemented by
247 sequences identified on the assemblies of the *Ovis aries* (ARS-UI_Ramb_v2.0) and
248 *Capra hircus* (ARS1.2) reference genomes [6].

249 These analyses revealed that the segregation of viral species varies considerably
250 depending on the region considered. The *pol* region did not permit a clear distinction
251 between XRV and ERV II.5 (Fig 2A). Conversely, the DR1-DR2 and DR3-DR4 regions
252 (Fig 2B-C) enabled the different viral forms to be separated into XRVs and ERVs.
253 However, 17 of the XRV sequences were grouped with the ERV sequences,
254 regardless of whether the trees were reconstructed using the DR1-DR2 or DR3-DR4
255 regions, this calls into question their exogenous origin. Some of the potentially
256 misidentified sequences presented a clear ERV pattern, including specific
257 insertions/deletions (indels) in the DR1-DR4 segment and the lack of the YXXM motif
258 in Env (e.g. MK164396 and MK210250 for ENTV-2 or DQ838494 and MZ931278 for
259 JSRV, Fig 3). The MK210250 sequence displayed a very strong sequence identity
260 (>96%) with the ERV II.5 *Capra hircus* consensus sequence in DR4, whereas the
261 MK164396 sequence did not show a clear pattern, being equally distant from the
262 exogenous and endogenous reference sequences (with an identity percentage

263 ranging from 80 to 83%). Two JSRV sequences, DQ838494 & MZ931278, which
264 cluster with ERVs had DR4/U3 sequences that are very different from the ERV II.5
265 consensus sequence with 35% and 71% identity respectively, as well as from the
266 JSRV-FR2054 (PP707050.1) reference sequence with 33% and 52% identity,
267 respectively. The BLAT of the DR4 sequence of DQ838494 against the sheep genome
268 (Oar_v4.0/oviAri4) resulted in multiple hits. The first hits (with over 95% identity)
269 corresponded to LINE sequences of the RTE-BovB family, suggested the unmonitored
270 amplification of another transposable element.

271 Different combinations of mixed XRV/ERV patterns can be observed, such as DR1
272 and DR3 insertions (“ERV like”) for ENTV-2 ON843769 and MZ931277, and DR3
273 deletion (“XRV like”) for PP682590. A group of sequences reported as ENTV-2
274 (KU258870 to KU258880) and integrated into the caprine ERV II.5 phylum does not
275 display the described typical ERV indels profile. Although only KU258870 lacks the
276 YXXM motif, they were still more similar to ERVs in DR1-DR2 and DR3-DR4 regions.
277 No accessible articles describing the conditions under which the sequences were
278 generated were available for any of these sequences except DQ838494 [16] and
279 MK164396 [25], which cluster with ERV II.5 sequences. In the cases of DQ838494
280 and MK164396, the associated primer sets were predicted to have very low to no
281 exogenous specificity for most primer pairs (S2 Table). This was likely the original
282 cause of the misidentification of these sequences. These genomes, which are most
283 probably ERVs rather than XRV or ERVs/XRV artificial chimeras, have been used as
284 reference sequences in other studies to develop molecular tools [20,21,25].

285 These analyses highlight the importance of selecting specific regions for detecting
286 XRVs without interfering with ERVs, which are present in multiple copies in ovine and
287 caprine genomes.

288 **Reliability of XRV detection in an ERV context**

289 Numerous published studies have reported the detection of JSRV and/or ENTV using
290 primers located in regions that are conserved between XRV and ERVs. This makes it
291 impossible to exclude the amplification of ERVs (S2 Table). To illustrate this challenge,
292 fourteen primer sets, defined as specific to JSRV or ENTVs were selected (S3 Table).
293 Specific primers for XRVs were included as specificity control [12].
294 The primers were defined by the following 3 criteria. The first one was the ability of
295 primers to detect either exogenous ENTV-2 (pairs 1-10) or JSRV (12-15) according to
296 the authors (S2-S3 Tables). The second criterion was the specificity of the primer
297 predicted through XRV and ERV sequences alignment. The third criterion relied on
298 the ability of primers to produce an amplicon of the expected size only from genomic
299 DNA extracted from JSRV- or ENTV-2-induced tumours and not from uninfected
300 ovine, caprine and human cells.

301 The theoretical specificity of the primers was predicted by aligning them with the XRV
302 and ERV sequences published from the public databases (S1-S2 Tables). Of the ten
303 pairs selected for ENTV-2, primer alignment identified seven pairs (#1, 3, 4, 5, 6, 7, 8)
304 that were non-specific for XRV, due to high homology with ERV sequences. Only three
305 pairs (2, 9, and 10) were specific for ENTV-2. Of the four pairs selected for JSRV, only
306 pair #14 was predicted to be specific for XRV. The primer pairs were then evaluated
307 for their ability to specifically amplify XRV, with amplicon generated only from gDNA
308 extracted from ENTV-2 or JSRV-induced tumours, and conversely to fail to produce

309 signal from gDNA extracted from IDO5, TiGEF or HeLa cells. Amplicons of the
310 expected size from all the gDNAs tested, whether extracted from tumours induced by
311 ENTV-2 as well as by JSRV or from uninfected ovine, caprine were generated for the
312 seven primer pairs # 1, 3, 4, 5, 6, 7, 8 predicted to be non-specific for ENTV-2. More
313 surprisingly, DNA was amplified with the primer pair #4 for human cells. Sequencing
314 confirmed this to be HERV-K (S4 Fig).

315 Similarly, of the four primer pairs tested for JSRV (12-15), only one pair (#15) was
316 found to be strictly specific, allowing amplification only from lung tumour gDNA; no
317 signal was observed from ENTV-2-induced nasal tumours or from uninfected human,
318 goat and sheep cells, confirming the perfect discrimination between ERVs and
319 exogenous retroviruses (Fig 4). In contrast, the other three primer pairs (#12, 13, 14)
320 produced an amplicon regardless of the gDNA or infection status. This confirms that
321 the published XRV primers are non-specific, enabling , indiscriminate amplification of
322 XRV and ERV sequences.

323 Amplicons obtained from DNA extracted from IDO5 sheep cells were sequenced.
324 Aligning the obtained sequences with the JSRV and ENTV-2 references sequences
325 (JSRV: FR2054 - PP707050; FR989 - PP707047; ENTV-2: FR3824 - PP669281) and
326 ERV II.5 confirmed their endogenous nature as evidenced by their high homology with
327 ERV II.5 family sequences (S5 Fig).

328 To unequivocally established their XRV or ERV nature, amplicons obtained with
329 primers containing the discriminating regions DR1 in *gag* (pair #1) and DR3 in *env*
330 (pair #13), were sequenced, and aligned against XRVs and ERV II.5 sequences (Fig
331 5). The primer pair #1 encompassed DR1 and amplified a region a 1130 bp U5/gag
332 region. The absence of the 12-nt deletion, present only in exogenous JSRV and ENTV

333 sequences, confirms that these primers failed to amplify XRVs (Fig 5A). The absence
334 of the YXXM motif, which is only present in oncogenic XRVs, confirms the amplification
335 of ERV sequences using primers pair #13 (Fig 5B). Phylogenetic reconstruction based
336 on amplicons generated with primer sets #1 (Fig 5C) and #13 (Fig 5D) from DNA of
337 non-infected ovine IDO5 cells or JSRV-induced cancers showed that the sequences
338 were closely related to the ERV II.5 *Ovis aries* consensus sequence (ERV II.5 *Oa*),
339 confirming the amplification of ERVs as part of the sheep genome. Symmetrically, the
340 sequences of amplicons obtained from DNA of non-infected caprine TiGEF cells or
341 ENTV-induced nasal tumours were highly similar to ERV II.5 *Capra hircus* (ERV II.5
342 *Ch*) consensus sequence (Fig 5C-D).

343 **Multiple tissues express endogenous sheep and goat retroviruses**

344 Several studies have been based on amplifying JSRV and ENTV from RNAs under
345 the assumption that only exogenous retroviruses are expressed. As RNA is an
346 appropriate target for retroviruses detection, the absence of ERV expression must be
347 confirmed to rule out dual expression of XRVs and ERVs at the same sites. We
348 analysed the expression of ERV II.5 using transcriptomic data from 60 tissues of the
349 Benz2616 ewe, the source of ARS-UI_Ramb_v2.0 reference ovine genome. Short
350 sequences or k-mers were defined in DR1 in *gag* and DR3 in *env* (Fig 6, Table 3 and
351 S1 Fig) from annotated ERV II.5 copies of the ARS-UI_Ramb_v2.0 genome [6]. JSRV
352 k-mers located in DR3 and DR4 were used to detect JSRV expression from RNA-Seq
353 data from JSRV-experimentally lambs and confirm the JSRV-negative status of the
354 Benz2616 ewe (Fig 6A, Table 3 and S1 Fig) [12,47]. For each of the RNA-Seq
355 datasets, sequences with specific ERV or JSRV k-mers were counted using the BBduk
356 tool. Conversely, ERV k-mers identified small quantities of ERV transcripts in most of

357 the tested tissues, with notably higher expression in the lung and kidney (particularly
358 the internal medullary part) and oviduct (Fig 6B). Sequences identified by the ERV k-
359 mers were mapped to the consensus sequence of ERV II.5 Oa, confirming their
360 endogenous origin proven the specificity of this approach to identify transcripts (Fig
361 6C). This demonstration in sheep shows that amplifying RNAs does not guarantee the
362 sole detection of JSRV or ENTV. If the primers are not highly XRV-specific, ERVs will
363 be amplified. Additionally, the absence of cellular DNA, even as traces, must be
364 carefully controlled to prevent the amplification of ERV copies, present in multiple
365 copies.

366 **The presence of JSRV and ENTV in arthropods cannot be not confirmed**

367 Six recent studies identified small ruminant XRV sequences [32–37] in insect vectors
368 using global approaches, such as metagenomics, transcriptomics and whole RNA-
369 Seq enriched for viral RNA. We used these datasets to challenge both our strategy for
370 specifically detecting XRVs in a context of coexistence of both XRV and ERVs in the
371 same individuals and the hypothesis of a potential vectorisation. Those approaches
372 rely on the BLASTn of the *de novo* assembled genomes to identify reference
373 sequences. Two studies made the reference matches available [33,36]. While
374 analysing the different datasets, we observed that identification of sequences
375 deposited in GenBank can be misleading. For example, MN564750.1 used as a
376 reference, is associated in GenBank with the organism "Enzootic nasal tumor virus 2"
377 (Taxonomy ID: 2913605), ID which also includes “Endogenous enzootic nasal tumor
378 virus”. MN564750.1 is identified as “Enzootic nasal tumor virus 2 isolate enENTV-FJ2
379 genomic sequence”. A scientist familiar with these viruses would recognise that the
380 name of the isolate ‘enENTV-FJ12’ corresponds to an endogenous (en) virus.

381 However, this is confusing, particularly in the absence of an associated article or any
382 further information about the origin in the deposit entry. We clearly identified ERV
383 sequences wrongly tagged as XRV in public databases such as MK210250 (Figs 2
384 and 3) or KT266728 (S6 Fig) used as a reference respectively for the detection of
385 ENTV-2 [33] and JSRV [36] which both lack the sequence coding for the YXXM motif.
386 Overall, this detailed initial analysis highlights the importance of carefully selecting the
387 reference sequence to exclude ERV sequences.

388 In their study [36], the authors identify both sequences of JSRV and ENTV-2 within
389 mosquitoes but only validated the presence of JSRV. While the amplification and
390 sequencing led to the identification of the sequence as “enJSRV”, the validation was
391 initially biased as primers (S2-S3 Tables) not specific for exogenous viruses were
392 designed based on the sequence initially matching with their reference. The non
393 specificity of primers #16 and #17 was confirmed using the same experimental design
394 with reference samples of DNA extracted from XRV-induced tumours, negative sheep
395 and goat controls, including DNA extracted from different mosquito species. The
396 amplification pattern was non-specific, with amplification in both uninfected and
397 infected samples (Fig 7A). The #16 and #17 primer sets amplified regions confirmed
398 to be of ERV origin by sequencing. The DR3 amplicon obtained with PCR #16 (Fig
399 7B) does not contain the insertion specific of XRVs nor the sequence coding for the
400 YXXM motif; the DR4 amplicon obtained with PCR #17 (Fig 7 A-C) has both the size
401 and sequence compatible with the ERV II.5 consensus. Neither the sequence nor the
402 accession number were made available for the “nearly complete sequence” of ENTV-
403 2 identified in mosquitoes. However the authors reported that the sequence was
404 closely related to the GDQY2017 Chinese strain (accession number MK164396), a

405 sequence that we identified as ERVs, questioning the XRV nature of the ENTV-2
406 sequence identified in this study (Figs 2 and 3) [36].

407 In a recent study [35], ENTV was amplified and sequenced from ticks collected in
408 Pakistan. A nearly full-length proviral genome (OR991120) was amplified (Fig 8A).
409 The sequence displayed 98% and 92 % overall nucleotide identity respectively with
410 the caprine and sheep ERV II.5 consensus, but less than 89% with JSRVs and ENTVs.
411 The DR3 and DR4 regions were unavailable. However, the DR1 and DR2 regions
412 displayed indels characteristics of ERVs (Fig 8C-D), as well as an ERV-specific partial
413 insertion in DR2 region. These findings suggest that the “provirus” was amplified from
414 a goat ERVs. The single set of primer targeting the *gag/ pol* was not predicted as
415 specific of XRVs, allowing an amplicon from both XRV infected and uninfected sources
416 (Fig 8B). The DR1 and DR2 sequencing of the amplicon confirmed the ERV nature of
417 the sequence (Fig 8C-D).

418 Three studies identified ENTV-2 in ticks isolated from cattle [32], JSRV from sheep
419 [37], ENTV-2 and JSRV from ticks isolated from yaks and ENTV-1 from Tibetan sheep
420 [34]. Unfortunately, neither data on the overall or DR regions coverage were available,
421 nor molecular validation of the presence of JSRV and ENTVs with XRV specific tools
422 were reported, only information regarding the abundance of the reads were available
423 for [32] and [34]. Therefore, the conclusions that can be drawn about the exogenous
424 nature of these sequences are limited, without further investigations. *A contrario* a
425 thorough checking based on the current recommendation can dispel doubts about the
426 nature of the strains detected by these high-throughput approaches. We therefore
427 collaborated with the authors to rework the preprint [33], which initially described
428 ENTV-2 within ticks. One of the tick metagenome samples displayed coverage of the

429 ENTV-2 provirus of over 80%, thus enabling to distinguish between ERV/XRV and
430 clearly identifying the DR1-3 regions associated with ERV. DR4 coverage was lower
431 but also suggested the detection of ERV. Phylogenetic analysis confirmed these
432 conclusions, as the consensus sequence clustered with the ERV sequences (S7 Fig).

433 Discussion

434 Accurate detection of β-retroviruses responsible for respiratory cancers in sheep and
435 goats is a critical prerequisite for controlling these diseases in livestock. This task is
436 complicated by the coexistence of exogenous retroviruses (JSRV and ENTV) with
437 numerous copies of endogenous retroviruses belonging to the ERV II.5 family in the
438 genome of sheep and goats. Molecular diagnostic approaches must therefore allow
439 unequivocal detection of XRV and therefore target regions that clearly discriminate
440 between ERVs and XRVs. Insufficient consideration of this constraint has directly
441 impacted on the specificity of molecular assays, the interpretation of phylogenetic
442 relationships, and the integrity of sequence annotations.

443 Early studies, based on a limited number of available sequences, suggested that
444 regions 1 to 4 could discriminate between ERVs and XRVs due to their low sequence
445 homology [5,19]. Building on these observations, we conducted an extensive analysis
446 integrating available XRV and ERV sequences with *in silico* annotation of ERV II.5 loci
447 in sheep and goat assemblies, that allowed us to define four discriminating segments
448 (DR1-DR4) that robustly discriminate exogenous from endogenous retroviruses. DR1
449 and DR2 localize in the 5‘gag region, DR3 in the 3’ terminus of *env*, and DR4 to U3.
450 Thus, they are prime targets for virus detection, reconstruction of phylogenetic
451 relationships and sequencing of near-full length of exogenous β-retroviruses. As we
452 established in this study from *in silico* analysis of primers reported in the literature,

453 targeting other regions with limited variation between ERVs and XRV forms poses a
454 risk of ERV detection.

455 As frequently reported in the literature, primers targeting conserved regions shared
456 between ERVs and XRVs are inherently non-specific and prone to amplifying
457 endogenous sequences. As we established in this study, most primers used to detect
458 JSRV or ENTV are unable to distinguish between exogenous and endogenous
459 sequences present in animal genomes and if used must be paired with XRV specific
460 primers. In a similar way, the growing use of high-throughput sequencing without
461 adequate discrimination criteria has amplified this problem. In recent years, this
462 methodological flaw has led to widespread mis-annotation of ERVs as infectious JSRV
463 or ENTV, distorting evolutionary analyses and misleading the development of
464 molecular tools.

465 Misinterpretation extends beyond host samples. Recent reports suggesting JSRV or
466 ENTV presence in ticks or mosquitoes may most likely reflect the detection of goat or
467 sheep DNA acquired from blood meal rather than true viral infection. As we have
468 shown there is no evidence that arthropods could serve as reservoirs or vectors for
469 JSRV and ENTV. Xenosurveillance may help to identify infection [36] but the clear
470 frontier with ERVs has to be drawn.

471 The distinction between XRV and ERV is often blurred in sequence databases, such
472 as GenBank, due to the presence of redundant and inconsistent taxonomic identifiers
473 that merge XRV and ERV. While a single taxonomic ID (69576) is defined for ENTV-
474 1, JSRV is classified within the taxonomic ID 47898 (recognised by the ICTV), which
475 also includes the “Endogenous Jaagsiekte retrovirus” (also classified with the *Ovis*
476 *aries*, taxonomic ID: 9940). A subclass “Exogenous Jaagsiekte sheep retrovirus”

477 (Taxonomic ID: 47898) is also reported, but this only contains 6 sequences. For ENTV-
478 2 the problem is even more complex. Recently, four taxonomic IDs coexisted for the
479 same virological entity, but taxonomic ID 2762664 (Endogenous enzootic nasal tumor)
480 and taxonomic ID 239365 (Enzootic nasal tumour virus of goats) were recently merged
481 with taxonomic ID 2913605 (Enzootic nasal tumor virus 2). The fusion of taxonomic ID
482 2762664 with 2913605 artificially grouped XRV and ERV sequences, whereas the
483 latter could be defined as part of the goat genome with identifier 9925 (*Capra hircus*).
484 Finally, the only singularity of taxonomic ID 2584748 is that it is defined by its common
485 name in US English, with “tumour” instead of “tumor”.

486 The classification of ERVs based on whether they are considered part of the virus or
487 the host is still a challenge, not limited to retroviruses in small ruminants. As an
488 example, porcine endogenous retroviruses are associated in GenBank either to their
489 host genome, *Sus scrofa* (Taxonomic ID 61673) or to the viral entity PERV (Porcine
490 endogenous retrovirus) (Taxonomic ID: 61673) subdivided in 11 subclasses with their
491 own taxonomic IDs. Brait and colleagues [38] suggest a rational solution to bypass
492 this challenge by adding a second taxonomic identifier. In this context, referring to ERV
493 II.5 with the taxonomy ID of the host in which they have been identified and a second
494 taxomic ID related to the closest XRV offers a pragmatic solution.

495 Finally, genome assembly approaches using short amplicon sequencing are ill-suited
496 to contexts where ERVs and XRVs coexist, as they cannot exclude chimeric
497 reconstructions. Long read sequencing technologies are therefore preferable an less
498 prone to generate *in silico* artefacts [12]. Regardless the approach used, the genetic
499 nature of the DR1 to DR4 discriminating regions is a minimal requirement for
500 classifying sequences as exogenous. Hallmarks features such as deletions in DR1/

501 DR2 regions as compared to ERV II.5, and the presence of the exogenous oncogenic
502 YXXM motif in DR3 (strictly absent in ERV) provide unambiguous evidence of
503 exogenous origin.

504 Although most of the complete sequences of JSRV and ENTV present in GenBank
505 possess these motifs indicating a probable exogenous origin, doubts remain about the
506 non-discriminatory regions such as the end of *gag*, *pro*, *pol* and the beginning of *env*.
507 For example, a recent study suggests possible recombination between different
508 strains of ENTV-2, precisely in the *pro* and *pol* regions, in which the absence of the
509 XRV/ ERV discriminatory region complicates amplification and the specific
510 identification of sequences of exogenous origin [21]. In this case, the reconstruction of
511 chimeric XRV/ ERV sequences cannot be ruled out.

512 The DR1-DR4 regions exhibit sufficient polymorphism to resolve viral clades and
513 geographic lineages. Updated diagnostic tools must therefore balance specificity for
514 exogenous viruses with coverage of known diversity.

515 In conclusion, we urge the stringent validation of the molecular tools used to detect
516 JSRV, ENTV-1, and ENTV-2. Primer pairs should include at least one primer located
517 in discriminating regions DR1–DR4, and the absence of amplification from DNA of
518 uninfected ovine and caprine cells specificity must be confirmed to establish
519 specificity. RNA-based detection alone is insufficient to exclude ERV contamination,
520 particularly in complex clinical samples and if we consider that ERV expression is
521 present in various tissues. Finally, assembled viral genomes lacking exogenous
522 hallmarks in DR1–DR4 using separated amplicons should be considered as potential
523 chimeric genomes, and not the reflect of the circulating strains. Only a rigorous,

524 discrimination-based framework will restore confidence in published sequences,
525 improve database reliability, and prevent further propagation of ERV/XRV confusion.

526 Acknowledgements

527 We thank Loïc Schwaller and Emmanuelle Lerat for the scientific advice; Vincent
528 Raquin and Barbara Viginier for providing us the mosquitoes. This work was
529 performed using the computing facilities of the CC LBBE/PRABI.

530 Funding

531 This research was funded by ANR, grant ANR-22-CE35-0002-01 and by INRAE "AAP-
532 Action incitative sur priorités SSD 2021-2025". BRV was the recipient of a VetAgro
533 Sup PhD grant. MV PhD fellowship was funded by ANR, grant ANR-22-CE35-0002-
534 01.

535 Data availability

536 This article used publicly available data. All the accession numbers from public
537 databases and sequences extracted from the articles were made available in the
538 supplementary tables 1 and 2.

539

540 Conflict of Interest

541 The authors declare that the research was conducted in the absence of any
542 commercial or financial relationships that could be construed as a potential conflict of
543 interest.

544

545 Author Contributions

BRV and CD realized the experiments. BRV, MV, JT, CL performed the analyses. BRV, VN, MV, CL, and JT designed the analysis. BRV, JT and CL wrote the manuscript. CL and JT conceived the project and secured the fundings. CL, JT, and VN supervised the work.

All participants contributed to the design and implementation of the research, participated to the interpretation of the data, and revised the manuscript.

Bibliography

1. Sigurdsson B. Adenomatosis of sheep's lungs; experimental transmission. Arch Gesamte Virusforsch. 1958;8: 51–58. doi:10.1007/BF01242312

2. De Las Heras M, Ortín A, Cousens C, Minguijón E, Sharp JM. Enzootic Nasal Adenocarcinoma of Sheep and Goats. In: Fan H, editor. Jaagsiekte Sheep Retrovirus and Lung Cancer. Berlin, Heidelberg: Springer Berlin Heidelberg; 2003. pp. 201–223. doi:10.1007/978-3-642-55638-8_8

3. Griffiths DJ, Martineau HM, Cousens C. Pathology and Pathogenesis of Ovine Pulmonary Adenocarcinoma. Journal of Comparative Pathology. 2010;142: 260–283. doi:10.1016/j.jcpa.2009.12.013

4. 江锦秀，张靖鹏，林裕胜，毛坤明，游伟，黄丽丽，et al. 山羊内源性鼻内肿瘤病毒enENTV-FJ1 株分享到：全基因组测序及生物信息学分析. Chinese Journal of Preventive Veterinary Medicine/Zhongguo Yufang Shouyi Xuebao. 2020;42.

5. Palmarini M, Hallwirth C, York D, Murgia C, De Oliveira T, Spencer T, et al. Molecular Cloning and Functional Analysis of Three Type D Endogenous Retroviruses of Sheep Reveal a Different Cell Tropism from That of the Highly Related Exogenous Jaagsiekte Sheep Retrovirus. J Virol. 2000;74: 8065–8076. doi:10.1128/JVI.74.17.8065-8076.2000

6. Verneret M, Leroux C, Faraut T, Navratil V, Lerat E, Turpin J. A genome-wide study of ruminants uncovers two endogenous retrovirus families recently active in goats. Mobile DNA. 2025;16: 4. doi:https://doi.org/10.1186/s13100-024-00337-6

7. Ortín A, Minguijón E, Dewar P, García M, Ferrer LM, Palmarini M, et al. Lack of a specific immune response against a recombinant capsid protein of Jaagsiekte sheep retrovirus in sheep and goats naturally affected by enzootic nasal tumour

578 or sheep pulmonary adenomatosis. Veterinary Immunology and
579 Immunopathology. 1998;61: 229–237. doi:10.1016/S0165-2427(97)00149-9

580 8. Hudachek SF, Kraft SL, Thamm DH, Bielefeldt-Ohmann H, DeMartini JC, Miller
581 AD, et al. Lung Tumor Development and Spontaneous Regression in Lambs
582 Coinfected With Jaagsiekte Sheep Retrovirus and Ovine Lentivirus. Vet Pathol.
583 2010;47: 148–162. doi:10.1177/0300985809352787

584 9. Caporale M, Martineau H, De las Heras M, Murgia C, Huang R, Centorame P, et
585 al. Host Species Barriers to Jaagsiekte Sheep Retrovirus Replication and
586 Carcinogenesis. J Virol. 2013;87: 10752–10762. doi:10.1128/JVI.01472-13

587 10. Walsh SR, Stinson KJ, Wootton SK. Seroconversion of sheep experimentally
588 infected with enzootic nasal tumor virus. BMC Res Notes. 2016;9: 15.
589 doi:10.1186/s13104-015-1824-2

590 11. Summers C, Neill W, Dewar P, Gonzalez L, van der Molen R, Norval M, et al.
591 Systemic immune responses following infection with Jaagsiekte sheep retrovirus
592 and in the terminal stages of ovine pulmonary adenocarcinoma. J Gen Virol.
593 2002;83: 1753–1757. doi:10.1099/0022-1317-83-7-1753

594 12. Riocreux-Verney B, Verneret M, Diesler R, Dolmazon C, Gineys B, Cadoré J-L,
595 et al. Association between genetic clades and cancer prevalence suggested by
596 French-wide study of oncogenic small ruminant β-retrovirus diversity. Front Cell
597 Infect Microbiol. 2024;14. doi:10.3389/fcimb.2024.1466333

598 13. York DF, Vigne R, Verwoerd DW, Querat G. Nucleotide sequence of the
599 jaagsiekte retrovirus, an exogenous and endogenous type D and B retrovirus of
600 sheep and goats. J Virol. 1992;66: 4930–4939. doi:10.1128/jvi.66.8.4930-
601 4939.1992

602 14. Palmarini M, Sharp JM, de las Heras M, Fan H. Jaagsiekte Sheep Retrovirus Is
603 Necessary and Sufficient To Induce a Contagious Lung Cancer in Sheep. J Virol.
604 1999;73: 6964–6972.

605 15. DeMartini JC, Bishop JV, Allen TE, Jassim FA, Sharp JM, de las Heras M, et al.
606 Jaagsiekte Sheep Retrovirus Proviral Clone JSRVJS7, Derived from the JS7
607 Lung Tumor Cell Line, Induces Ovine Pulmonary Carcinoma and Is Integrated
608 into the Surfactant Protein A Gene. J Virol. 2001;75: 4239–4246.
609 doi:10.1128/JVI.75.9.4239-4246.2001

610 16. Shuying L, Xueen M, Jingwei Q, Yu W. Cloning and sequencing analysis of the
611 complete genome of jaagsiekte sheep retrovirus inner mongolia strain. Zhongguo
612 Bing du xue= Virologica Sinica. 2006;21: 443–448.

613 17. Yadav EJ, Prabhakar YK, Shanmugam G, Vijayakumar P, Kumar MVJ,
614 Vijayakumar N, et al. Whole genome characterization of Jaagsiekte Sheep
615 Retrovirus isolated from OPA infected sheep (India) and structural assessment
616 (In-silico) of proteins of JSRV. The Microbe. 2024;3: 100068.
617 doi:10.1016/j.microb.2024.100068

618 18. Cousens C, Minguijon E, Dalziel RG, Ortin A, Garcia M, Park J, et al. Complete
619 Sequence of Enzootic Nasal Tumor Virus, a Retrovirus Associated with
620 Transmissible Intranasal Tumors of Sheep. Journal of Virology. 1999;73: 3986
621 LP – 3993. doi:10.1128/JVI.73.5.3986-3993.1999

622 19. Walsh SR, Linnerth-Petrik NM, Laporte AN, Menzies PI, Foster RA, Wootton SK.
623 Full-length genome sequence analysis of enzootic nasal tumor virus reveals an
624 unusually high degree of genetic stability. Virus research. 2010;151: 74–87.
625 doi:10.1016/j.virusres.2010.04.002

626 20. Ye C, Huang Q, Chen T, Jiang J, Hou F, Xu D, et al. First detection and genotypic
627 analysis of goat enzootic nasal tumor virus 2 in Chongqing, China. Arch Virol.
628 2019;164: 1647–1650. doi:10.1007/s00705-019-04211-2

629 21. Li Y, Niu J, Liu Y, Dai Y, Ni H, Wang J, et al. Genomic Sequencing and Analysis
630 of Enzootic Nasal Tumor Virus Type 2 Provides Evidence for Recombination
631 within the Prevalent Chinese Strains. Veterinary Sciences. 2024;11: 248.
632 doi:10.3390/vetsci11060248

633 22. Li P, Cao X, Wu J, Liu X, Mao S, Yuan L, et al. Detection and analysis of enzootic
634 nasal tumor virus 2 in China. J Vet Diagn Invest. 2025;37: 244–251.
635 doi:10.1177/10406387241310204

636 23. He Y, Zhang Q, Wang J, Zhou M, Fu M, Xu X. Full-length genome sequence
637 analysis of enzootic nasal tumor virus isolated from goats in China. Virol J.
638 2017;14: 141. doi:10.1186/s12985-017-0795-4

639 24. Ortín A, Cousens C, Minguijón E, Pascual Z, Villarreal MP de, Sharp JM, et al.
640 Characterization of enzootic nasal tumour virus of goats: complete sequence and
641 tissue distribution. The Journal of general virology. 2003;84: 2245–2252.
642 doi:10.1099/vir.0.19125-0

643 25. Zhai S-L, Lv D-H, Xu Z-H, Yu J-S, Wen X-H, Zhang H, et al. A Novel Enzootic
644 Nasal Tumor Virus Circulating in Goats from Southern China. Viruses. 2019;11:
645 956. doi:10.3390/v11100956

646 26. Li P, Yin H, Cao X, Lan X, Wu J, He J, et al. Development and Application of a
647 TaqMan-Based qPCR Assay for Detecting ENTV-2 in Goats. Genes (Basel).
648 2025;16: 529. doi:10.3390/genes16050529

649 27. Arnaud F, Caporale M, Varela M, Biek R, Chessa B, Alberti A, et al. A paradigm
650 for virus-host coevolution: sequential counter-adaptations between endogenous
651 and exogenous retroviruses. PLoS pathogens. 2007;3: e170.
652 doi:10.1371/journal.ppat.0030170

653 28. Zhang K, Kong H, Liu Y, Shang Y, Wu B, Liu X. Diagnosis and phylogenetic
654 analysis of ovine pulmonary adenocarcinoma in China. Virus Genes. 2014;48:
655 64–73. doi:10.1007/s11262-013-0988-x

656 29. Huang Q, Ye C, Chen T, Jiang J, Peng Y, Chen J, et al. EvaGreen-based real-
657 time PCR assay for sensitive detection of enzootic nasal tumor virus 2. Molecular
658 and Cellular Probes. 2019;44: 51–56. doi:10.1016/j.mcp.2019.02.003

659 30. Shi W, Jia S, Guan X, Yao X, Pan R, Huang X, et al. A survey of jaagsiekte sheep
660 retrovirus (JSRV) infection in sheep in the three northeastern provinces of China.
661 Arch Virol. 2021;166: 831–840. doi:10.1007/s00705-020-04919-6

662 31. Abass FA, Khudhair YI. Clinical, molecular, and pathological investigations of
663 ovine pulmonary adenocarcinoma in the middle of Iraq. Open Vet J. 2022;12:
664 264–272. doi:10.5455/OVJ.2022.v12.i2.15

665 32. Zhang X-L, Deng Y-P, Yang T, Li L-Y, Cheng T-Y, Liu G-H, et al. Metagenomics
666 of the midgut microbiome of Rhipicephalus microplus from China. Parasit
667 Vectors. 2022;15: 48. doi:10.1186/s13071-022-05161-6

668 33. Atim SA, Ashraf S, Niebel M, Ssekagiri A, Hardy MN, Shepherd JG, et al. Ticks;
669 a reservoir for virus emergence at the human-livestock interface in Uganda.
670 bioRxiv. 2023; 2023.03.10.532017. doi:10.1101/2023.03.10.532017

671 34. Zhang Y, Cheng T-Y, Liu G-H, Liu L, Duan D-Y. Metagenome reveals the midgut
672 microbial community of Haemaphysalis qinghaiensis ticks collected from yaks
673 and Tibetan sheep. Parasit Vectors. 2024;17: 370. doi:10.1186/s13071-024-
674 06442-y

675 35. Jamil A, Yu Z, Wang Y, Xin Q, Gao S, Abdul Wahab M, et al. Tick-borne
676 Rickettsia, Anaplasma, Theileria, and enzootic nasal tumor virus in ruminant,
677 PET, and poultry animals in Pakistan. Front Microbiol. 2024;15: 1359492.
678 doi:10.3389/fmicb.2024.1359492

679 36. Ndione MHD, Ndiaye EH, Dieng M, Diouf B, Sankhé S, Diallo D, et al. Mosquito-
680 based detection of retroviruses and arboviruses in Senegal: expanding the scope
681 of xenosurveillance. One Health Outlook. 2025;7: 32. doi:10.1186/s42522-025-
682 00155-7

683 37. Zhou K, Zhang L, Li W, Feng Q, Zhao W, He W, et al. Metagenomic sequencing
684 of epidemic pathogens in Ornithodoros lahorensis from southern Xinjiang.
685 Medycyna Weterynaryjna. 2026;82: 135–141. doi:10.21521/mw.7113

686 38. Brait N, Hackl T, Morel C, Exbrayat A, Gutierrez S, Lequime S. A tale of caution:
687 How endogenous viral elements affect virus discovery in transcriptomic data.
688 Virus Evolution. 2024;10: vead088. doi:10.1093/ve/vead088

689 39. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version
690 7: Improvements in Performance and Usability. Molecular Biology and Evolution.
691 2013;30: 772–780. doi:10.1093/molbev/mst010

692 40. Canty A, Ripley B, maintainer 1999--2022, support) author of parallel, fixes) ARB
693 (minor bug. boot: Bootstrap Functions (Originally by Angelo Canty for S). 2024.
694 Available: https://cran.r-project.org/web/packages/boot/index.html

41. Da Silva Teixeira MF, Lambert V, Mselli-Lakahl L, Chettab A, Chebloune Y, Mornex JF. Immortalization of caprine fibroblasts permissive for replication of small ruminant lentiviruses. Am J Vet Res. 1997;58: 579–584.

42. SCHERER WF, SYVERTON JT, GEY GO. Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. J Exp Med. 1953;97: 695–710. doi:10.1084/jem.97.5.695

43. Kent WJ. BLAT—The BLAST-Like Alignment Tool. Genome Res. 2002;12: 656–664. doi:10.1101/gr.229202

44. Salavati M, Caulton A, Clark R, Gazova I, Smith TPL, Worley KC, et al. Global Analysis of Transcription Start Sites in the New Ovine Reference Genome (Oar rambouillet v1.0). Front Genet. 2020;11: 580580. doi:10.3389/fgene.2020.580580

45. Davenport KM, Massa AT, Bhattarai S, McKay SD, Mousel MR, Herndon MK, et al. Characterizing Genetic Regulatory Elements in Ovine Tissues. Front Genet. 2021;12: 628849. doi:10.3389/fgene.2021.628849

46. Davenport KM, Bickhart DM, Worley K, Murali SC, Salavati M, Clark EL, et al. An improved ovine reference genome assembly to facilitate in-depth functional annotation of the sheep genome. GigaScience. 2022;11: giab096. doi:10.1093/gigascience/giab096

47. Karagianni AE, Vasoya D, Finlayson J, Martineau HM, Wood AR, Cousens C, et al. Transcriptional Response of Ovine Lung to Infection with Jaagsiekte Sheep Retrovirus. J Virol. 2019;93: e00876-19. doi:10.1128/JVI.00876-19

48. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170

49. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324

50. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29: 24–26. doi:10.1038/nbt.1754

51. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution. 2015;32: 268–274. doi:10.1093/molbev/msu300

Legends

733 **Fig 1. XRV are genetically close to ERVs.** The identity score was calculated from a
734 multiple alignment (MAFFT) of XRV sequences against a consensus sequence of ERV
735 II.5 Oa for JSRV and ENTV-1 and ERV II.5 Ch for ENTV-2. The average identity
736 percentage was calculated using a 100-base pair window and plotted in steps of 10.
737 A 95% confidence interval was represented for each virus, using a bootstrap-type
738 repetition approach. The 286 primers (grey triangles ▼), reported in publications (S2
739 Table) and used for the detection of XRV, were positioned with a tolerance of 4
740 mismatches. Primers tested in this study are identified by green triangles (▼). DR1-
741 DR4: discriminating regions 1-4.

742

743 **Fig 2. DR1-DR2 et DR3-DR4 are sufficient to classify JSRV, ENTVs and ERV**
744 **sequences.** Full-length sequences XRVs and ERVs available on GenBank by July
745 2025 and ERV copies annotated in *Capra hircus* and *Ovis aries* were used for
746 phylogenetic reconstructions. The maximum likelihood trees generated using 10,000
747 ultrafast bootstraps were rooted on the Bt ERV (Bos taurus ERV) family II.4 consensus
748 sequence [6] for **A.** the *pol* gene, **B.** the 360-nt DR1 - DR2 region and **C.** the 449-nt
749 DR3 to DR4 region. The scale bar indicates the number of substitutions per site. Black
750 triangles indicate sequences that have been referred to exogenous JSRV or ENTV
751 despite the absence of the YXXM oncogenic motif in the TM region.

752

753 **Fig 3. Example of wrongly annotated XRV retroviruses** Viral genome sequences
754 of JSRV (DQ838494 ; MZ931278) or ENTV (MK164396 ; MK210250) strains were
755 aligned (MAFFT) with viral genomic reference sequence of *Capra hircus* and *Ovis*
756 *aries* II.5 ERV consensus sequences [6], JSRV clade I (FR2054; PP707050.1), clade

757 II (FR989; PP707047.1), ENTV-1 (FR4403, PP669280.1) and ENTV-2 (FR3824,
758 PP669281.1) in **A.** DR1, **B.** DR2 , **C.** DR3 and **D.** DR4. Only nucleotides distinct from
759 the ERV II.5 Ch (top sequence) are indicated, identical nucleotides are marked by a
760 dot. The position on the alignment is indicated on top.

761

762 **Fig 4. Lack of specificity of selected XRV primers.** We analysed the 'XRV
763 specificity' of fourteen published primer pairs (01 to 10 and 12 to 15) according to the
764 authors' statement ('Publication'), our *in silico* analysis ('Prediction') and their profil of
765 detection in PCR ('Experiment'). The "XRV infected" reference panel consisted of
766 genomic DNAs extracted from a nasal tumour induced by ENTV-2 (FR3824, GenBank
767 accession PP669281.1) and lung cancers induced by clade I (FR2054; GenBank
768 accession PP707050.1) or clade II (FR989; GenBank accession PP707047.1) JSRV.
769 The "XRV non-infected" reference panel consisted of genomic DNAs extracted from
770 goat (TiGEF), sheep (IDO5) and human (HeLa) cell lines. Y: Yes; N: No. NTC: No
771 Template Control.

772

773 **Fig 5. Absence of exogenous DR1 and DR3 discriminatory regions in amplicons**
774 **obtained with primer pairs #01 (ENTV-2) and #13 (JSRV).** Amplicons obtained with
775 primer pairs #1 (ENTV-2) (**A** and **C**) and #13 (JSRV) (**B** and **D**) were sequenced and
776 aligned (MAFFT) with a reference sequence of ENTV-2 (FR3824, GenBank accession
777 PP669281.1), JSRV clade I (FR2054; GenBank accession PP707050.1) & II (FR989;
778 GenBank accession PP707047.1) and the consensus sequences of ERV II.5 Oa and
779 ERV II.5 Ch [4]. Only nucleotides distinct from the ERV II.5 *Ch* consensus sequence
780 (on top) have been indicated, identical nucleotides are marked by a dot. Maximum

781 likelihood (ML) trees were reconstructed using PhyML from **C.** 826-nt of the U5/gag or
782 **D.** 864-nt of the *env* amplicons with ENTV1-FR4403 sequence (GenBank accession
783 PP707036) as an outgroup for rooting. Bootstrap values (1000 replicates) are
784 expressed as a percentage and indicated at the node level. The scale bar indicates
785 the number of substitutions per site.

786

787 **Fig 6. Expression of ERV family II.5 in tissues. A**. K-mers localization on the JSRV
788 provirus. kERV-DR1 and kERV-D3 are specific to ERV II.5 Oa and located in the DR1
789 (*gag*) and DR3 (*env*) discriminating regions, respectively. kJSRV-DR3 and JSRV-DR4
790 are specific to JSRV and located in DR3 (*env*) and DR4 (U3). **B**. The number of reads
791 containing one of the four selected k-mers was calculated for each tissue or organ.
792 The samples were grouped by physiological system and visualised using a colour
793 code. JSRV-infected and uninfected control sample [47] were added and labelled as
794 "Lung_JSRV_infected" for lung RNA-Seq data of JSRV-JS21-infected lamb or
795 "Lung_JSRV_uninfected" for uninfected control lambs. **C.** Reads containing kERV-
796 DR1 and kERV-DR3 k-mers were mapped on the ERV II.5 Oa consensus sequence
797 and visualised on IGV.

798

799 **Fig 7. Molecular detection of JSRV and ENTVs in mosquitoes**. **A.** Evaluation of
800 the specificity of primer pairs # 16 and #17 on genomic DNA extracted from JSRV or
801 ENTV tumours, from uninfected *sheep (IDO5) and goat (TIGEF)* cell lines, and from
802 *Aedes aegypti (Ae)* and *Culex Quique/asciatus (Cq/a*). The amplicons obtained with
803 primer pairs #16 (**B**) and #17 (**C**) were sequenced and aligned against ERV II.5, ENTV
804 and JSRV reference sequences. Only nucleotides distinct from the ERV II.5 *Ch*

805 consensus sequence (on top) have been indicated, identical nucleotides are marked
806 by a dot.

807

808 **Fig 8. Molecular detection of ENTV in ticks. A**. The partial OR991120 proviral
809 sequence, missing the LTRs and the DR3 region, reported in ticks [35] was aligned
810 with II.5 ERV Oa and Ch consensus sequences [6], JSRV clade I (FR2054; GenBank
811 accession PP707050.1), clade II (FR989; GenBank accession PP707047.1), ENTV-
812 1(FR4403, GenBank accession PP669280.1) and ENTV-2 (FR3824, GenBank
813 accession PP669281.1). The black and grey bars correspond respectively to the
814 differential or identical nucleotides. **B**. Evaluation of the primer specificity. The
815 amplicon #11 was sequenced and aligned against proviral sequences of JSRV, ENTV
816 and ERV II.5. **C.** Alignment of the DR1 region. **D.** Alignment of the DR2 region.

817

818 **S1 Table. Accession number list of all full-length and near full sequences of**
819 **JSRV, ENTV-1, ENTV-2 and ERVs used in the study.**

820

821 **S2 Table. Description of the 286 JSRV and ENTV published primers used in the**
822 **study**.
823 The 'Publication specificity' column relies on the authors' statement of whether the
824 primer pair was specific ("S") or not ("NS") for JSRV or ENTVs in the corresponding
825 article(s). The 'Predicted XRV' column is based on *in silico* control of the specificity for
826 XRV (Yes) or non-XRV (No). "ND" indicates that the XRV or ERV specificity could not
827 be determined.

828

829 **S3 Table. Sets of primers tested for XRV detection and PCR conditions.**

830

831 **S1 Fig. Characterisation of the k-mers used to detect ERV II.5 and JSRV reads.**

832 The k-mers were identified from multiple alignments (MAFFT) of JSRV strains present
833 in GenBank, whose exogenous nature was validated, and the copies of ERV II.5 Oa
834 annotated in the ARS-UI_Ramb_v2.0 genome which contains the DR1 region (A), the
835 DR3 region (B) or the DR4 region (C). Mismatches against the chosen k-mers are
836 highlighted.

837

838 **S2 Fig. Individual identity scores per XRV sequence**. The individual identity scores
839 for all the XRV sequences were calculated using a multiple alignment (MAFFT) of the
840 XRV sequences against an ERV II.5 Oa consensus for JSRV (A) and ENTV-1 (B), and
841 against an ERV II.5 Ch consensus for ENTV-2 (C). The average identity percentage
842 was calculated using a 100-base pair window and plotted in steps of 10.

843

844 **S3 Fig. DR1-DR4 discriminating regions between endogenous retroviruses and**
845 **JSRV & ENTVs.** Viral genomic sequence of ERV II.5 Oa and Ch consensus
846 sequences [6] and JSRV clade I (FR2054; PP707050.1), clade II (FR989;
847 PP707047.1), ENTV-1 (FR4403, PP669280.1) and ENTV-2 (FR3824, PP669281.1)
848 were aligned using MAFFT. **A.** DR1 and **B**. DR2 in the *gag* gene. For DR2 the
849 percentage of each nucleotide (A,T,G,C) per position for all the II.5 copies in the Goat
850 (ARS1.2) and Sheep (ARS-UI_Ramb_v2.0) references was plotted in the deletion
851 area. **C.** DR3 in *env* and **D.** DR4 in U3. Only distinct nucleotides with the first sequence

852 of the alignment (ERV II.5 Oa) have been indicated, identical nucleotides are marked
853 by a dot. Numbering corresponds to the position on the alignment.

854

855 **S4 Fig. HERV-k sequence obtained with primer pair #4, supposedly specific of**
856 **ENTV-2 from genomic DNA of human HeLa cells.** Alignment (MAFFT) of the
857 sequence of PCR fragment #4 obtained by the Sanger method with the closest β-
858 HERV (HERVK-13) identified by BLAT.

859

860 **S5 Fig. Sequencing amplicons obtained from ovine IDO5 cells.** MAFFT alignment
861 of the amplicon sequence with ERV II.5 Oa consensus sequence. **A.** Primers set #3.
862 **B.** Primers set #4. **C.** Primers set #5. **D.** Primers set #6. **E.** Primers set #7. **F.** Primers
863 set #8. **G.** Primers set #14.

864

865 **S6 Fig. KT266728.1 Env sequence does not contain the oncogenic « YXXM »**
866 **motif.** The JSRV viral sequences KT266728.1, reported as the best match in [36], was
867 aligned (MAFFT) with the ERV II.5 Oa consensus sequence [6] as well as with JSRV
868 clade I (FR2054; PP707050.1) and clade II (FR989; PP707047.1) sequences. The
869 nucleotides encoding YXXM are not present in KT266728.1; this strain has a HKXM
870 predicted sequence that is common to ERV II.5.

871

872 **S7 Fig. Verification of metagenomic sequences identified as exogenous.** The
873 consensus sequence (TickP118) obtained from one tick metagenome from [33] and
874 the viral genomic sequence of ERV II.5 Oa and Ch consensus sequences [6] and

875 JSRV clade I (FR2054; PP707050.1), clade II (FR989; PP707047.1), ENTV-1
876 (FR4403, PP669280.1) and ENTV-2 (FR3824, PP669281.1) were aligned using
877 MAFFT in A. DR1, B. DR2 , C. DR3 and D. DR4. Only distinct nucleotides with the first
878 sequence of the alignment (ERV II.5 Ch) have been indicated, identical nucleotides
879 are marked by a dot. Numbering corresponds to the position on the alignment. Full-
880 length sequences of ENTV-2 and enENTV available on GenBank and the consensus
881 sequence obtained from one tick metagenome from [33] were used for phylogenetic
882 reconstructions. Maximum likelihood (ML) trees were reconstructed using PhyML **F.**
883 the 449-nt DR3 to DR4 region. Bootstrap values (1000 replicates) are expressed as a
884 percentage and indicated at the node level. To help the visualisation the tree was
885 rooted on the ERV branch. The scale bar indicates the number of substitutions per
886 site.

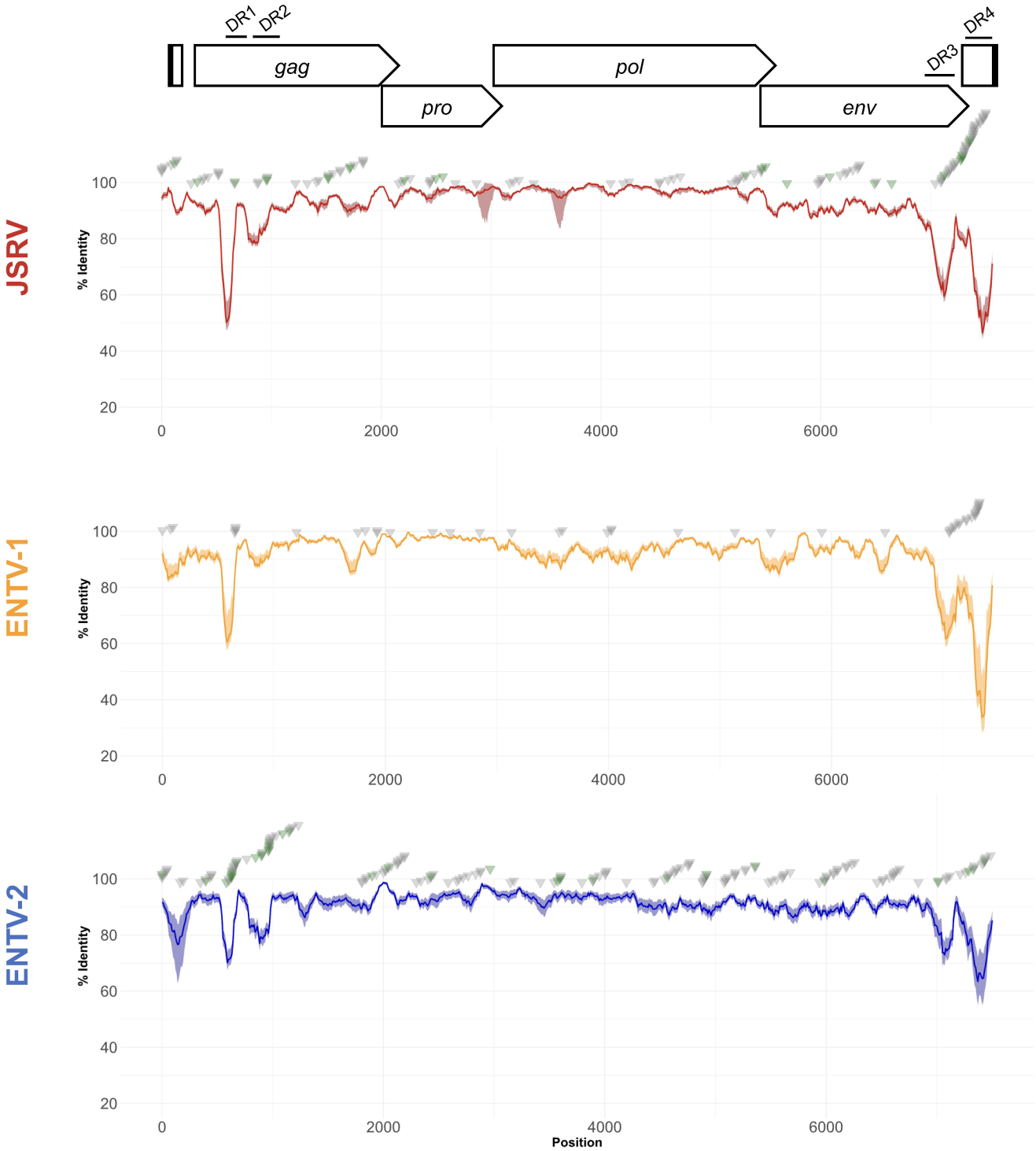
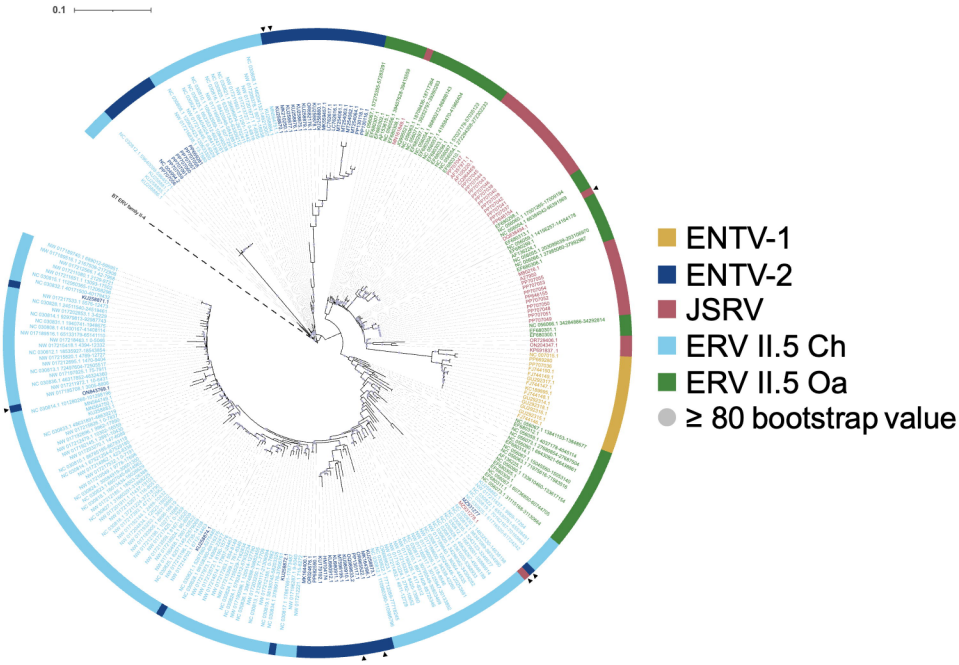
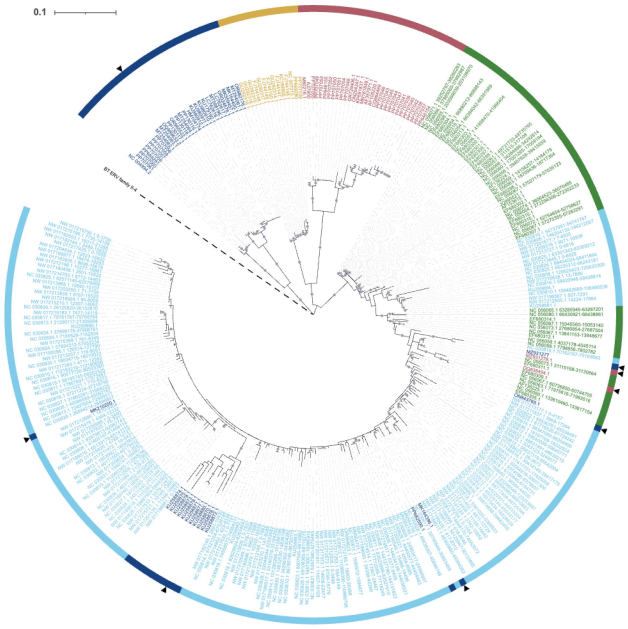

DR1
DR2
DR3
DR4
gag
pro
pol
env
JSRV
ENTV-1
ENTV-2
% Identity
100
80
60
40
20
0
2000
4000
6000
Position

Figure 1

A. *pol*



B. DR1 to DR2



C. DR3 to DR4

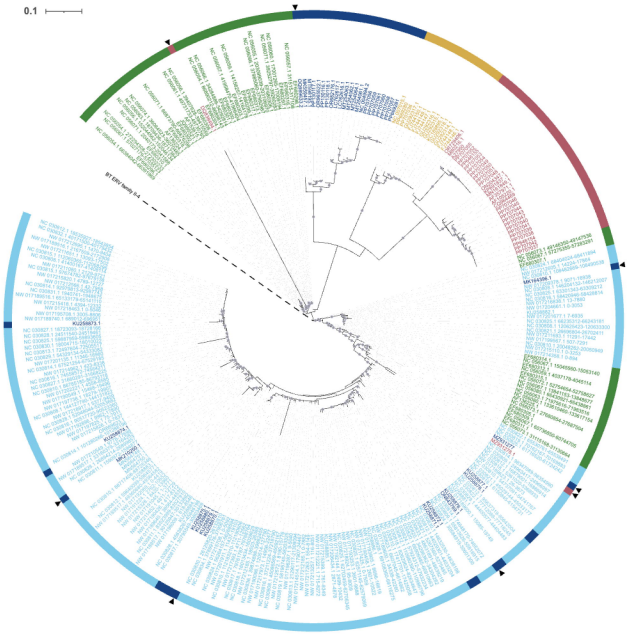


Figure 2

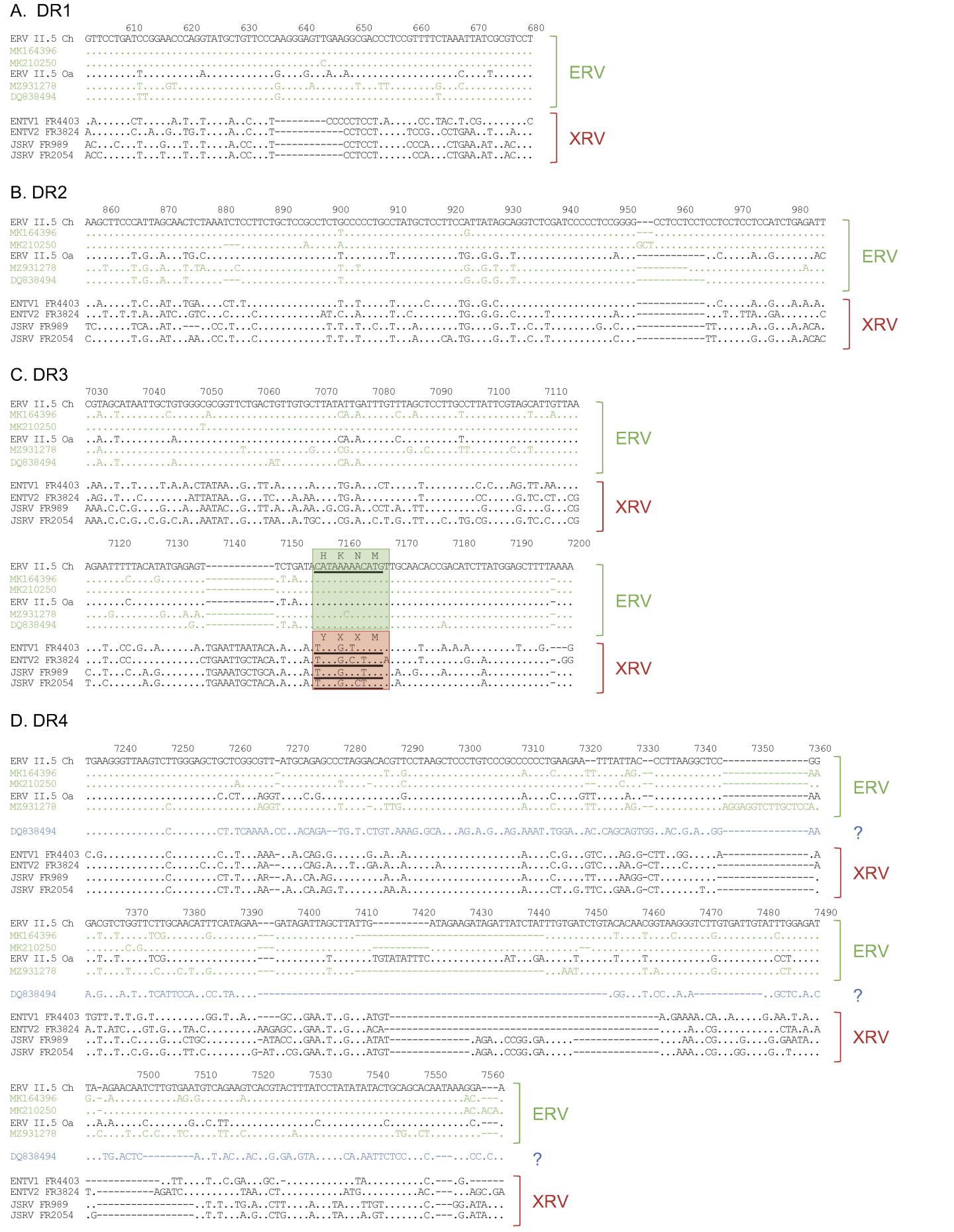


Figure 3

	PCR primer pairs #	Target	XRV specificity		
			Publication	Prediction	Experiment
ENTV-2	01	U5/ *gag*	Y	N	N
	02	*gag*	Y	Y	Y
	03	*gag/ pro*	Y	N	N
	04	*gag/ pol*	Y	N	N
	05	*pro/ pol*	Y	N	N
	06	*pol*	Y	N	N
	07	*pol*	Y	N	N
	08	*env*	Y	N	N
	09	*env*	Y	Y	Y
	10	LTR	Y	Y	Y
JSRV	12	LTR	Y	N	N
	13	*env*	Y	N	N
	14	*env*	Y	N	N
	15	LTR	Y	Y	Y

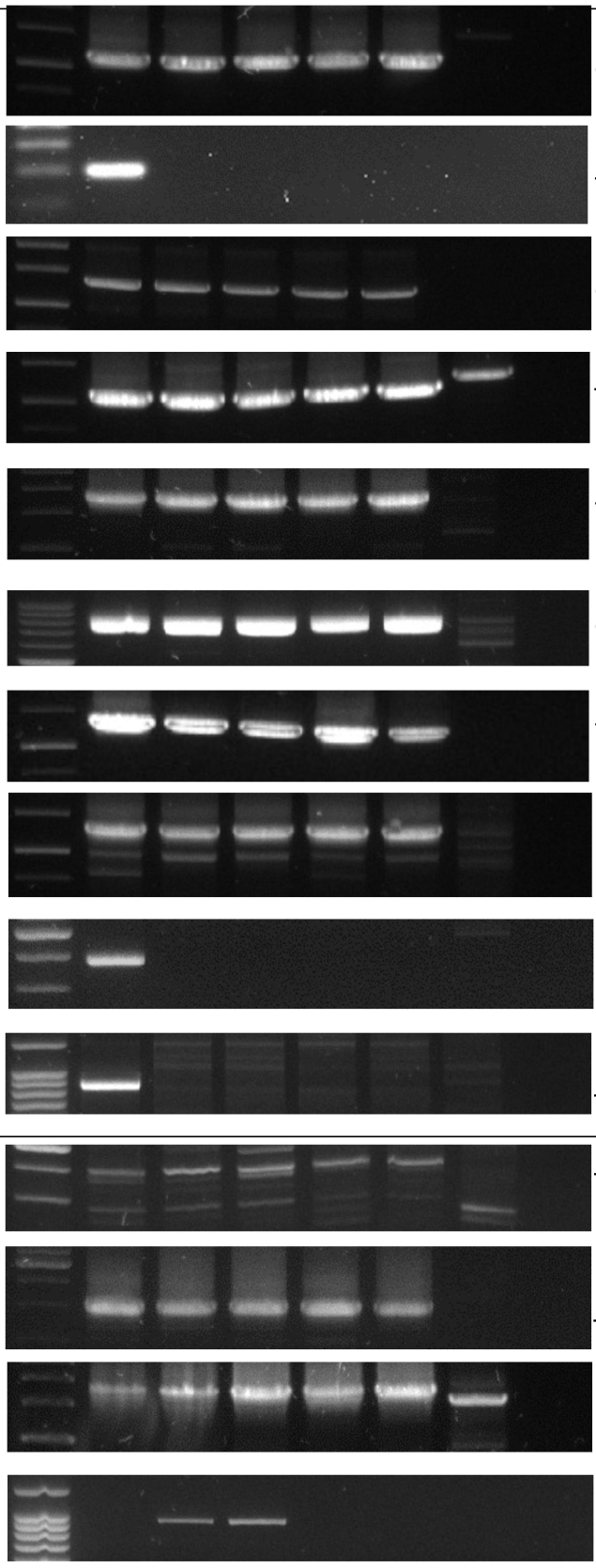


Figure 4

A. Alignment in the U5/gag DR1 region (amplicon PCR #01)

```
                        930       940       950       960       970       980       990      1000
ERV II.5 Ch         GTTCCTGATCCGGAACCCAGGTATGCTGTTCCCAAGGGAGTTGAAGGCGACCCTCCGTTTTCTAAATTATCGCGTCCT
Amplicon ENTV2      ..........................................................................
Amplicon TIGEF      ..........................................................................
ERV II.5 Oa         .........T..........A............G....G...A..A...................C....T.......
Amplicon JSRV-I     .........T.......................G....G...A..A..........T........C....T.......
Amplicon JSRV-II.   .........T.......................G....G...A..A..........T........T....T..A....
Amplicon IDO5       .........T.......................G....G...A..A...................C....T..A....

ENTV2 FR3824        A.......C..A..G..TG.T....A..C...T------------CCTCCT.....TCCG..CCTGAA..T...A...
JSRV FR2054         ACC......T...T...T..T....A.CC...T------------CCTCCT......CCA...CTGAA.AT..AC...
JSRV FR989          AC...C...T...G...T..T....A.CC...T------------CCTCCT.....CCCA...CTGAA.AT..AC...
```

ERV

XRV

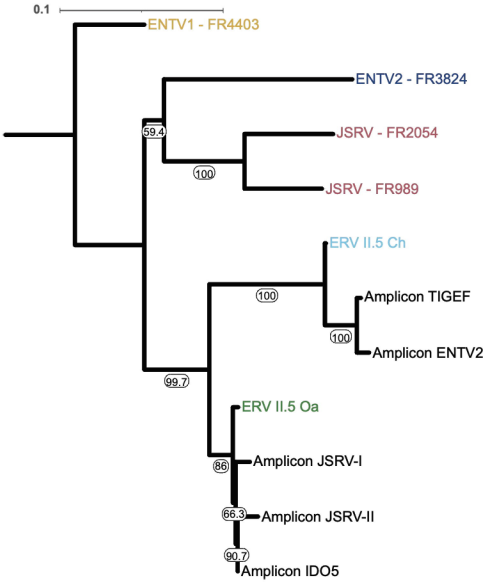
B. Alignment in the *env* DR3 region (amplicon PCR #13)

```
                     7360      7370      7380      7390      7400      7410      7420      7430      7440
ERV  II.5 Ch      CGTAGCATAATTGCTGTGGGCGCGGTTCTGACTGTTGTGCTTATATTGATTTGTTTAGCTCCTTGCCTTATTCGTAGCATTGTTAA
Amplicon ENTV2    ..A..T...............A.......................CA.A......C..A.......T...........T........
Amplicon TIGEF    ..A..T...............A.......................CA.A......C..A.......T...........T........
ERV  II.5 Oa      ..A..T.........A............................CA.A......C..........T....................
Amplicon JSRV-I   ..A..T.........A............................CA.A................T...........C.........
Amplicon JSRV-II  ..A..T.........A............................CA.A.......................G...C.........
Amplicon IDO5     ..A..T.........A............................CA.A................T....................

ENTV2 FR3824      .AG..T...C........ATTATAA..G...TC...A.AA....TG.A..........T.........CC.....G.TC.CT..CG
JSRV FR2054       AAA.C.C.G..C.G.C.A..AATAT..G...TAA..A.TGC...CG.A..C.T.G..TT...C..TG.CG.....G.TC.C...CG
JSRV FR989        AAA.C.C.G....G...A..AATAC..G..TT.A..A.AA..G.CG.A..CCT.A..TT..........G.....G....G...CG

                          7450      7460      7470      7480      7490      7500      7510      7520
                                                          H  K  N  M
ERV  II.5 Ch      AGAATTTTTACATATGAGAGT------------TCTGATACATAAAAACATGTTGCAACACCGACATCTTATGGAGCTTTTAAAA
Amplicon ENTV2    .......C....G........------------.T.A...............................................-...
Amplicon TIGEF    .......C....G........------------.T.A...............................................-...
ERV  II.5 Oa      .......C.............------------.T.A...............................................-...
Amplicon JSRV-I   ......................------------.T.A..........C.........C....AC.................-...
Amplicon JSRV-II  ......................------------.T.A..........C.........C....AC.......G.........-...
Amplicon IDO5     ......................------------.T.A..........C.............AC.................-...

                                                          Y  X  X  M
ENTV2 FR3824      ...T..CC............CTGAATTGCTACA.T...A.T...G.C.T...A.....T.......G..A...........-.GG
JSRV FR2054       T..C......A.G........TGAAATGCTACA.A...A.T...G..CT.....A..............A...........-...
JSRV FR989        C..T...C..A.G........TGAAATGCTGCA.A...A.T...G...T.....A..G....A......A...........-...
```

ERV

XRV

C. ENTV2 U5/*gag* (PCR #01)



D. JSRV *env* (PCR #13)

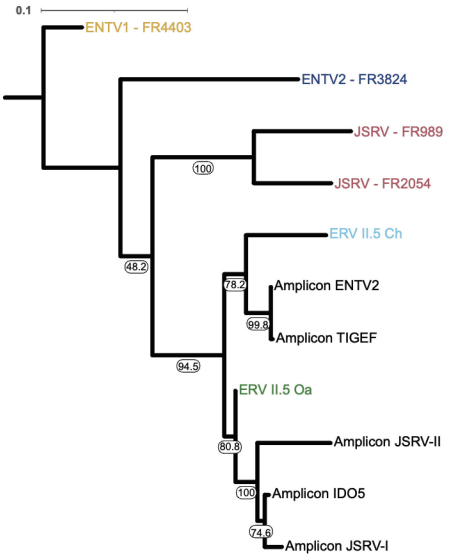
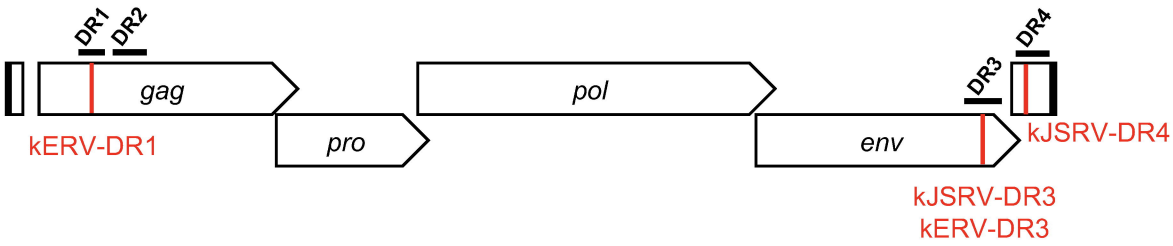
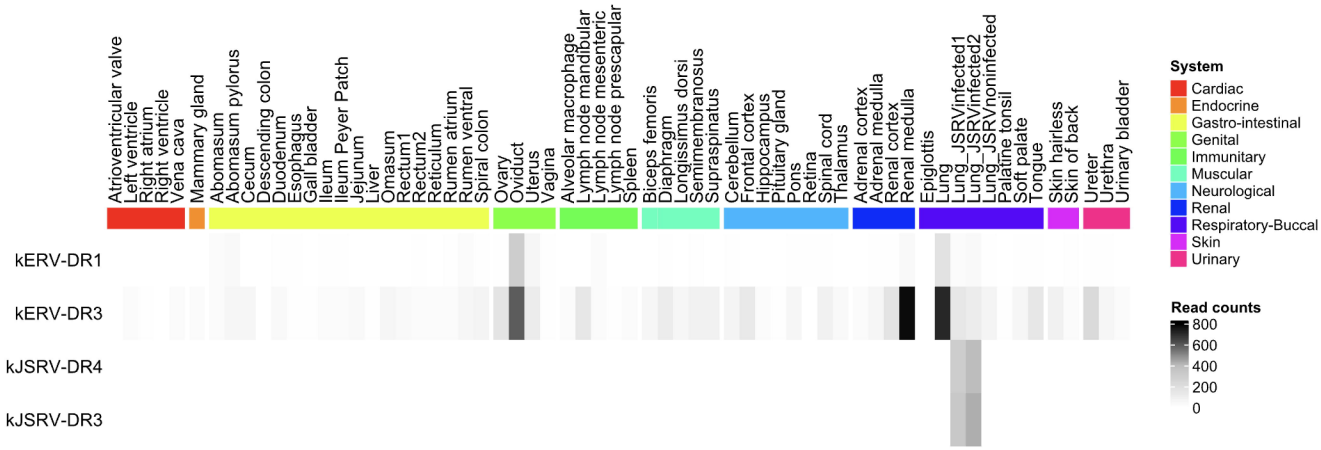


Figure 5

A. Localization of kmers



B. Expression of ERV II-5 in organs and tissues



C. Mapping of selected reads on ERV II.5 Oa consensus sequence

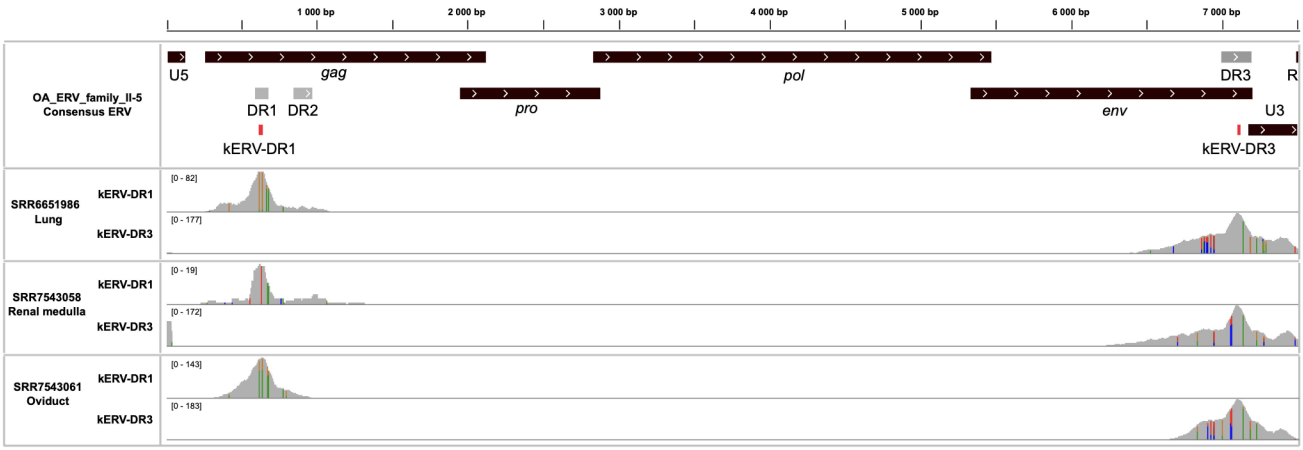
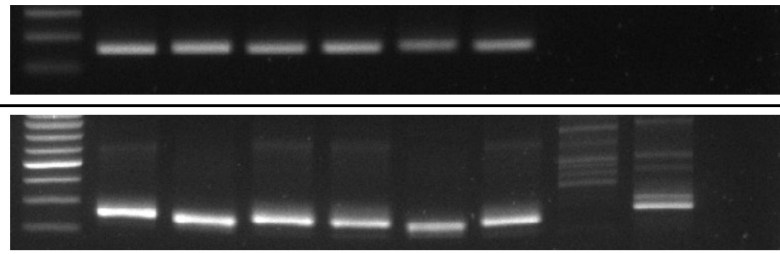


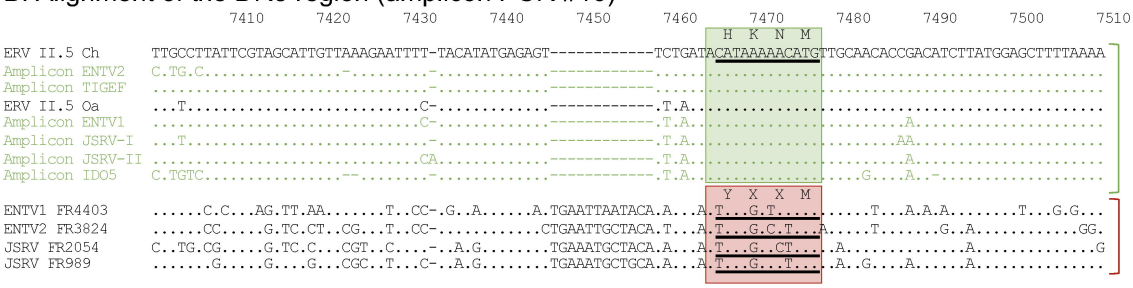
Figure 6

A. Primer specifcity

PCR primer pair #	Target	XRV specificity: Publication	Prediction	Experiment
16	*env*	Y	N	N
17	*U3/ R*	Y	N	N



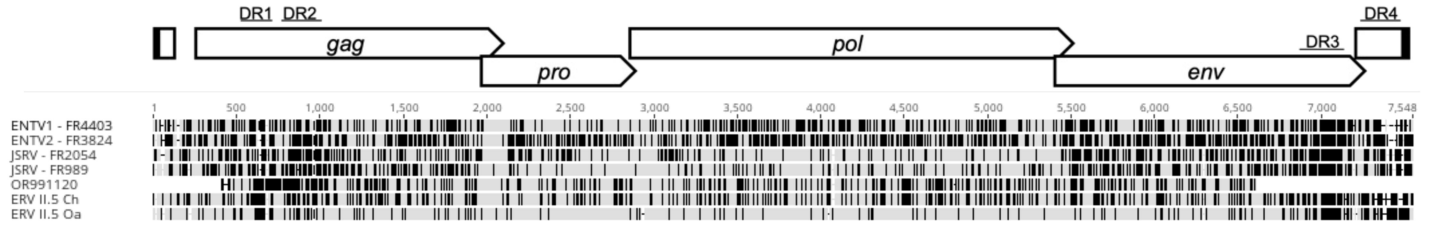
B. Alignment of the DR3 region (amplicon PCR #16)



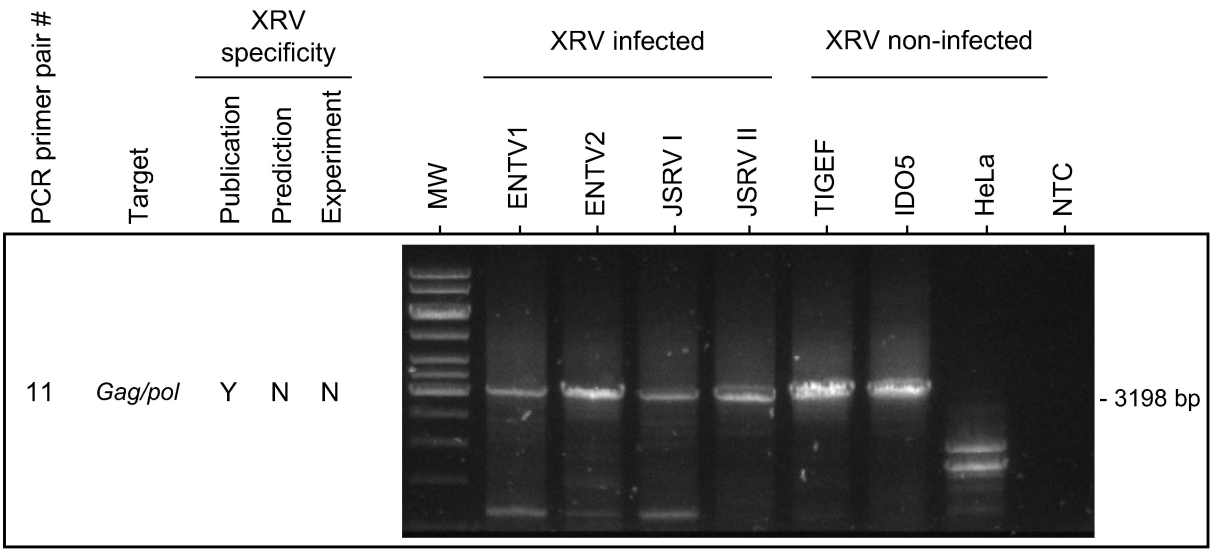
C. Alignment of the DR4 region (amplicon PCR #17)

Figure 7

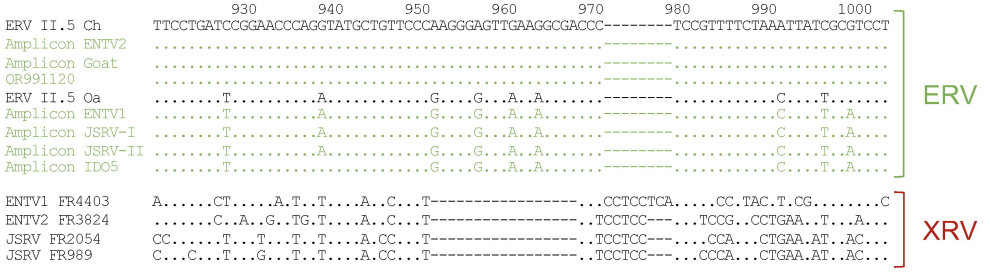
A. Alignment of OR991120 with reference sequences



B. Primer specificity



C. Alignment in the DR1 region (amplicon PCR #11)



D. Alignment in the DR2 region (amplicon PCR #11)

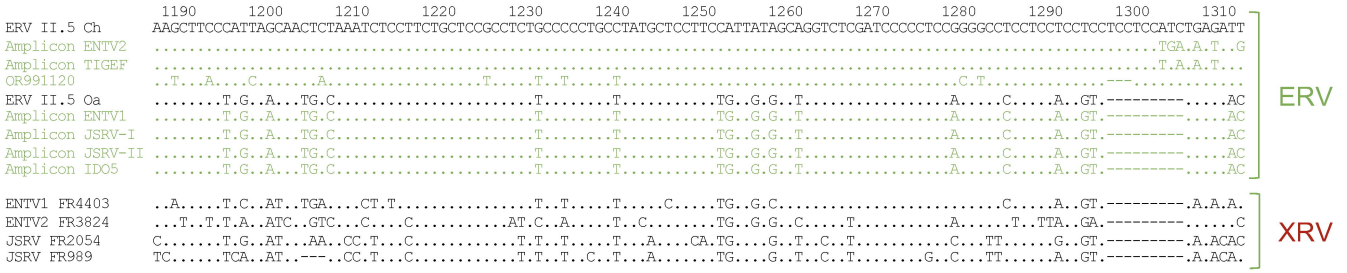


Figure 8