Virus Evolution, 2026, **12(1)**, veag028
https://doi.org/10.1093/ve/veag028
Advance Access Publication 8 May 2026
Research Article

Unravelling viral identity: avoiding the trap of endogenous sequences for viral surveillance of small ruminant oncogenic retroviruses

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

[1]Université Lyon 1, INRAE, EPHE, Université PSL, IVPC, UMR 754, 50 Avenue Tony Garnier, 69007 Lyon, France
[2]Université Lyon 1, LBBE, UMR 5558, CNRS, VetAgro Sup, 43 bd du 11 novembre 1918, 69622 Villeurbanne, France
[3]MRC-University of Glasgow Centre for Virus Research (CVR), 464 Bearsden Road, G13 2QJ Glasgow, United Kingdom
[4]MRC/UVRI and LSHTM Uganda Research Unit, Plot 51-59, Nakiwogo Road, Entebbe, Uganda
[5]Ministry of Agriculture, Animal Industry and Fisheries, P.O. Box 102, Entebbe, Uganda
[6]PRABI, Pôle Rhône-Alpes Bioinformatics Center, Université Lyon 1, 16 rue Raphael Dubois, 69622 Villeurbanne, France
[7]European Virus Bioinformatics Center, Leutragraben 1, 07743 Jena, Germany
*Corresponding authors. Université Lyon 1, INRAE, EPHE, Université PSL, IVPC, UMR 754, 50 Avenue Tony Garnier, 69007 Lyon, France, Caroline Leroux, E-mail: caroline.leroux@inrae.fr; Jocelyn Turpin, E-mail: jocelyn.turpin@inrae.fr.
†Caroline Leroux and Jocelyn Turpin contributed equally to this work and share last authorship.

Abstract

Small ruminants (sheep and goats) are one of the few mammals in which an exogenous retrovirus (XRV) and closely related endogenous retroviral (ERV) elements coexist within the same host genome. The betaretroviruses Jaagsiekte sheep retrovirus (JSRV) and Enzootic Nasal Tumour 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 resulted in the population of databases with dubious entries, fostering erroneous hypotheses such as vector-borne transmission of JSRV and ENTV. To address this issue, we propose a concise set of criteria for assay design, validation, and database annotation, emphasizing DR targeting, specificity testing against endogenous templates, and transparent reporting.

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

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

© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

1. Introduction

Cancers induced by the exogenous retroviruses (XRVs), Jaagsiekte Sheep RetroVirus (JSRV) and Enzootic Nasal Tumour Virus (ENTV), have been documented in all regions where sheep and goats are reared, except in Australia, New Zealand, the Falkland Islands and Iceland, where JSRV has been eradicated (Sigurdsson 1958, De Las Heras et al. 2003, Griffiths et al. 2010). Multiple copies of endogenous retrovirus (ERV) family II.5, previously named enJSRV and enENTV (Palmarini et al. 2000, 江锦秀 et al. 2020), are present in goats (ERV II.5 Ch for *Capra hircus*) and sheep (ERV II.5 Oa for *Ovis aries*), with an average of 223 and 98 *loci* respectively (Verneret et al. 2025). The presence of whole copies with little divergence and with preserved open reading frames (*gag, pro, pol,* and *env*) suggests the recent activity of ERV II.5 Ch in goats (Verneret et al. 2025). Despite their close genetic relationship with domestic sheep and goats XRVs, ERVs are not implicated in the development of respiratory cancers in domestic small ruminants. Therefore, the XRVs JSRV and ENTV remain the only known aetiological agents of respiratory cancers in domestic small ruminants.

Developing robust virus detection tools is essential for managing the infectious risk posed by these incurable, fatal, and transmissible cancers that threaten small-ruminant farming. Serological detection of antibodies is a tool widely used in both veterinary and human medicine. Apart from the fact that it only allows indirect detection of infection through the demonstration of antibody production, it requires a detectable and stable humoral response. In a context of multiple ERV copies impairing the development of a XRV-specific humoral response, antibody detection in JSRV or ENTV-infected animals is not a reliable tool to establish the infectious status (Ortín et al. 1998, Hudachek et al. 2010, Caporale et al. 2013, Walsh et al. 2016).

Therefore, infection can only be detected *ante or post mortem* using molecular approaches, i.e. Polymerase Chain Reaction (PCR) and/or Reverse Transcription Polymerase Chain Reaction (RT-PCR) (Ortín et al. 1998, Summers et al. 2002). The genetic diversity of JSRV and ENTV is only partially known, and before our recent study reporting on 29 JSRV and 24 ENTV partial and full proviral sequences (Riocreux-Verney et al. 2024), only a few dozen complete sequences were available in the databases, i.e. nine complete sequences of JSRV isolated in Africa, North America, Europe, and Asia (York et al. 1992, Palmarini et al. 1999, DeMartini et al. 2001, Shuying et al. 2006, Yadav et al. 2024), and 50 complete or near-full-length ENTV genomes isolated in Canada and Europe for ENTV-1 (Cousens et al. 1999, Walsh et al. 2010) and in Europe and Asia for ENTV-2 (Ortín et al. 2003, He et al. 2017, Ye et al. 2019, Zhai et al. 2019, Li et al. 2024, 2025a, 2025b), including sequences not associated with accessible publications or unpublished.

The widespread use of high-throughput sequencing tools has made pathogen sequences widely available. In theory, detecting viral or proviral genomes using targeted molecular approaches is straightforward, given the multiple bioinformatic tools available for designing primers and probes. However, when dealing with the coexistence of genetically closely related ERV family II.5 and XRVs, it is crucial to ensure the strict XRV specificity of the primers or probes. To optimize the molecular XRV detection tools in regards to published studies, we reviewed multiple reports, focusing on the molecular detection and sequencing tools used and their relevance in the context of the coexistence of XRVs with ERV II.5 *loci* (Verneret et al. 2025). A recent study of all available sequences in databases allowed us to identified sequences wrongly identified as exogenous JSRV or ENTV despite the absence of the YXXM motif in the region of the *env* gene encoding the intracytoplasmic tail (CT), suggesting their endogenous origin (Riocreux-Verney et al. 2024). The YXXM motif is an essential oncogenic determinant and a genetic marker of exogenous sequences, which is strictly absent from all the ERV II.5 Oa and Ch sequences identified up to date (Cousens et al. 1999, Arnaud et al. 2007, Riocreux-Verney et al. 2024, Verneret et al. 2025). On the other hand, some studies have relied on RNAs to specifically detect XRVs (Zhang et al. 2014, He et al. 2017, Huang et al. 2019, Ye et al. 2019, Shi et al. 2021, Abass and Khudhair 2022). This approach assumes that ERVs are not expressed in tissues, which is risky, as we will show in this study.

In a large retrospective study based on published and unpublished sequences included in the public archives, we questioned the truly exogenous nature of the detected and sequenced viral genomes, after highlighting errors in the annotation of sequences unduly identified as XRV. To this end, we defined the regions of the JSRV and ENTV genomes specific to XRVs. We then reanalysed the available sequences and published amplification tools to determine their specificity, verified the XRV specificity of primers reported in the literature, and proposed a systematic approach with rigorous controls to ensure the exogenous nature of the sequences. We apply our methodology to case studies reporting on the detection of JSRV or ENTV in ticks and mosquitoes (Zhang et al. 2022, Atim et al. 2023, Jamil et al. 2024, Zhang et al. 2024, Ndione et al. 2025, Zhou et al. 2026) using metagenomics or virome approaches and their potential transmission by those vectors.

Our study raises questions about the robustness of the virus discovery in the context of XRVs/ERVs coexistence in small ruminants and the overall challenge in metagenomic and transcriptomic studies in the presence of ERVs and, more globally, endogenous viral elements (EVEs) (Brait et al. 2024).

2. Materials and methods

2.1 Determination of discriminating regions between XRVs et ERVs

Complete nucleic sequences from 30 JSRV, 14 ENTV-1, and 35 ENTV-2 sequences available in GenBank (Table S1, sequences identified with an asterisk *) were aligned with the ERV II.5 consensus for *O. aries* for JSRV and ENTV-1, and for *C. hircus* for ENTV-2 (Verneret et al. 2025). Nucleotide alignments were performed with the 'MAFFT v7.490' software (Katoh and Standley 2013) and processed with the 'R v4.4.1' suite to picture the variability of the identity scores obtained from the comparison between XRV sequences and the ERV consensus and to calculate the percentage of identity for each position in the alignment, averaged over a 100 bp window with a 10 bp step. The analysis was performed individually for each sequence of the alignment, and the mean percentage of identity was plotted. A 95% confidence interval was estimated using a bootstrap approach with the 'boot v1.3–31' suite (Canty et al. 2024), debiased and corrected for asymmetric data. The limited number of samples used and the potential lack of independence between sequences originated from the same origin (herd, geographical location, etc.) are potential biases that we have identified. Four discriminant regions [referred further as discriminating regions (DRs) DR1, DR2, DR3, DR4], were identified with a percentage of identity lower than 80%. DR2 sequence variability was evaluated by aligning [MAFFT Alignment v7.490 (Katoh and Standley 2013)] copies containing the *gag* region, i.e. 28 ERV II.5 copies annotated from the *O. aries* assembly (ARS-UI_Ramb_v2.0 -GCA_016772045.1), 155 ERV

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

II.5 copies from the *C. hircus* assembly (ARS1.2 -GCA_001704415.2), 27 ovine and 10 caprine ERV sequences recovered from GenBank (Verneret et al. 2025) (Table S1). The percentage of each nucleotide among all copies was calculated at each position.

2.2 *In silico* and experimental evaluation of XRV primers specificity

The localization prediction of 286 published PCR primers and probes (Table S2) reported for the detection or amplification of JSRV et ENTVs was determined by mapping them onto the respective alignments for each virus. Publications not written in English were automatically translated with DeepL. A maximum of four mismatches from the XRV sequences was allowed, and the position of each primer was extracted from the alignment, retaining the multi-mapped positions. Primers including additional sequences, e.g. restriction sites, were not included. Primers only mapping on XRV were defined as XRV-specific, whereas primers mapping on a least one copy of ERV were defined as non-specific to exogenous sequences. A selection of seventeen primer pairs has been tested on various DNA templates, covering different parts of the viral genomes. Fourteen of them were tested by PCR on mammalian DNAs; they targeted ENTV-2 for primer pairs # 01–10 or JSRV for primers pairs 12–15 (Table S3). Their ability to detect only XRV was tested using genomic DNAs from JSRV-induced lung tumours of clades I (FR2054) and II (FR989), and ENTV-2-induced nasal tumours (FR3824) (Riocreux-Verney et al. 2024) (Table 1). To evaluate their potential for cross-hybridization with ERV sequences, PCRs were performed using DNAs from ovine IDO5 cells (Institut Mérieux, Lyon, France), caprine TiGEF cells (Da Silva Teixeira et al. 1997) and human HeLa cells (Scherer et al. 1953) (Table 1). Primer pairs 10 (ENTV-2 U3/gag) and 15 (JSRV U3/gag) have been recently published by our team and shown to be highly specific for JSRV ($n = 22$) and ENTV-2 ($n = 10$), with no cross-hybridization with ERV sequences and validated by sequencing (Riocreux-Verney et al. 2024). The amplicons obtained with pairs 01 and 13 for each condition (JSRV, ENTV, and cell references), with pairs 03 to 08 and 14 for the IDO5 cell reference, and with pair 04 for HeLa cells, were subjected to Sanger sequencing (TubeSeq Supreme Eurofins) and aligned against XRV and ERV reference proviral sequences. A BLAT search (Kent 2002) on the human genome (hg38) was performed to identify the nature of the amplicon sequence obtained with pair 04 from human HeLa cells.

A similar approach was taken for the primer pairs published to target ENTV (primer pair #11) and JSRV (primer pair 16 and 17) in arthropods (Jamil et al. 2024, Ndione et al. 2025) by adding DNA extracted from an ENTV-1-induced nasal tumour (FR4403) (Riocreux-Verney et al. 2024) and mosquitoes (*Aedes aegypti and Culex Quique/asciatus*) that have fed on human blood. All the amplicons were subjected to Sanger sequencing (TubeSeq Supreme Eurofins).

2.3 ERV II.5 expression in sheep tissues

Data of mRNA-seq from 60 tissues (Table 2, Ovine FAANG project (Salavati et al. 2020, Davenport et al. 2021)) of the Benz2616 ewe, used to assemble the ARS-UI_Ramb_v2.0 ovine genome [GCA_016772045.1, (Davenport et al. 2022)], and of lung RNA-seq of $JSRV_{JS21}$-experimentally infected lambs and the associated uninfected controls (Karagianni et al. 2019) were collected from the ENA (European Nucleotide Archive) database. ERV II.5 family specific k-mers of 21 base pairs were defined in the DR1 and DR3 regions in order to cover the largest number of copies of ERVs annotated in the ARS-UI_Ramb_v2.0 genome (Table 3 and Fig. S1) (Verneret et al. 2025). These k-mers were used to identify and count the reads corresponding to ERV sequences in the selected RNAseq data using the BBduk analysis tool (Bbtools v38.95, https://quay.io/repository/biocontainers/bbmap) by allowing two mismatches. As a control, k-mers specific to JSRV were defined in DR3 and DR4 and searched in the same RNAseq data (Table 3 and Fig. S1), as previously described (Riocreux-Verney et al. 2024). The reads containing the kmers were extracted from the raw data (FastQC file), low-quality sequences and adapters were removed with Trimmomatic (Bolger et al. 2014). The sequences were then mapped [Bwa-mem, (Li and Durbin 2009)] on the *O. aries* family II.5 consensus sequence (Verneret et al. 2025). The alignment was visualized with IGV [Integrative Genomics Viewer, (Robinson et al. 2011)].

2.4 Alignment and phylogenetic analyses

The alignments of *pol*, DR1-DR2, DR3-DR4, complete genomes, and sequenced amplicons were carried out using MAFFT Alignment v7.490 (Katoh and Standley 2013) and Geneious Prime software (v2025.0.2). Maximum Likelihood (ML) phylogenetic trees in the *pol*, DR1-DR2, and DR3-DR4 regions were generated using IQ-TREE v1.5.3 (Nguyen et al. 2015) with 10 000 ultrafast bootstrap and rooting on the BT ERV (*Bos taurus* ERV) family II.4 consensus sequence (Verneret et al. 2025). ML-type phylogenetic trees (PhyML 3.3.20180621, Jukes Cantor substitution model) of the *gag* (pair #01, 826 nt) and *env* (pair #13, 864 nt) sequences were created using the ENTV1-FR4403 sequence (GenBank #PP669280, (Riocreux-Verney et al. 2024)) for rooting, with 1 000 iterations (bootstraps) and edited through the iTOL software.

Publicly available sequences used in the study are reported in Table S1. ERV copies from the family II.5, annotated in the *C. hircus*

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

System	Tissue	ENA accession #	System	Tissue	ENA accession #
Cardiac	Right ventricle	ERR3728426	Neurological	Retina	ERR3728428
	Atrioventricular valve	ERR3728429		Spinal cord	ERR3728434
	Right atrium	ERR3665709		Pons	ERR3728433
	Left ventricle	SRR7612070		Thalamus	ERR3728435
	Vena cava	ERR3728430		Hippocampus	ERR3665708
Immunitary	Spleen	ERR3728431		Pituitary gland	ERR3650379
	Mesenteric lymph node	ERR3665711		Cerebellum	SRR6651979
	Prescapular lymph node	ERR3665712		Frontal cortex	SRR7612068
	Mandibular lymph node	SRR7543059	Renal	Adrenal medulla	SRR7543054
	Alveolar macrophage	SRR6651987		Adrenal cortex	SRR6651976
Gastro-intestinal	Ileum	ERR3728427		Renal medulla	SRR7543058
	Ileum Peyer patches	ERR3665710		Renal cortex	SRR7612071
	Cecum	ERR3665706	Respiratory and	Tongue	SRR7612076
	Jejunum	SRR7543057	buccal	Soft palate	ERR3650378
	Duodenum	ERR3650374		Epiglottis	SRR6651978
	Rectum	ERR3728432		Palatine tonsil	SRR7612075
	Rectum	ERR3665713		Lung	SRR6651986
	Descending colon	SRR7543055	Urinary	Urinary bladder	ERR3728424
	Spiral colon	ERR3665714		Urethra	ERR3665716
	Oesophagus	ERR3665707		Ureter	SRR7543063
	Abomasum	ERR3665705	Endocrine	Mammary gland	ERR3728425
	Abomasum pylorus	ERR3650373	Genital	Uterus	ERR3665717
	Reticulum	ERR3650375		Oviduct	SRR7543061
	Rumen atrium	ERR3650376		Ovary	SRR6651991
	Rumen ventral	ERR3650377		Vagina	SRR7612077
	Omasum	SRR7543060	Muscular	Biceps femoris	SRR6651988
	Gall bladder	SRR7543056		Semimembranosus	SRR6651990
	Liver	SRR7612072		Longissimus dorsi	SRR7612073
Skin	Skin hairless	ERR3665715		Supraspinatus	SRR7612074
	Skin of back	SRR7543062		Diaphragm	SRR7612069

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

Target	DR	Name	5′—3′sequence
ERV	DR1	kERV-DR1	TATGCTGTTCCCGAGGGGGTT
	DR3	kERV-DR3	GAGAGTTTTAATACATAAAAA
JSRV	DR3	kJSRV-DR3	GAGAGTTGAAATGCTACATAT
	DR4	kJSRV-DR4	TCTGATTATGTAAGAATCCGG

assembly ARS1.2 (GCA_001704415.2) and the *O. aries* assembly ARS-UI_Ramb_v2.0 (GCA_016772045.1), were retrieved from (Verneret et al. 2025). The partial proviral sequence OR991120 described in (Jamil et al. 2024) and the partial *env* sequence KT266728 (Ndione et al. 2025) were also analysed.

3. Results

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

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

Of the 286 PCR primers/probes listed in the 109 published studies, only 28% (79 out of 286) were qualified as discriminating between XRVs and ERVs by our *in silico* analysis (Table S2). Notably, among

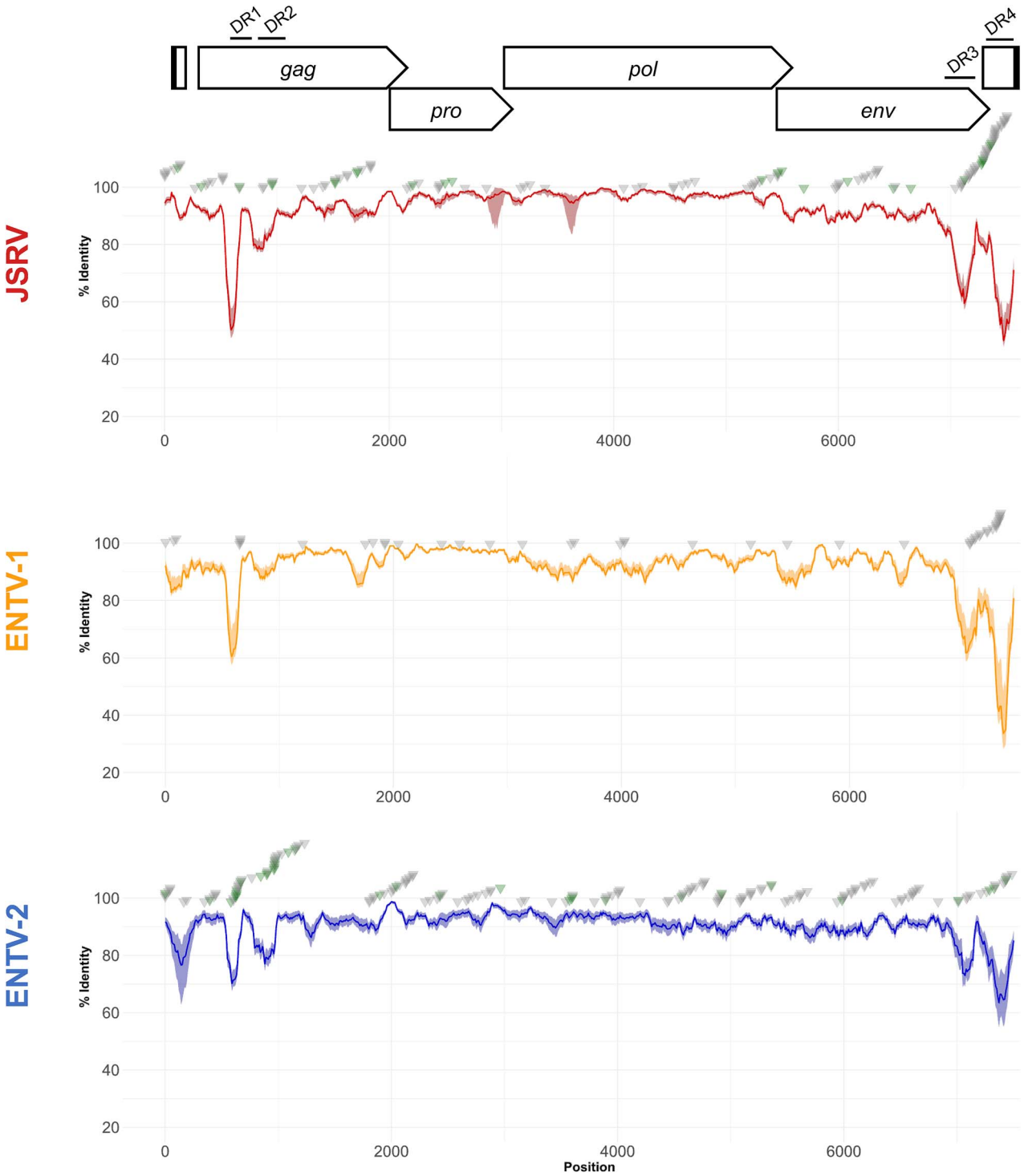

Figure 1 XRV are genetically close to ERVs. The identity score was calculated from a multiple alignment (MAFFT) of XRV sequences against a consensus sequence of ERV II.5 Oa for JSRV and ENTV-1 and ERV II.5 Ch for ENTV-2. The average identity percentage was calculated using a 100-base pair window and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

the 286 primers, 74 (26%) are in the *pro* or *pol* genes, which are regions that are highly similar between XRV and ERV sequences. Similarly, less than one-third of the primers in R/U5 region are specific. Therefore, in order to minimize the risk of amplifying ERVs instead of XRV, they must be combined with XRV-specific primers.

3.2 Phylogenetic reconstructions reveal the challenge of correctly classifying XRV

Phylogenetic trees were reconstructed in the *pol* region (Fig. 2), which weakly discriminates between XRVs and ERVs, as well as in the DR1-DR2 (Fig. 2B) and the DR3-DR4 (Fig. 2C) segments from 32 JSRV, 14 ENTV-1, 45 ENTV-2, 27 ERV II.5 Oa and 10 ERV II.5 Ch sequences extracted from GenBank and supplemented by sequences identified on the assemblies of the *O. aries* (ARS-UI_Ramb_v2.0) and *C. hircus* (ARS1.2) reference genomes (Verneret et al. 2025).

These analyses revealed that the segregation of viral species varies considerably depending on the region considered. The *pol* region did not permit a clear distinction between XRV and ERV II.5 (Fig. 2A). Conversely, the DR1-DR2 and DR3-DR4 regions (Fig. 2B-C) enabled the different viral forms to be separated into XRVs and ERVs. However, 17 of the XRV sequences were grouped with the ERV sequences, regardless of whether the trees were reconstructed using the DR1-DR2 or DR3-DR4 regions, which calls into question their exogenous origin. Some of the potentially misidentified sequences presented a clear ERV pattern, including specific indels in the DR1-DR4 segment and the lack of the YXXM motif in Env (e.g. MK164396 and MK210250 for ENTV-2 or DQ838494 and MZ931278 for JSRV, Fig. 3). The MK210250 sequence displayed a very strong sequence identity (>96%) with the ERV II.5 *C. hircus* consensus sequence in DR4, whereas the MK164396 sequence did not show a clear pattern, being equally distant from the exogenous and endogenous reference sequences (with an identity percentage ranging from 80% to 83%). Two JSRV sequences, DQ838494 & MZ931278, which cluster with ERVs, had DR4/U3 sequences that are very different from the ERV II.5 consensus sequence with 35% and 71% identity, respectively, as well as from the JSRV-FR2054 (PP707050.1) reference sequence with 33% and 52% identity, respectively. The BLAT of the DR4 sequence of DQ838494 against the sheep genome (Oar_v4.0/oviAri4) resulted in multiple hits. The first hits (with over 95% identity) corresponded to LINE sequences of the RTE-BovB family, suggested the unmonitored amplification of another transposable element.

Different combinations of mixed XRV/ERV patterns can be observed, such as DR1 and DR3 insertions ('ERV like') for ENTV-2 ON843769 and MZ931277, and DR3 deletion ('XRV like') for PP682590. A group of sequences reported as ENTV-2 (KU258870 to KU258880) and integrated into the caprine ERV II.5 phylum does not display the described typical ERV indels profile. Although only KU258870 lacks the YXXM motif, they were still more similar to ERVs in DR1-DR2 and DR3-DR4 regions. No accessible articles describing the conditions under which the sequences were generated were available for any of these sequences except DQ838494 (Shuying et al. 2006) and MK164396 (Zhai et al. 2019), which cluster with ERV II.5 sequences. In the case of DQ838494 and MK164396, the associated primer sets were predicted to have very low to no exogenous specificity for most primer pairs (Table S2). This was likely the original cause of the misidentification of these sequences. These genomes, which are most probably ERVs rather than XRVs or ERVs/XRVs artificial chimeras, have been used as reference sequences in other studies to develop molecular tools (Ye et al. 2019, Zhai et al. 2019, Li et al. 2024).

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

3.3 Reliability of XRV detection in an ERV context

Numerous published studies have reported the detection of JSRV and/or ENTV using primers located in regions that are conserved between XRVs and ERVs. This makes it impossible to exclude the amplification of ERVs (Table S2). To illustrate this challenge, fourteen primer sets, defined as specific to JSRV or ENTVs, were selected (Table S3). Specific primers for XRVs were included as specificity control (Riocreux-Verney et al. 2024).

The primers were defined by the following 3 criteria. The first one was the ability of primers to detect either exogenous ENTV-2 (pairs 1–10) or JSRV (12–15) according to the authors (Tables S2 and S3). The second criterion was the specificity of the primer predicted through XRV and ERV sequences alignment. The third criterion relied on the ability of primers to produce an amplicon of the expected size only from genomic DNA extracted from JSRV- or ENTV-2-induced tumours and not from uninfected ovine, caprine, and human cells.

The theoretical specificity of the primers was predicted by aligning them with the XRV and ERV sequences published from the public databases (Tables S1 and S2). Of the ten pairs selected for ENTV-2, primer alignment identified seven pairs (#1, 3, 4, 5, 6, 7, 8) that were non-specific for XRV, due to high homology with ERV sequences. Only three pairs (2, 9, and 10) were specific for ENTV-2. Of the four pairs selected for JSRV, only pair #14 was predicted to be specific for XRV. The primer pairs were then evaluated for their ability to specifically amplify XRV, with amplicon generated only from gDNA extracted from ENTV-2 or JSRV-induced tumours, and conversely to fail to produce signal from gDNA extracted from IDO5, TiGEF, or HeLa cells. Amplicons of the expected size from all the gDNAs tested, whether extracted from tumours induced by ENTV-2 as well as by JSRV or from uninfected ovine, caprine were generated for the seven primer pairs # 1, 3, 4, 5, 6, 7, 8 predicted to be non-specific for ENTV-2. More surprisingly, DNA was amplified with the primer pair #4 for human cells. Sequencing confirmed this to be HERV-K (Fig. S4).

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

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

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

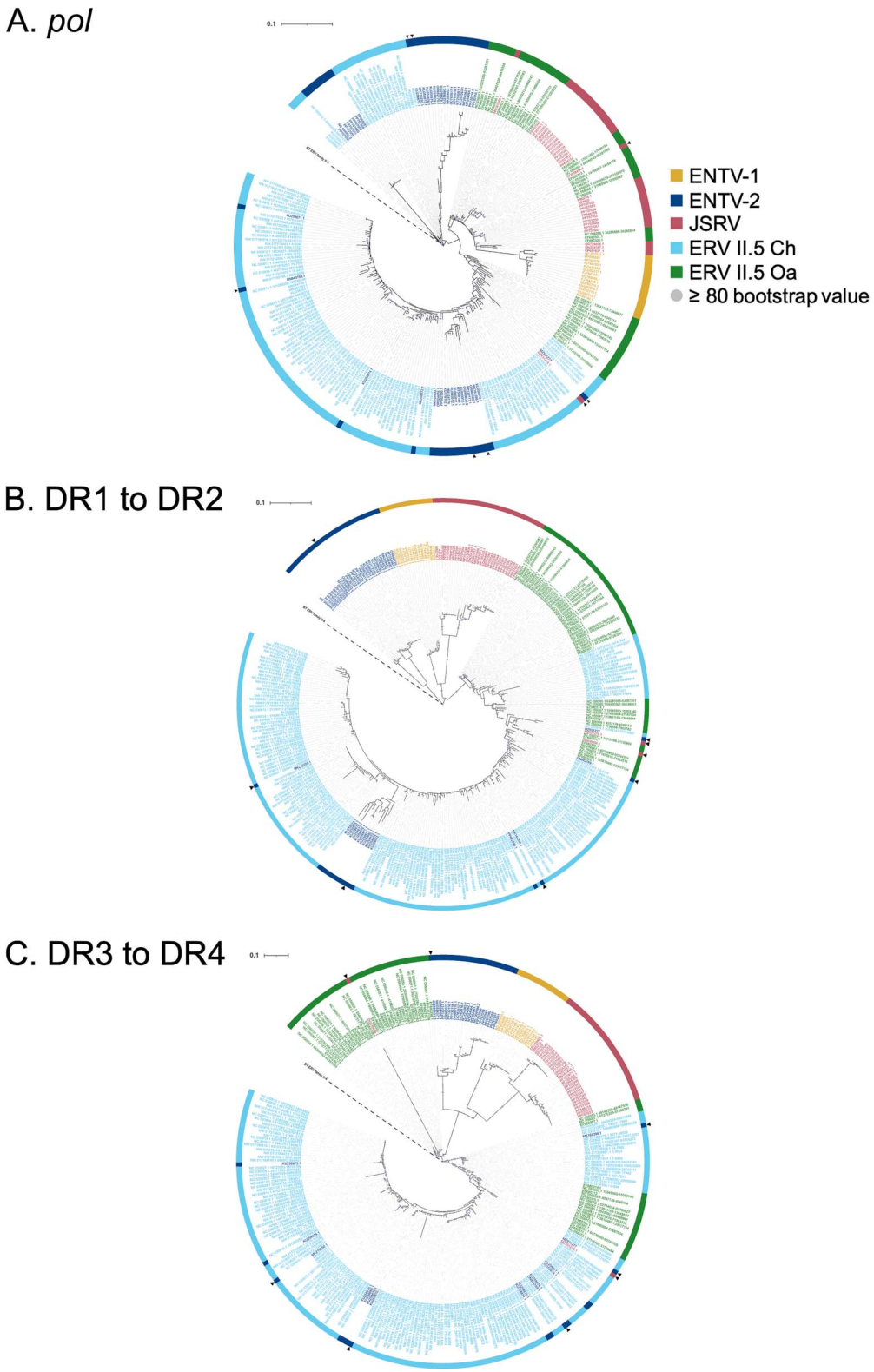

Figure 2 DR1-DR2 et DR3-DR4 are sufficient to classify JSRV, ENTVs, and ERV sequences. Full-length sequences XRVs and ERVs available on GenBank by July 2025, and ERV copies annotated in *C. hircus* and *O. aries* were used for phylogenetic reconstructions. The maximum likelihood trees generated using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

Figure 3 Example of wrongly annotated XRV retroviruses viral genome sequences of JSRV (DQ838494; MZ931278) or ENTV (MK164396; MK210250) strains were aligned (MAFFT) with viral genomic reference sequence of *C. hircus* and *O. aries* II.5 ERV consensus sequences (Verneret et al. 2025) JSRV clade I (FR2054; PP707050.1), clade II (FR989; PP707047.1), ENTV-1 (FR4403, PP669280.1), and ENTV-2 (FR3824, PP669281.1) in A. DR1, B. DR2, C. DR3, and D. DR4. Only nucleotides distinct from the ERV II.5 Ch (top sequence) are indicated; identical nucleotides are marked by a dot. The position on the alignment is indicat d on top.

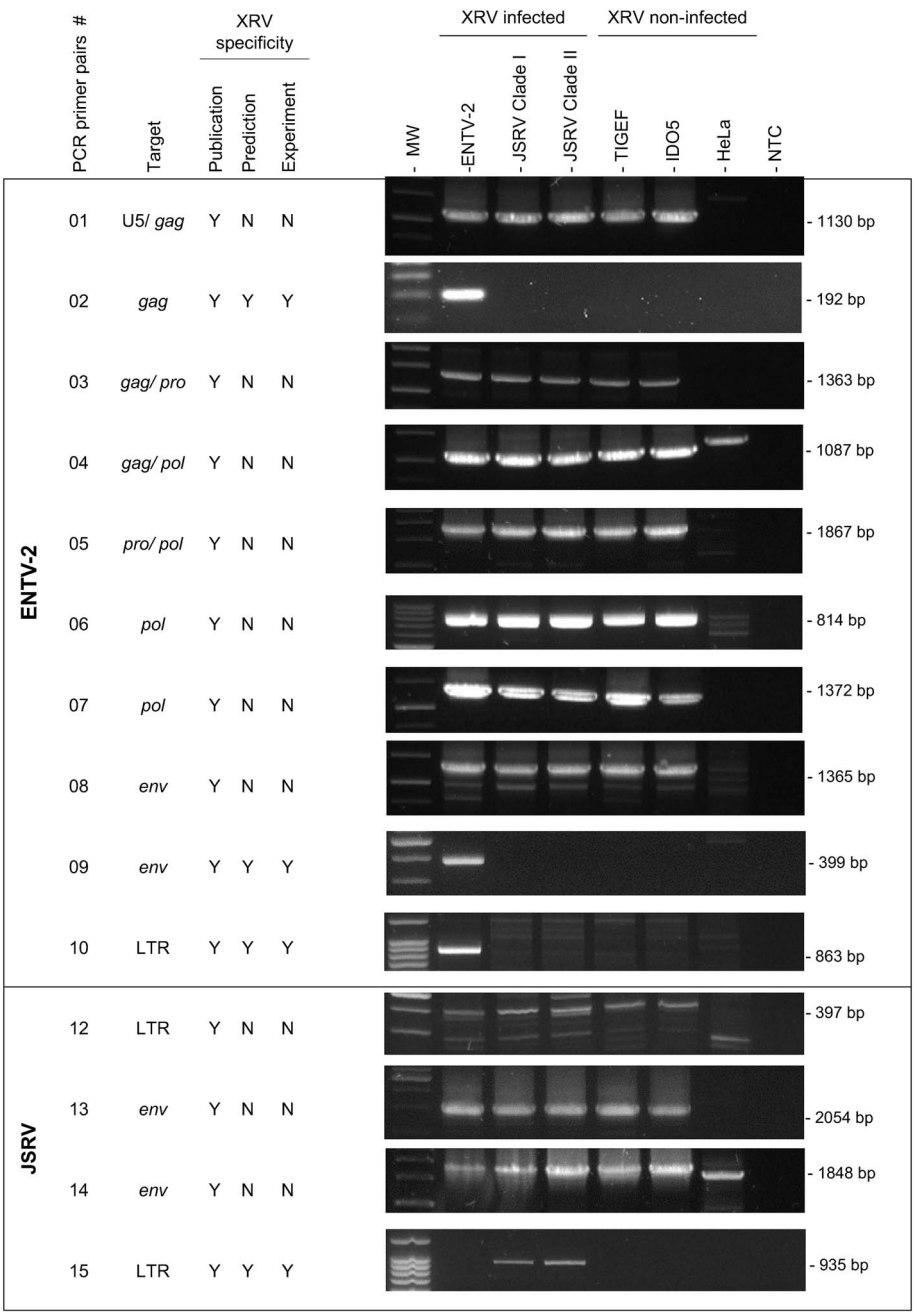

Figure 4 Lack of specificity of selected XRV primers. We analysed the 'XRV specificity' of fourteen published primer pairs (01–10 and 12–15) according to the authors' statement ('publication'), our *in silico* analysis ('prediction'), and their profile of detection in PCR ('experiment'). The 'XRV infected' reference panel consisted of genomic DNAs extracted from a nasal tumour induced by ENTV-2 (FR3824, GenBank accession PP669281.1) and lung cancers induced by clade I (FR2054; GenBank accession PP707050.1) or clade II (FR989; GenBank accession PP707047.1) JSRV. The 'XRV non-infected' reference panel consisted of genomic DNAs extracted from goat (TiGEF), sheep (IDO5), and human (HeLa) cell lines. Y: Yes; N: No. NTC: No template control.

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

3.4 Multiple tissues express endogenous sheep and goat retroviruses

Several studies have been based on amplifying JSRV and ENTV from RNAs under the assumption that only XRVs are expressed. As RNA is an appropriate target for retroviruses detection, the absence of

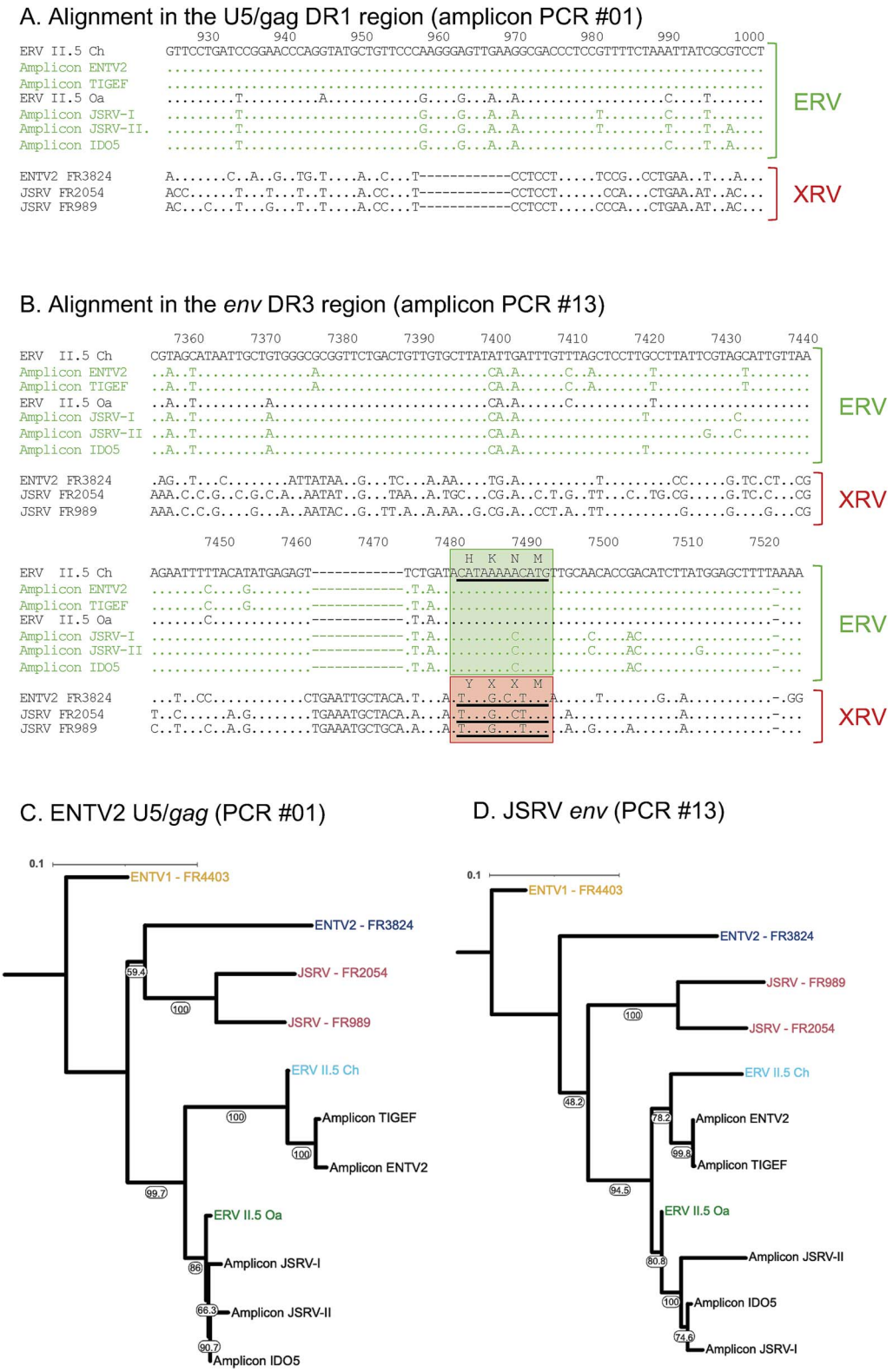

Figure 5 Absence of exogenous DR1 and DR3 discriminatory regions in amplicons obtained with primer pairs #01 (ENTV-2) and #13 (JSRV). Amplicons obtained with primer pairs #1 (ENTV-2) (A and C) and #13 (JSRV) (B and D) were sequenced and aligned (MAFFT) with a reference sequence of ENTV-2 (FR3824, GenBank accession PP669281.1), JSRV clade I (FR2054; GenBank accession PP707050.1) & II (FR989; GenBank accession PP707047.1) and the consensus sequences of ERV II.5 Oa and ERV II.5 Ch [4]. Only nucleotides distinct from the ERV II.5 *Ch* consensus sequence (on top) have been indicated; identical nucleotides are marked by a dot. Maximum likelihood (ML) trees were reconstructed using PhyML from C. 826-nt of the U5/gag or D. 864-nt of the *env* amplicons with ENTV1-FR4403 sequence (GenBank accession PP707036) as an outgroup for rooting. Bootstrap values (1 000 replicates) are expressed as a percentage and indicated at the node level. The scale bar indicates the number of substitutions per site.

ERV expression must be confirmed to rule out dual expression of XRVs and ERVs at the same sites. We analysed the expression of ERV II.5 using transcriptomic data from 60 tissues of the Benz2616 ewe, the source of ARS-UI_Ramb_v2.0 reference ovine genome. Short sequences or k-mers were defined in DR1 in *gag* and DR3 in *env* (Fig. 6, Table 3 and Fig. S1) from annotated ERV II.5 copies of the ARS-UI_Ramb_v2.0 genome (Verneret et al. 2025). JSRV k-mers located in DR3 and DR4 were used to detect JSRV expression from RNA-Seq data from JSRV-experimentally lambs and confirm the JSRV-negative status of the Benz2616 ewe (Fig. 6A, Table 3 and Fig. S1) (Karagianni et al. 2019, Riocreux-Verney et al. 2024). For each of the RNA-Seq datasets, sequences with specific ERV or JSRV k-mers were counted using the BBduk tool. Conversely, ERV k-mers identified small quantities of ERV transcripts in most of the tested tissues, with notably higher expression in the lung and kidney (particularly the internal medullary part) and oviduct (Fig. 6B). Sequences identified by the ERV k-mers were mapped to the consensus sequence of ERV II.5 Oa, confirming their endogenous origin proven the specificity of this approach to identify transcripts (Fig. 6C). This demonstration in sheep shows that amplifying RNAs does not guarantee the sole detection of JSRV or ENTV. If the primers are not highly XRV-specific, ERVs will be amplified. Additionally, the absence of cellular DNA, even as traces, must be carefully controlled to prevent the amplification of ERV copies, present in multiple copies.

3.5 The presence of JSRV and ENTVs in arthropods cannot be confirmed

Six recent studies identified small ruminant XRV sequences (Zhang et al. 2022, Atim et al. 2023, Jamil et al. 2024, Zhang et al. 2024, Ndione et al. 2025, Zhou et al. 2026) in insect vectors using global approaches, such as metagenomics, transcriptomics, and whole RNA-Seq enriched for viral RNA. We used these datasets to challenge both our strategy for specifically detecting XRVs in a context of coexistence of both XRV and ERVs in the same individuals and the hypothesis of a potential vectorisation. Those approaches rely on the BLASTn of the *de novo* assembled genomes to identify reference sequences. Two studies made the reference matches available (Atim et al. 2023, Ndione et al. 2025). While analysing the different datasets, we observed that identification of sequences deposited in GenBank can be misleading. For example, MN564750.1 used as a reference, is associated in GenBank with the organism 'Enzootic nasal tumor virus 2' (Taxonomy ID: 2913605), ID which also includes 'Endogenous enzootic nasal tumor virus'. MN564750.1 is identified as 'Enzootic nasal tumor virus 2 isolate enENTV-FJ2 genomic sequence'. A scientist familiar with these viruses would recognize that the name of the isolate 'enENTV-FJ12' corresponds to an endogenous (en) virus. However, this is confusing, particularly in the absence of an associated article or any further information about the origin in the deposit entry. We clearly identified ERV sequences wrongly tagged as XRV in public databases such as MK210250 (Figs 2 and 3) or KT266728 (Fig. S6), used as a reference, respectively, for the detection of ENTV-2 (Atim et al. 2023) and JSRV (Ndione et al. 2025), which both lack the sequence coding for the YXXM motif. Overall, this detailed initial analysis highlights the importance of carefully selecting the reference sequence to exclude ERV sequences.

In their study (Ndione et al. 2025), the authors identify both sequences of JSRV and ENTV-2 within mosquitoes, but only validated the presence of JSRV. While the amplification and sequencing led to the identification of the sequence as 'enJSRV', the validation was initially biased as primers (Tables S2 and S3) not specific for exogenous viruses were designed based on the sequence initially matching with their reference. The non-specificity of primers #16 and #17 was confirmed using the same experimental design with reference samples of DNA extracted from XRV-induced tumours, negative sheep and goat controls, including DNA extracted from different mosquito species. The amplification pattern was non-specific, with amplification in both uninfected and infected samples (Fig. 7A). The #16 and #17 primer sets amplified regions confirmed to be of ERV origin by sequencing. The DR3 amplicon obtained with PCR #16 (Fig. 7B) does not contain the insertion specific of XRVs nor the sequence coding for the YXXM motif; the DR4 amplicon obtained with PCR #17 (Fig. 7A-C) has both the size and sequence compatible with the ERV II.5 consensus. Neither the sequence nor the accession number was made available for the 'nearly complete sequence' of ENTV-2 identified in mosquitoes. However, the authors reported that the sequence was closely related to the GDQY2017 Chinese strain (accession number MK164396), a sequence that we identified as ERVs, questioning the XRV nature of the ENTV-2 sequence identified in this study (Figs 2 and 3) (Ndione et al. 2025).

In a recent study (Jamil et al. 2024), ENTV was amplified and sequenced from ticks collected in Pakistan. A nearly full-length proviral genome (OR991120) was amplified (Fig. 8A). The sequence displayed 98% and 92% overall nucleotide identity, respectively, with the caprine and sheep ERV II.5 consensus, but <89% with JSRVs and ENTVs. The DR3 and DR4 regions were unavailable. However, the DR1 and DR2 regions displayed indel characteristics of ERVs (Fig. 8C-D), as well as an ERV-specific partial insertion in DR2 region. These findings suggest that the 'provirus' was amplified from a goat ERVs. The single set of primer targeting the *gag/pol* was not predicted as specific of XRVs, allowing an amplicon from both XRV-infected and uninfected sources (Fig. 8B). The DR1 and DR2 sequencing of the amplicon confirmed the ERV nature of the sequence (Fig. 8C-D).

Three studies identified ENTV-2 in ticks isolated from cattle (Zhang et al. 2022), JSRV from sheep (Zhou et al. 2026), ENTV-2 and JSRV from ticks isolated from yaks and ENTV-1 from Tibetan sheep (Zhang et al. 2024). Unfortunately, neither data on the overall nor DR regions coverage were available, nor molecular validation of the presence of JSRV and ENTVs with XRV-specific tools was reported; only information regarding the abundance of the reads was available for (Zhang et al. 2022) and (Zhang et al. 2024). Therefore, the conclusions that can be drawn about the exogenous nature of these sequences are limited without further investigations. *A contrario*, a thorough checking based on the current recommendation can dispel doubts about the nature of the strains detected by these high-throughput approaches. We therefore collaborated with the authors to rework the preprint (Atim et al. 2023), which initially described ENTV-2 within ticks. One of the tick metagenome samples displayed coverage of the ENTV-2 provirus of over 80%, thus enabling to distinguish between ERV/XRV and clearly identifying the DR1–3 regions associated with ERV. DR4 coverage was lower, but also suggested the detection of ERV. Phylogenetic analysis confirmed these conclusions, as the consensus sequence clustered with the ERV sequences (Fig. S7).

4. Discussion

Accurate detection of β-retroviruses responsible for respiratory cancers in sheep and goats is a critical prerequisite for controlling these

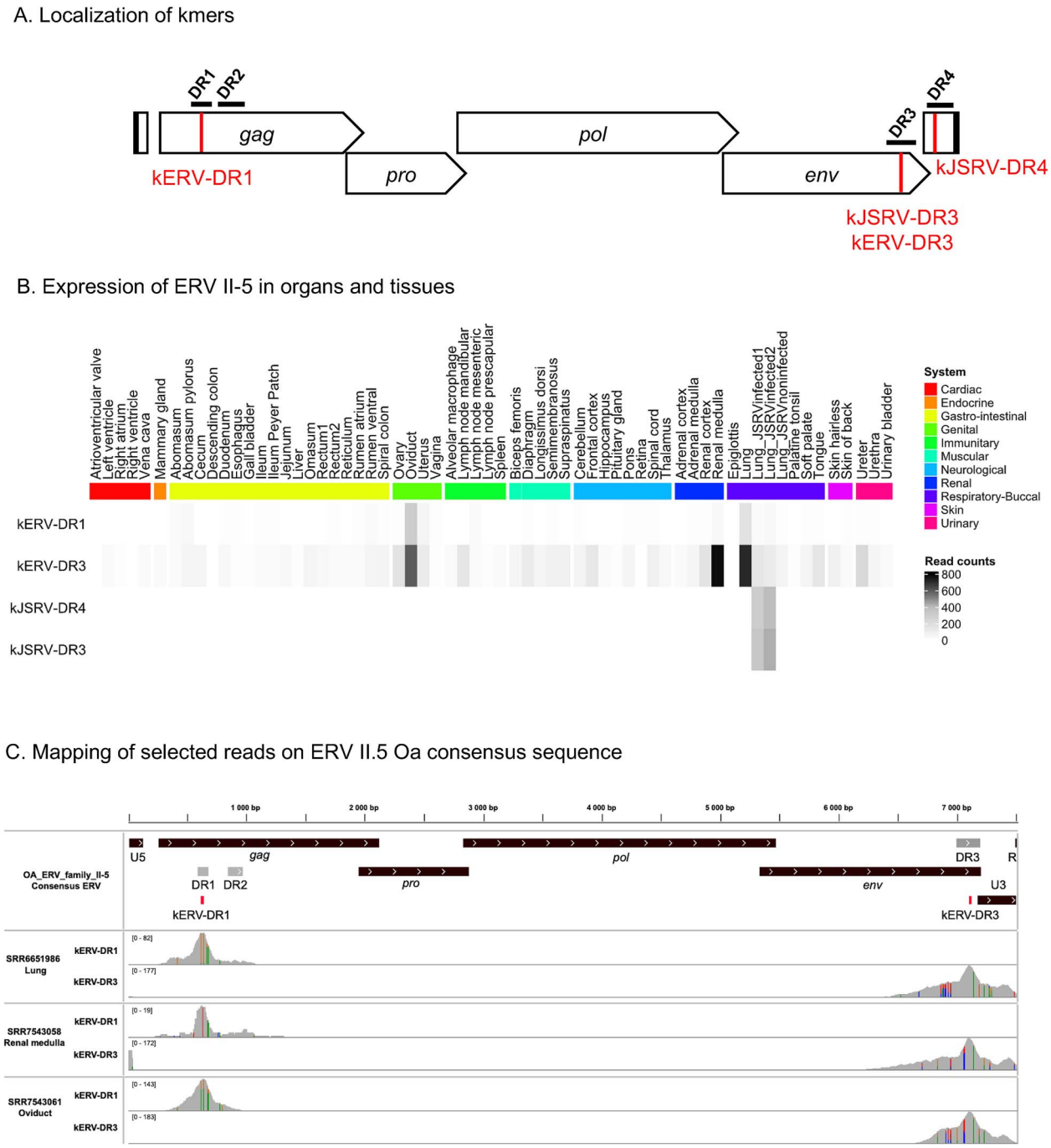

Figure 6 Expression of ERV family II.5 in tissues. A. K-mers localization on the JSRV provirus. kERV-DR1 and kERV-D3 are specific to ERV II.5 Oa and located in the DR1 (*gag*) and DR3 (*env*) DRs, respectively. kJSRV-DR3 and kJSRV-DR4 are specific to JSRV and located in DR3 (*env*) and DR4 (U3). B. The number of reads containing one of the four selected k-mers was calculated for each tissue or organ. The samples were grouped by physiological system and visualized using a colour code. JSRV-infected and uninfected control sample (Karagianni et al. 2019) were added and labelled as 'lung_JSRV_infected' for lung RNA-Seq data of JSRV-JS21-infected lamb or 'Lung_JSRV_uninfected' for uninfected control lambs. C. Reads containing kERV-DR1 and kERV-DR3 k-mers were mapped on the ERV II.5 Oa consensus sequence and visualized on IGV.

diseases in livestock. This task is complicated by the coexistence of XRVs (JSRV and ENTV) with numerous copies of endogenous retroviruses belonging to the ERV II.5 family in the genomes of sheep and goats. Molecular diagnostic approaches must therefore allow unequivocal detection of XRV and therefore target regions that clearly discriminate between ERVs and XRVs. Insufficient consideration of this constraint has directly impacted the specificity of molecular assays, the interpretation of phylogenetic relationships, and the integrity of sequence annotations.

Early studies, based on a limited number of available sequences, suggested that regions 1–4 could discriminate between ERVs and XRVs due to their low sequence homology (Palmarini et al. 2000, Walsh et al. 2010). Building on these observations, we conducted an extensive analysis integrating available XRV and ERV sequences with

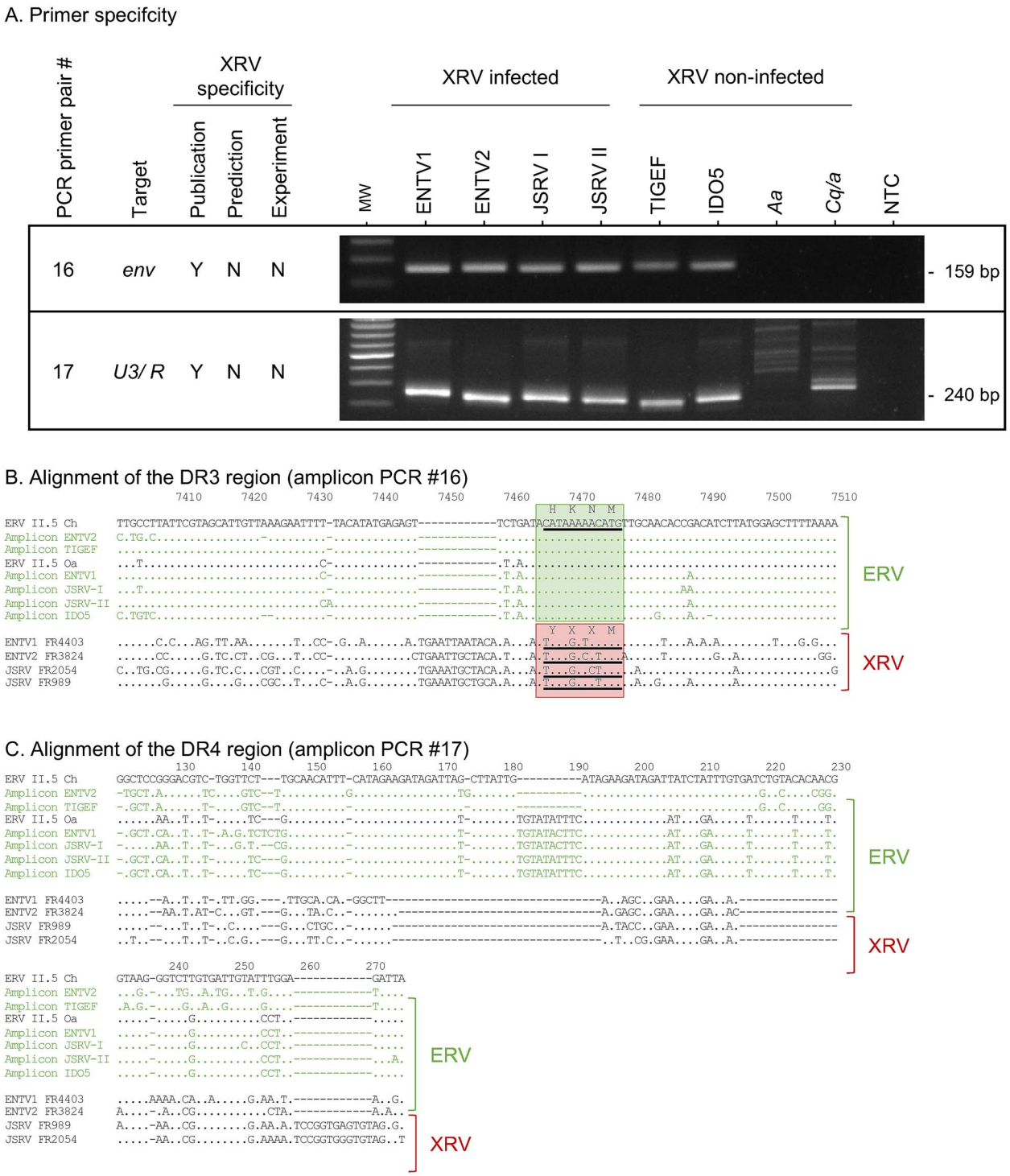

Figure 7 Molecular detection of JSRV and ENTVs in mosquitoes. A. Evaluation of the specificity of primer pairs # 16 and #17 on genomic DNA extracted from JSRV or ENTV tumours, from uninfected sheep (IDO5) and goat (TIGEF) cell lines, and from *A. aegypti (ae)* and *Culex Quique/asciatus (Cq/a)*. The amplicons obtained with primer pairs #16 (B) and #17 (C) were sequenced and aligned against ERV II.5, ENTV and JSRV reference sequences. Only nucleotides distinct from the ERV II.5 *Ch* consensus sequence (on top) have been indicated; identical nucleotides are marked by a dot.

in silico annotation of ERV II.5 *loci* in sheep and goat assemblies, which allowed us to define four discriminating segments (DR1–DR4) that robustly discriminate exogenous from endogenous retroviruses. DR1 and DR2 localize in the 5'gag region, DR3 in the 3' terminus of *env*, and DR4 to U3. Thus, they are prime targets for virus detection, reconstruction of phylogenetic relationships, and sequencing of near-full-length exogenous β-retroviruses. As we established in this study from *in silico* analysis of primers reported in the literature, targeting other regions with limited variation between ERVs and XRV forms poses a risk of ERV detection.

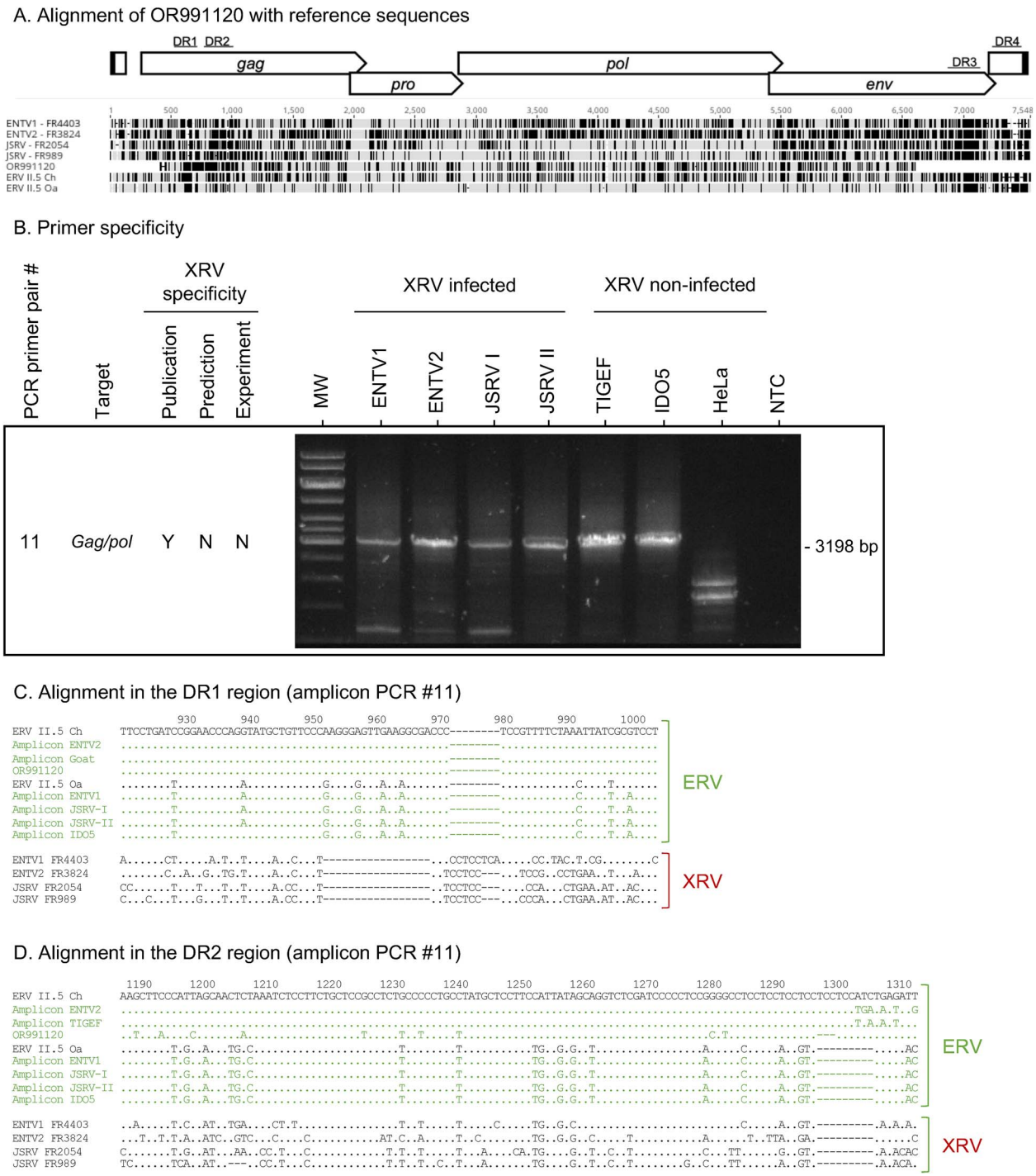

Figure 8 Molecular detection of ENTV in ticks. A. The partial OR991120 proviral sequence, missing the LTRs and the DR3 region, reported in ticks (Jamil et al. 2024) was aligned with II.5 ERV Oa and Ch consensus sequences (Verneret et al. 2025), JSRV clade I (FR2054; GenBank accession PP707050.1), clade II (FR989; GenBank accession PP707047.1), ENTV-1(FR4403, GenBank accession PP669280.1), and ENTV-2 (FR3824, GenBank accession PP669281.1). The black and grey bars correspond respectively to the differential or identical nucleotides. B. Evaluation of the primer specificity. The amplicon #11 was sequenced and aligned against proviral sequences of JSRV, ENTV, and ERV II.5. C. Alignment of the DR1 region. D. Alignment of the DR2 region.

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

Misinterpretation extends beyond host samples. Recent reports suggesting JSRV or ENTV presence in ticks or mosquitoes may most likely reflect the detection of goat or sheep DNA acquired from blood meal rather than true viral infection. As we have shown, there is no evidence that arthropods could serve as reservoirs or vectors for JSRV and ENTV. Xenosurveillance may help to identify infection (Ndione et al. 2025) but the clear frontier with ERVs has to be drawn.

The distinction between XRV and ERV is often blurred in sequence databases, such as GenBank, due to the presence of redundant and inconsistent taxonomic identifiers that merge XRV and ERV. While a single taxonomic ID (69576) is defined for ENTV-1, JSRV is classified within the taxonomic ID 47898 (recognized by the ICTV), which also includes the 'Endogenous Jaagsiekte retrovirus' (also classified with the *O. aries*, taxonomic ID: 9940). A subclass 'Exogenous Jaagsiekte sheep retrovirus' (Taxonomic ID: 47898) is also reported, but this only contains six sequences. For ENTV-2, the problem is even more complex. Recently, four taxonomic IDs coexisted for the same virological entity, but taxonomic ID 2762664 (Endogenous enzootic nasal tumour) and taxonomic ID 239365 (ENTV of goats) were recently merged with taxonomic ID 2913605 (ENTV 2). The fusion of taxonomic ID 2762664 with 2913605 artificially grouped XRV and ERV sequences, whereas the latter could be defined as part of the goat genome with identifier 9925 (*C. hircus*). Finally, the only singularity of taxonomic ID 2584748 is that it is defined by its common name in US English, with 'tumour' instead of 'tumour'.

The classification of ERVs based on whether they are considered part of the virus or the host is still a challenge, not limited to retroviruses in small ruminants. As an example, porcine endogenous retroviruses are associated in GenBank either to their host genome, *Sus scrofa* (Taxonomic ID 61673) or to the viral entity PERV (Porcine endogenous retrovirus) (Taxonomic ID: 61673), subdivided in 11 subclasses with their own taxonomic IDs. Brait and colleagues (Brait et al. 2024) suggest a rational solution to bypass this challenge by adding a second taxonomic identifier. In this context, referring to ERV II.5 with the taxonomy ID of the host in which they have been identified and a second taxonomic ID related to the closest XRV offers a pragmatic solution.

Finally, genome assembly approaches using short amplicon sequencing are ill-suited to contexts where ERVs and XRVs coexist, as they cannot exclude chimeric reconstructions. Long read sequencing technologies are therefore preferable and less prone to generate *in silico* artefacts (Riocreux-Verney et al. 2024). Regardless of the approach used, the genetic nature of the DR1 to DR4 DRs is a minimal requirement for classifying sequences as exogenous. Hallmarks features such as deletions in DR1/DR2 regions, as compared to ERV II.5, and the presence of the exogenous oncogenic YXXM motif in DR3 (strictly absent in ERV) provide unambiguous evidence of exogenous origin.

Although most of the complete sequences of JSRV and ENTV present in GenBank possess these motifs, indicating a probable exogenous origin, doubt remains about the non-discriminatory regions, such as the end of *gag*, *pro*, *pol*, and the beginning of *env*. For example, a recent study suggests possible recombination between different strains of ENTV-2, precisely in the *pro* and *pol* regions, in which the absence of the XRV/ERV discriminatory region complicates amplification and the specific identification of sequences of exogenous origin (Li et al. 2024). In this case, the reconstruction of chimeric XRV/ERV sequences cannot be ruled out.

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

In conclusion, we urge the stringent validation of the molecular tools used to detect JSRV, ENTV-1, and ENTV-2. We propose the following recommendation for assay development:

1—Primer pairs should include at least one primer located in DRs DR1–DR4,

2—The absence of amplification from DNA of uninfected ovine and caprine cells must be confirmed to establish specificity.

3—RNA-based detection alone is insufficient to exclude ERV contamination, particularly in complex clinical samples, and if we consider that ERV expression is present in various tissues.

4—Assembled viral genomes lacking exogenous hallmarks in DR1–DR4 using separated amplicons should be considered as potential chimeric genomes, and not the reflect of the circulating strains.

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

Acknowledgements

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

Author contributions

BRV, CD, and SAA realized the experiments. BRV, MV, SA, JT, and 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.

Supplementary material

Supplementary material is available at *VEVOLU Journal* online.

Conflicts of interest

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

Funding

This research was funded by ANR, grant ANR-22-CE35–0002-01, by INRAE 'AAP-Action incitative sur priorités SSD 2021-2025' and by European Union's Horizon Europe Project 101136346 European Partnership in Animal Health and Welfare (EUPAHW). BRV was the recipient of a VetAgro Sup PhD grant. MV PhD fellowship was funded by ANR, grant ANR-22-CE35–0002-01.

Data availability

The data underlying this article are available in the article and in its online supplementary material.

References

Abass FA, Khudhair YI. Clinical, molecular, and pathological investigations of ovine pulmonary adenocarcinoma in the middle of Iraq. *Open Vet J* 2022;**12**:264–72. https://doi.org/10.5455/OVJ.2022.v12.i2.15

Arnaud F, Caporale M, Varela M *et al.* A paradigm for virus-host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses. *PLoS Pathog* 2007;**3**:e170. https://doi.org/10.1371/journal.ppat.0030170

Atim SA, Ashraf S, Niebel M *et al.* Ticks; a reservoir for virus emergence at the human-livestock interface in Uganda. bioRxiv. 2023;2023.03.10.532017. https://doi.org/10.1101/2023.03.10.532017

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20. https://doi.org/10.1093/bioinformatics/btu170

Brait N, Hackl T, Morel C *et al.* A tale of caution: how endogenous viral elements affect virus discovery in transcriptomic data. *Virus Evol* 2024;**10**:vead088. https://doi.org/10.1093/ve/vead088

A. Canty, B. Ripley, maintainer 1999–2022, support, author of parallel, and fixes, A. R. B. minor bug (2024). boot: Bootstrap Functions (Originally by Angelo Canty for S). Available at: https://cran.r-project.org/web/packages/boot/index.html (Accessed February 11, 2025).

Caporale M, Martineau H, De las Heras M *et al.* Host species barriers to Jaagsiekte sheep retrovirus replication and carcinogenesis. *J Virol* 2013;**87**:10752–62. https://doi.org/10.1128/JVI.01472-13

Cousens C, Minguijon E, Dalziel RG *et al.* Complete sequence of enzootic nasal tumor virus, a retrovirus associated with transmissible intranasal Tumors of sheep. *J Virol* 1999;**73**:3986. https://doi.org/10.1128/JVI.73.5.3986-3993.1999

Da Silva Teixeira MF, Lambert V, Mselli-Lakahl L *et al.* Immortalization of caprine fibroblasts permissive for replication of small ruminant lentiviruses. *Am J Vet Res* 1997;**58**:579–84. https://doi.org/10.2460/ajvr.1997.58.06.579

Davenport KM, Massa AT, Bhattarai S *et al.* Characterizing genetic regulatory elements in ovine tissues. *Front Genet* 2021;**12**:628849. https://doi.org/10.3389/fgene.2021.628849

Davenport KM, Bickhart DM, Worley K *et al.* An improved ovine reference genome assembly to facilitate in-depth functional annotation of the sheep genome. *GigaScience* 2022;**11**:giab096. https://doi.org/10.1093/gigascience/giab096

De Las Heras M, Ortín A, Cousens C *et al.* Enzootic nasal adenocarcinoma of sheep and goats. In: Fan H (ed.), *Jaagsiekte Sheep Retrovirus and Lung Cancer*, pp. 201–23. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003.

DeMartini JC, Bishop JV, Allen TE *et al.* Jaagsiekte sheep retrovirus Proviral clone JSRVJS7, derived from the JS7 lung tumor cell line, induces ovine pulmonary carcinoma and is integrated into the surfactant protein a gene. *J Virol* 2001;**75**:4239–46. https://doi.org/10.1128/JVI.75.9.4239-4246.2001

Griffiths DJ, Martineau HM, Cousens C. Pathology and pathogenesis of ovine pulmonary adenocarcinoma. *J Comp Pathol* 2010;**142**:260–83. https://doi.org/10.1016/j.jcpa.2009.12.013

He Y, Zhang Q, Wang J *et al.* Full-length genome sequence analysis of enzootic nasal tumor virus isolated from goats in China. *Virol J* 2017;**14**:141. https://doi.org/10.1186/s12985-017-0795-4

Huang Q, Ye C, Chen T *et al.* EvaGreen-based real-time PCR assay for sensitive detection of enzootic nasal tumor virus 2. *Mol Cell Probes* 2019;**44**:51–6. https://doi.org/10.1016/j.mcp.2019.02.003

Hudachek SF, Kraft SL, Thamm DH *et al.* Lung tumor development and spontaneous regression in lambs coinfected with Jaagsiekte sheep retrovirus and ovine lentivirus. *Vet Pathol* 2010;**47**:148–62. https://doi.org/10.1177/0300985809352787

Jamil A, Yu Z, Wang Y *et al.* Tick-borne rickettsia, Anaplasma, Theileria, and enzootic nasal tumor virus in ruminant, PET, and poultry animals in Pakistan. *Front Microbiol* 2024;**15**:1359492. https://doi.org/10.3389/fmicb.2024.1359492

Karagianni AE, Vasoya D, Finlayson J *et al.* Transcriptional response of ovine lung to infection with Jaagsiekte sheep retrovirus. *J Virol* 2019;**93**:e00876–19. https://doi.org/10.1128/JVI.00876-19

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–80. https://doi.org/10.1093/molbev/mst010

Kent WJ. BLAT—the BLAST-like alignment tool. *Genome Res* 2002;**12**:656–64. https://doi.org/10.1101/gr.229202

Li H, Durbin R. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* 2009;**25**:1754–60. https://doi.org/10.1093/bioinformatics/btp324

Li Y, Niu J, Liu Y *et al.* Genomic sequencing and analysis of enzootic nasal tumor virus type 2 provides evidence for recombination within the prevalent Chinese strains. *Vet Sci* 2024;**11**:248. https://doi.org/10.3390/vetsci11060248

Li P, Cao X, Wu J *et al.* Detection and analysis of enzootic nasal tumor virus 2 in China. *J Vet Diagn Invest* 2025a;**37**:244–51. https://doi.org/10.1177/10406387241310204

Li P, Yin H, Cao X *et al.* Development and application of a TaqMan-based qPCR assay for detecting ENTV-2 in goats. *Genes (Basel)* 2025b;**16**:529. https://doi.org/10.3390/genes16050529

Ndione MHD, Ndiaye EH, Dieng M *et al.* Mosquito-based detection of retroviruses and arboviruses in Senegal: expanding the scope of xeno-surveillance. *One Health Outlook* 2025;**7**:32. https://doi.org/10.1186/s42522-025-00155-7

Nguyen L-T, Schmidt HA, von Haeseler A *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;**32**:268–74. https://doi.org/10.1093/molbev/msu300

Ortín A, Minguijón E, Dewar P *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 or sheep pulmonary adenomatosis. *Vet Immunol Immunopathol* 1998;**61**:229–37. https://doi.org/10.1016/S0165-2427(97)00149-9

Ortín A, Cousens C, Minguijón E *et al.* Characterization of enzootic nasal tumour virus of goats: complete sequence and tissue distribution. *J Gen Virol* 2003;**84**:2245–52. https://doi.org/10.1099/vir.0.19125-0

Palmarini M, Sharp JM, de las Heras, M., and Fan, H. Jaagsiekte sheep retrovirus is necessary and sufficient to induce a contagious lung cancer in sheep. *J Virol* 1999;**73**:6964–72. https://doi.org/10.1128/JVI.73.8.6964-6972.1999

Palmarini M, Hallwirth C, York D *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**:8065–76. https://doi.org/10.1128/JVI.74.17.8065-8076.2000

Riocreux-Verney B, Verneret M, Diesler R *et al.* Association between genetic clades and cancer prevalence suggested by French-wide study of oncogenic small ruminant β-retrovirus diversity. *Front Cell Infect Microbiol* 2024;**14**. https://doi.org/10.3389/fcimb.2024.1466333

Robinson JT, Thorvaldsdóttir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;**29**:24–6. https://doi.org/10.1038/nbt.1754

Salavati M, Caulton A, Clark R *et al.* Global analysis of transcription start sites in the new ovine reference genome (oar Rambouillet v1.0). *Front Genet* 2020;**11**:580580. https://doi.org/10.3389/fgene.2020.580580

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. https://doi.org/10.1084/jem.97.5.695

Shi W, Jia S, Guan X *et al.* A survey of jaagsiekte sheep retrovirus (JSRV) infection in sheep in the three northeastern provinces of China. *Arch Virol* 2021;**166**:831–40. https://doi.org/10.1007/s00705-020-04919-6

Shuying L, Xueen M, Jingwei Q *et al.* Cloning and sequencing analysis of the complete genome of jaagsiekte sheep retrovirus inner Mongolia strain. *Zhongguo Bing du xue=Virologica Sinica* 2006;**21**:443–8.

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

Summers C, Neill W, Dewar P *et al.* Systemic immune responses following infection with Jaagsiekte sheep retrovirus and in the terminal stages of ovine pulmonary adenocarcinoma. *J Gen Virol* 2002;**83**:1753–7. https://doi.org/10.1099/0022-1317-83-7-1753

Verneret M, Leroux C, Faraut T *et al.* A genome-wide study of ruminants uncovers two endogenous retrovirus families recently active in goats. *Mob DNA* 2025;**16**:4. https://doi.org/10.1186/s13100-024-00337-6

Walsh SR, Linnerth-Petrik NM, Laporte AN *et al.* Full-length genome sequence analysis of enzootic nasal tumor virus reveals an unusually high degree of genetic stability. *Virus Res* 2010;**151**:74–87. https://doi.org/10.1016/j.virusres.2010.04.002

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

Yadav EJ, Prabhakar YK, Shanmugam G *et al.* Whole genome characterization of Jaagsiekte sheep retrovirus isolated from OPA infected sheep (India) and structural assessment (In-silico) of proteins of JSRV. *The Microbe* 2024;**3**:100068. https://doi.org/10.1016/j.microb.2024.100068

Ye C, Huang Q, Chen T *et al.* First detection and genotypic analysis of goat enzootic nasal tumor virus 2 in Chongqing, China. *Arch Virol* 2019;**164**:1647–50. https://doi.org/10.1007/s00705-019-04211-2

York DF, Vigne R, Verwoerd DW *et al.* Nucleotide sequence of the jaagsiekte retrovirus, an exogenous and endogenous type D and B retrovirus of sheep and goats. *J Virol* 1992;**66**:4930–9. https://doi.org/10.1128/jvi.66.8.4930-4939.1992

Zhai S-L, Lv D-H, Xu Z-H *et al.* A novel enzootic nasal tumor virus circulating in goats from southern China. *Viruses* 2019;**11**:956. https://doi.org/10.3390/v11100956

Zhang K, Kong H, Liu Y *et al.* Diagnosis and phylogenetic analysis of ovine pulmonary adenocarcinoma in China. *Virus Genes* 2014;**48**:64–73. https://doi.org/10.1007/s11262-013-0988-x

Zhang X-L, Deng Y-P, Yang T *et al.* Metagenomics of the midgut microbiome of Rhipicephalus microplus from China. *Parasit Vectors* 2022;**15**:48. https://doi.org/10.1186/s13071-022-05161-6

Zhang Y, Cheng T-Y, Liu G-H *et al.* Metagenome reveals the midgut microbial community of Haemaphysalis qinghaiensis ticks collected from yaks and Tibetan sheep. *Parasit Vectors* 2024;**17**:370. https://doi.org/10.1186/s13071-024-06442-y

Zhou K, Zhang L, Li W *et al.* Metagenomic sequencing of epidemic pathogens in Ornithodoros lahorensis from southern Xinjiang. *Medycyna Weterynaryjna* 2026;**82**:135–41. https://doi.org/10.21521/mw.7113

江锦秀, 张靖鹏, 林裕胜 *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**.

© The Author(s) 2026. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.
Virus Evolution, 2026, **12**, veag028
https://doi.org/10.1093/ve/veag028
Research Article