VIRUS EVOLUTION

Virus Evolution, 2023, **9(1)**, 1–12
DOI: https://doi.org/10.1093/ve/vead005
Advance access publication date: 18 January 2023
Research Article

Emergence and pandemic spread of small ruminant lentiviruses

Maria-Luisa Carrozza,[1] Anna-Maria Niewiadomska,[2] Maurizio Mazzei,[3,†] Mounir R. Abi-Said,[4] Stéphane Hué,[5,‡] Joseph Hughes,[6,§] Anna Gatseva,[6,**] and Robert J. Gifford[6,††,*]

[1]Scuola Normale Superiore, Piazza dei Cavalieri 7, Pisa 56126, Italy, [2]Virus Pathogen Resource, J. Craig Venter Institute, 9605 Medical Center Drive, Suite 150, Rockville, MD 20850, USA, [3]Department of Veterinary Science, University of Pisa, Viale delle piagge 2, Pisa 56124, Italy, [4]Faculty of Sciences II, Lebanese University, Campus Pierre Gemayel Fanar, Jdeidet 90656, Lebanon, [5]London School of Hygiene and Tropical Medicine, Room 13, Keppel Street, London WC1E 7HT, UK and [6]MRC - Centre for Virus Research, Sir Michael Stoker Building, 464 Bearsden Road, Glasgow G61 1QH, UK
[†]https://orcid.org/0000-0003-3797-8320
[‡]https://orcid.org/0000-0002-8580-6905
[§]https://orcid.org/0000-0003-2556-2563
[**]https://orcid.org/0000-0003-1601-5277
[††]https://orcid.org/0000-0003-4028-9884
*Corresponding author: E-mail: robert.gifford@glasgow.ac.uk

Abstract

Small ruminant lentiviruses (SRLVs) cause chronic, persistent infections in populations of domestic sheep (*Ovis aries*) and goats (*Capra hircus*) worldwide. The vast majority of SRLV infections involve two genotypes (A and B) that spread in association with the emergence of global livestock trade. However, SRLVs have likely been present in Eurasian ruminant populations since at least the early Neolithic period. Here, we use phylogenetic and phylogeographic approaches to reconstruct the origin of pandemic SRLV strains and infer their historical pattern of global spread. We constructed an open computational resource ('Lentivirus-GLUE') via which an up-to-date database of published SRLV sequences, multiple sequence alignments (MSAs), and sequence-associated metadata can be maintained. We used data collated in Lentivirus-GLUE to perform a comprehensive phylogenetic investigation of global SRLV diversity. Phylogenies reconstructed from genome-length alignments reveal that the deep divisions in the SRLV phylogeny are consistent with an ancient split into Eastern (A-like) and Western (B-like) lineages as agricultural systems disseminated out of domestication centres during the Neolithic period. These findings are also consistent with historical and phylogeographic evidence linking the early 20th century emergence of SRLV-A to the international export of Central Asian Karakul sheep. Investigating the global diversity of SRLVs can help reveal how anthropogenic factors have impacted the ecology and evolution of livestock diseases. The open resources generated in our study can expedite these studies and can also serve more broadly to facilitate the use of genomic data in SRLV diagnostics and research.

Key words: retrovirus; lentivirus; evolution; emergence; iatrogenesis

Introduction

Small ruminant lentiviruses (SRLVs) are retroviruses (family *Retroviridae*) that cause chronic, persistent infections in domestic sheep (*Ovis aries*) and goats (*Capra hircus*) worldwide, resulting in considerable economic losses (Keen et al. 1997). There are multiple possible routes of SRLV transmission, but the most important are thought to be ingestion of infected colostrum and/or milk and inhalation of respiratory secretions (Blacklaws et al. 2004). SRLV infection causes multi-system disease—typical symptoms include pneumonia, wasting, paralysis, polyarthritis, and mastitis. Disease manifestations can ultimately lead to organ failure and death. However, infection is usually only apparent following an incubation period of at least 3–4 years, and lifelong subclinical infections often occur (Thormar 2005; Blacklaws 2012). General susceptibility and specific disease symptoms are influenced by genetic factors and vary between small ruminant species and breeds (Straub 2004; Herrmann-Hoesing et al. 2008; Highland 2017).

The study of SRLVs played an important role in the discovery of an association between persistent viral infections and chronic progressive diseases. In the 1930s a wasting disease of sheep, called '*maedi-visna*', emerged in Iceland. Sustained efforts to understand the cause of this disease ultimately led to the isolation of '*ovine maedi-visna virus*' (OMVV) in the 1950s (Sigurdsson 1954). A virus closely related to OMVV was isolated from a North American dairy goat in 1974 and named '*caprine arthritis-encephalitis virus*' (CAEV) (Crawford et al. 1980). It has subsequently become clear that each of these viruses occurs in both sheep and goats, and they are now considered to represent two distinct genotypes (A and B respectively) of a single virus species (SRLV) (Shah et al. 2004).

The A and B genotypes are widespread in sheep and goat populations and have been reported in countries throughout the world (Adams et al. 1984; Shah et al. 2004). Subsequently, additional, diverse genotypes (B3, C, and E) have been identified, which appear to have relatively restricted distributions in Europe and Western Asia (Gjerset, Storset, and Rimstad 2006; Grego et al. 2009; Bertolotti et al. 2011).

Genomic data show that lentiviruses (genus *Lentivirus*) have circulated in mammals for millions of years (Gifford 2012), and SRLVs likely have similarly ancient associations with small ruminants. However, uncertainty surrounds the deeper origins of the pandemic SRLV strains and their longer-term, ancestral distribution among domestic ruminant populations and breeds (Gifford 2012; Molaee et al. 2020). Notably, SRLVs do not appear to have spread in parallel with the first European colonists during the 'Age of Discovery' as seems to be the case for other livestock pathogens (Smith et al. 2009; Capomaccio et al. 2012). This suggests that SRLV infection was not widespread in Europe during the 15th—17th century when the first exports of European small ruminants to Africa, Oceania, and the Americas occurred. Consistent with this, the phylogenetic analysis of SRLVs indicates that the pandemic spread of the A and B genotypes occurred relatively recently, enabled by international livestock trade (Shah et al. 2004).

Archaeological and genetic evidence show that sheep and goats were domesticated in the 'Fertile Crescent' region of Western Asia 10–11 thousand years ago (Zeder 2008). Domesticated animals disseminated into Europe, Asia, and Africa during the Neolithic period, in parallel with the spread of agro-pastoralism (Amills, Capote, and Tosser-Klopp 2017; Taylor et al. 2021). Subsequently, vicariance and selective breeding gave rise to region-specific sheep and goat breeds (Ryder 1983).

In Europe, sheep diversity is thought to reflect two waves of diffusion—the first involved more primitive breeds reared primarily for meat, and the second involved breeds specialised for secondary products such as wool and milk. In most areas, primitive breeds have been replaced by (or subsumed into) the 'improved' breeds that spread during the second wave (Chessa et al. 2009). The presence of a divergent SRLV genotype (C) in isolated, remnant populations of primitive sheep is consistent with the notion that SRLVs spread into Europe in parallel with early Neolithic agricultural systems, although other explanations cannot currently be ruled out (Gjerset, Storset, and Rimstad 2006).

In Asia and North Africa, most native sheep are 'fat-tailed' breeds that are genetically and phenotypically distinct from sheep found in Europe (Ryder 1983). The longer-term historical associations of fat-tailed sheep with SRLVs are unknown. However, they are implicated in the emergence of SRLV-A because the Icelandic outbreak from which the prototypic SRLV-A isolate was obtained was unambiguously traced to a 1933 import of 'Karakul' sheep (Thormar 2013). Karakul sheep are an ancient, fat-tailed breed native to Central Asia (Ryder 1983; Rocha et al. 2011). The Karakul that were imported to Iceland in the 1930s were obtained from a flock maintained at the Institute for Animal Breeding in Halle, Germany (Thormar 2013), but this flock had only been established quite recently (in 1903) by importing animals directly from Bukhara, Uzbekistan (Saraiva 2013). Shortly prior to this, no European populations of Karakul existed.

During the early 20th century, however, Karakul breeding briefly became a focus of intense interest in several European countries, primarily due to the breed's suitability for rearing in arid regions of colonial Africa (Saraiva 2013). In 1907, the Karakul from the Halle flock were transported to German South West Africa (present day Namibia) where they quickly became the basis of a highly profitable pelt industry (Krogh 1955; Saraiva 2013). Following this, experiments with Karakul farming took place in several European countries, including Italy (Saraiva 2013), Switzerland (Schmid 1934), Greece (Seimenis et al. 1982), and the UK (Holm 1916). Exports of Uzbek Karakul to the USA and Canada also took place around this time (Young 1914). From ~1920 onwards, however, changing constraints on international animal trade effectively prevented further export of Karakul sheep from Central Asia (Saraiva 2013).

Acknowledging that the available historical evidence is to some extent circumstantial, there appears to be a striking association between the dissemination of Karakul sheep into Europe, Africa, and North America and the early emergence of SRLV-associated diseases (Fig. 1).

In this study we use molecular sequence data to explore the origin and history of the SRLV pandemic. We reveal phylogenetic evidence for an ancient divergence among SRLVs in Eastern versus Western regions of the Eurasian continent and show how this supports a role for Karakul in the emergence of SRLV-A. In addition, we construct an openly accessible, data-oriented resource for comparative genomic analysis of lentiviruses ('Lentivirus-GLUE') that can facilitate further investigation of SRLV history while also supporting efforts to control and monitor the SRLV spread.

Results

Construction of open resources for comparative genomic analysis of SRLVs

We used a previously developed bioinformatics software framework (GLUE—Genes Linked by Underlying Evolution) (Singer et al. 2018) to create Lentivirus-GLUE, an openly accessible online resource for comparative analysis of lentivirus genomes. Lentivirus-GLUE not only contains the items required for comparative analysis—e.g. lentivirus sequences, MSAs, and isolate-associated data—it also maps the semantic links between these items via a relational database (Fig. S1) (Gifford 2022). The Lentivirus-GLUE resource is instantiated using GLUE's native command layer and a project-specific build file. GLUE's command layer can then be used to co-ordinate interactions between the Lentivirus-GLUE project database and bioinformatics software tools (e.g. Basic Local Alignment Search Tool (BLAST); (Camacho et al. 2009), Multiple Alignment using Fast Fourier Transform; (Katoh and Standley 2013), and Randomized Axelerated Maximum Likelihood (RAxML); (Stamatakis 2006)). Selection of sequence data for analysis can be conditioned on isolate-associated data (e.g. host species and country/year of isolation) or alignment-associated information (e.g. genome feature coverage). This computational framework allows comparative genomic analysis to be performed in a reproducible, standardised way. It supports the reuse of data curated in this study and can also provide a flexible foundation for the development of information systems that use comparative genomic approaches to inform SRLV diagnostics and surveillance.

To support a lightweight, flexible approach to project development, Lentivirus-GLUE has a multi-layered structure. The 'base layer' represents the genomic diversity of exogenous lentivirus species using a minimal set of components, while an SRLV-focussed extension layer represents all published SRLV sequences ($n = 4{,}653$). The base project contains (1) a set of six reference genome sequences each representing a distinct exogenous lentivirus species (Table S1); (2) a standardised set of sixty-six lentivirus genome features (Table S2); (3) genome annotations specifying the co-ordinates of these genome features within

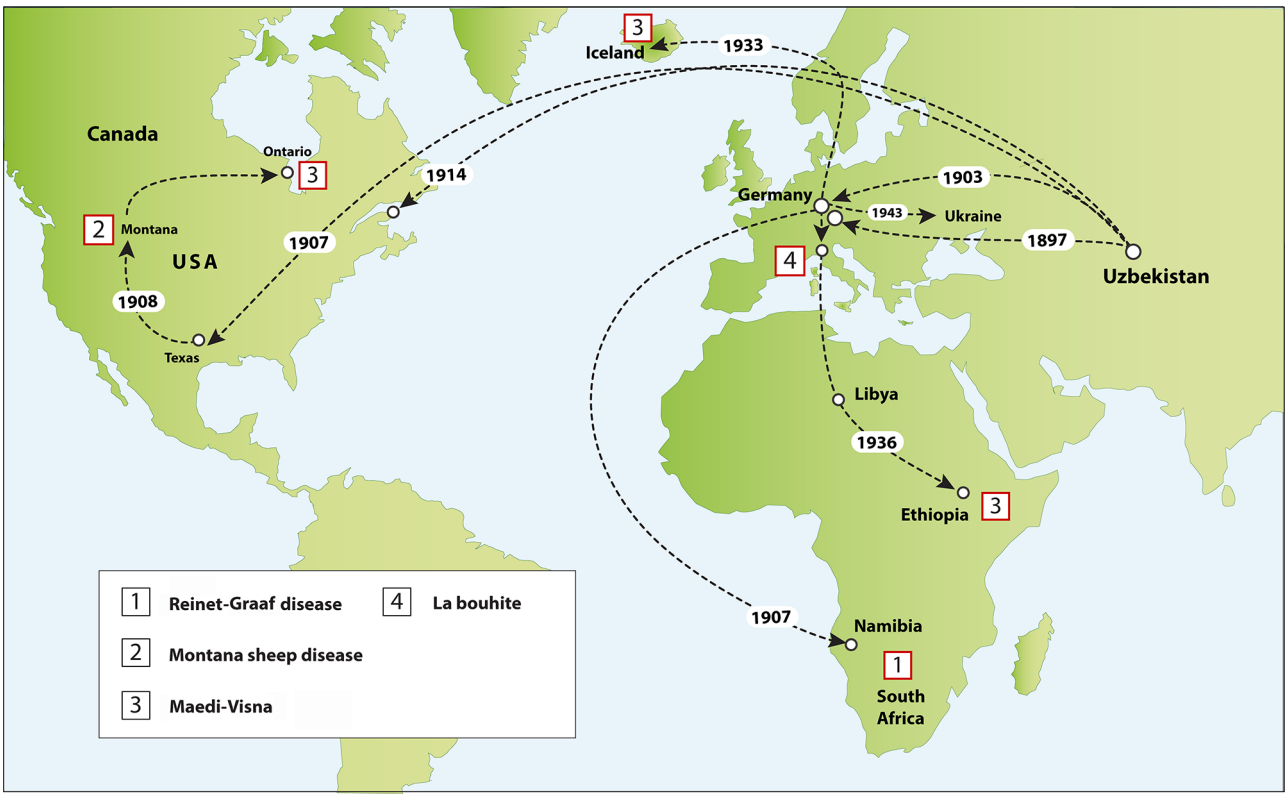

Figure 1. Dissemination of Karakul and early SRLV outbreaks. Arrows on the map indicate documented international exportations of Karakul during the period 1900–70. During this period, a cluster of similar disease syndromes emerged in sheep, all of which are either known or considered likely to have been caused by SRLV-A infection (see inset key). Their geographic locations are approximately indicated on the map following the key. '*Graaff-Reinet disease*' was first reported in South Africa in 1915 (De Kock 1929), and the same year, '*Montana sheep disease*' was reported for the first time in the USA (Marsh 1923). In Europe, '*maedi-visna*' (Sigurdsson 1954) and '*La Bouhite*' (France; Lucam 1942) both emerged during the 1930s and 40s. Further outbreaks of *maedi-visna* were reported in Canada in 1956 and in East Africa (Kenya) in 1966 (Table S4).

reference genome sequences (Gifford 2022); and (4) a hierarchically linked set of MSAs representing inferred sequence homology across distinct taxonomic levels within the *Lentivirus* genus. This 'alignment tree' links MSAs in a way that reflects their evolutionary relationships and thereby maximises the representation of homology among all lentivirus sequences (Fig. 2, Table 1).

The SRLV extension layer incorporates all published SRLV sequence data (Table S3) linked to associated metadata, including information that is (1) sequence-associated (length, publication date); (2) taxonomic (viral genotype, subtype); and (3) isolate-associated (isolation host species, date and location of sampling, isolation source). We also collated information on the breed associations of early SRLV outbreaks (Tables S4 and S5).

Phylogenetic analysis of complete SRLV genomes reveals an ancient East–West split

We used Lentivirus-GLUE to investigate the history of SRLV spread, reconstructing maximum likelihood (ML) phylogenies for a range of distinct alignment partitions (Fig. 2, Table 1). The majority of SRLV isolates in our dataset was only represented by sub-genomic sequences (Table S3). We found that inclusion of these sequences in the phylogenetic analysis revealed little about the deeper evolutionary origins of SRLVs, as the internal nodes of resultant phylogenies typically had only weak support (Fig. S2A–I). Furthermore, via analysis of complete genome sequences we found that the structure of the SRLV epidemic has been shaped by ancestral as well as recent recombination (Fig. S3, Table S6). Thus, complete SRLV genome sequences are required to reveal the longer-term evolutionary history of SRLVs.

Phylogenetic investigation across all lentiviruses revealed that the bovine lentiviruses are most closely related to the SRLVs (Fig. 3A). We therefore used bovine immunodeficiency virus (BIV) as an out-group taxon for rooting SRLV phylogenies. The divergent 'E' genotype of SRLVs was observed to group basally in trees reconstructed from an alignment of Gag–Pol proteins and rooted on BIV (Fig. 3B). We therefore used SRLV-E to root phylogenies constructed from a nucleotide-level, genome-length alignment of eighty-four SRLV isolates (Fig. 3C–E). In these phylogenies, robust support is obtained for a deep ancestral division of Eurasian SRLVs into two major sub-lineages—an Eastern-origin 'A-like' clade and a Western-origin 'B-like' clade that includes genotypes B, B3, and C (Fig. 3C–E). Both of these clades contain taxa sampled in Eastern and Western regions of the Eurasian continent, as well as in other continents throughout the world. Crucially, however, the basal taxa in each clade reflect SRLV diversity found near the extreme Eastern and Western ends of the Eurasian continent. Among A-like SRLVs, basal isolates are linked to regions where fat-tailed breeds have traditionally predominated and include one isolate 'NM1111' sampled in the Inner Mongolia region of China (Zhao et al. 2021) (Fig. 3D). Among B-like SRLVs, the most basal lineage is genotype C, which is found only in primitive sheep in Northwestern Europe (Norway) (Fig. 3E). By contrast, all 'non-Asian A-like' viruses and all 'non-European B-like' isolates occur in derived clusters with relatively short branch lengths—yet incorporating isolates from geographically distant

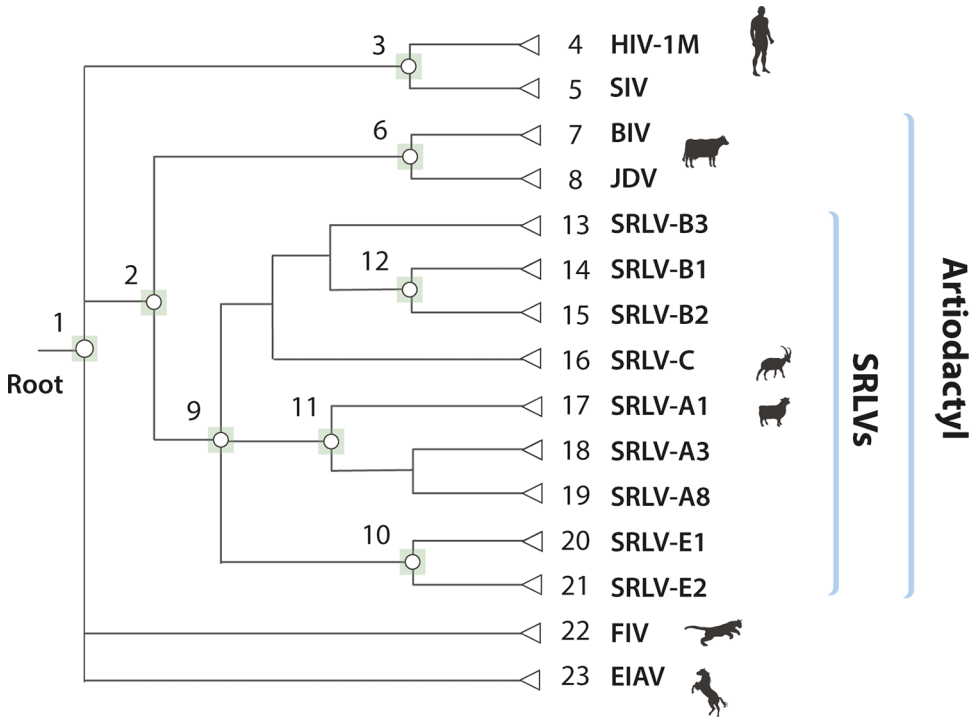

Figure 2. The MSA hierarchy in Lentivirus-GLUE. The GLUE software framework allows hierarchical linking of MSAs to reflect taxonomic relationships (Singer et al. 2018). The figure shows a schematic representation of the hierarchical 'MSA tree' data structure in Lentivirus-GLUE, which recapitulates the relationships among lentivirus sequences. Numbers shown adjacent to nodes/tips correspond to rows in Table 1. Abbreviations: SIV = simian immunodeficiency virus; JDV = Jembrana disease virus; FIV = feline immunodeficiency virus; EIAV = equine infectious anaemia virus.

countries—consistent with recent introduction via international livestock trade (Shah et al. 2004) (Fig. 3D, E).

Phylogenetic evidence supports a role for Karakul export in the emergence of SRLV-A

The presence of distinct 'A-like' and 'B-like' clades in SRLV phylogenies, each with ancestral associations in Western versus Eastern Eurasia, respectively, lends support to the notion that SRLV-A could have been initially introduced into Europe, Africa, and the Americas via Karakul exports from Central Asia (Fig. 1). Furthermore, full-genome phylogenies show European, African, and North American isolates clustering in robustly supported, derived clades, each of which could plausibly correspond to founder events accompanying Karakul introductions in these regions (Fig. 3D). However, the number of founder events and their timing remain uncertain. To investigate this further, we cross-referenced between phylogenies reconstructed using complete genome sequences and those reconstructed using sub-genomic regions, but including a more broadly sampled range of taxa (Fig. S2A–E). While these phylogenies reveal relatively little about the deeper origins of SRLV genotypes (as discussed earlier), they can provide insights into more recent spread.

All European, African, and American isolates of SRLV-A form an in-group relative to basal, Asian isolates. However, the North American clade is clearly separate, consistent with distinct sampling from a source population. There are no complete genome sequences from Canadian SRLV-A isolates, but phylogenies support a common origin with the SRLV-A strains found in the USA (Fig. S2C). In early 20th century North America interest in Karakul focussed on cross-breeding with local breeds, such as Lincoln Longwool, to improve fleece and mutton quality (Hunter 2016), but these experiments had mixed results and the interest waned from ~1950 onwards. Consequently, North American Karakul are now relatively rare and their history has been quite well recorded (Hunter 2016). Intriguingly, SRLVs seem to be prevalent in areas where Karakul herds have been introduced or maintained—e.g. Ontario (Campbell et al. 1994) and Quebec (L'Homme et al. 2011) (Fig. 1).

Among European isolates, robust support was obtained for several sub-clades containing isolates from Iceland, Switzerland and Spain, and Piedmont in Italy. The Piedmont cluster is interesting because this region has a documented history of Karakul farming—during the 1930s–40s the Alessandria Institute of Rabbit Breeding (Piedmont) became a centre of Karakul breeding and production, supplying rams to settler communities in Libya and Italian East Africa. Institute experiments involved cross-breeding with North African fat-tailed breeds and the use of artificial insemination (Saraiva 2013).

Another intriguing observation, although it relies on sub-genomic data, is the extraordinarily close genetic relationship between (1) capsid sequences recently obtained from Polish SRLV isolates and (2) the prototypic OMVV isolate from the Icelandic outbreak of the 1930s–50s (Fig. S4)—this observation is hard to explain without an explicit connection. Interestingly, wartime records describe how, during the 1940s, the Halle Karakul flock (the source of the Icelandic outbreak) was transported East as they were incorporated into German plans for the settlement of Poland and Ukraine (Heim 2008; Saraiva 2013) (Fig. 1).

Within the 'Europe-and-Africa' clade of SRLV-A isolates, the sole African sequence derives from South Africa and falls into a derived clade with sequences from the UK and Belgium (Fig. 3D). Notably, several of the European countries represented in this paraphyletic clade have colonial history in Africawith Italy and Portugal also having attempted to establish Karakul farming in

Table 1. The MSA hierarchy in Lentivirus-GLUE.

No.[a]	Taxonomic scope	Name	Parent[b]	Children[c]	Constraining reference[d]	Genome coverage[e]	Sequence count
Genus							
1	*Lentivirus*	Root	N/A	5	FIV	*gag–pol*	5
Supergroup							
2	Artiodactyl	Artiodactyl		2			2
Subgroup							
3	Primate lentiviruses	Primate	Root	0	SIVcpz		1
6	Bovine lentiviruses	Bovine	Artiodactyl	0	BIV	Genome	2
7	SRLVs	Ovine–caprine	Artiodactyl	1	SRLV-A	Genome	5
Species							
4	HIV	HIV	Primate		HIV-1	Genome	1
5	SIV	SIV	Primate		SIV	Genome	1
7	BIV	BIV	Bovine		BIV	Genome	1
8	JDV	JDV	Bovine		JDV	Genome	1
9	SRLV	SRLV	Ovine–caprine	5	SRLV-A	Genome	5
23	Equine lentiviruses	Equine	Root	0	EIAV	Genome	1
22	Feline lentiviruses	Feline	Root	0	FIV	Genome	2
Clade							
	SRLV-B + SRLV-C	SRLC-BC	Ovine–caprine	4	SRLV-C	Genome	4
Genotype							
10	SRLV-E	SRLV-E	Ovine–caprine	2	SRLV-E	Genome	2
11	SRLV-A	SRLV-A	Ovine–caprine	3	SRLV-A	Genome	3
12	SRLV-B	SRLV-B	SRLC-BC	3	SRLV-B	Genome	3
16	SRLV-C	SRLV-C	SRLC-BC	0	SRLV-C	Genome	67
Subtype							
14	SRLV-B1	SRLV-B1	SRLC-BC	0	SRLV-B1	Genome	1,212
15	SRLV-B2	SRLV-B2	SRLC-BC	0	SRLV-B2	Genome	113
14	SRLV-B3	SRLV-B3	SRLC-BC	0	SRLV-B3	Genome	79
17	SRLV-A1	SRLV-A1	SRLV-A	0	SRLV-A1	Genome	464
18	SRLV-A3	SRLV-A3	SRLV-A	0	SRLV-A3	Genome	1,870
19	SRLV-A8	SRLV-A8	SRLV-A	0	SRLV-A8	Genome	229
21	SRLV-E1	SRLV-E1	SRLV-E	0	SRLV-E1	Genome	38
22	SRLV-E2	SRLV-E2	SRLV-E	0	SRLV-E2	Genome	18

[a] Numbers correspond to labelled nodes in Fig. 2. [b] Name of the parent MSA within the hierarchy. [c] Number of child MSAs belonging to this MSA tree node. [d] Reference sequence that constrains the genomic co-ordinate space in the MSA. [e] Regions of the constraining reference genome spanned by the MSA. Note that the GLUE framework allows parent MSAs to be recursively populated with member sequences from any or all of their child alignments, so that for example the root alignment can be exported with all taxa above the root included.

their African colonies (Saraiva 2013). However, present data do not allow us to ascertain whether the sole African isolate derives directly from SRLV-A imported to Africa in 1907 with German Karakul or with later imports of European sheep.

Unfortunately, we lack any sequence data from the outbreaks of maedi-visna reported in the highlands of Ethiopia (Ayelet et al. 2001; Mekonnen, Sirak, and Chaka 2010; Woldemeskel and Tibbo 2010) or elsewhere in East Africa (Wandera 1970; Rwambo, Brodie, and DeMartini 2001). These are described as emerging in association with 'exotic sheep' but might not derive from Italian efforts to establish Karakul farming in East Africa in the 1930s (Fig. 1). Rather, they could have arisen in association with later efforts to improve Ethiopian sheep stocks via cross-breeding with animals imported from the UK and Israel. These imports occurred in the 1970s, by which time SRLV infection had spread far and wide among small ruminant breeds and was prevalent throughout much of Europe and the Middle East (Straub 2004) (Tables S4 and S5).

We used Bayesian phylogeographic approaches to estimate the timing of SRLV-A spread. However, only two partitions were amenable to Bayesian analysis—those corresponding to the reverse transcriptase (*RT*) and integrase (*IN*) genes. Both yielded 'time-to-most-recent-common-ancestor' (tMRCA) estimates in the late 19th century—shortly predating the first Karakul exports from Central Asia (Fig. S5, Table S7).

Ancestors of pandemic SRLV-B strains are associated with Central Europe

Phylogenetic analyses of all available sequence data suggest that initial dissemination of SRLV-B involved a European source population—the deeper nodes in SRLV-B phylogenies were placed in Central Europe, whereas clades sampled in Asia, South America, and around the northern and western periphery of Europe (i.e. Ireland, Norway, and Finland) have more recent origins and reflect country-specific radiation of imported viruses (Fig. S2G–I). Outbreaks of SRLV-B-associated disease emerge from ~1960 onwards, and consistent with this, estimates based on Bayesian phylogeographic analysis indicate that the dissemination of SRLV-B1 stemmed from a Central European source in the mid-20th century (Fig. S6, Table S8). We obtained tMRCA estimates of 1966 for MSA partitions based on the *RT* gene (1958–74, 95 per cent highest probability density (HPD)) and 1952 for *IN* (1939–64, 95 per cent HPD). Interestingly, epidemiological data show that the pandemic spread of SRLV-B1 is conspicuously associated with Saanen goats—a popular Swiss dairy breed (Table S5)—consistent with other data implicating a Central European source. However, as noted by Shah et al. (2004), the ultimate origins of this virus strain may lie in other Central European small ruminant species (e.g. chamois goats). Furthermore, it should be noted that Switzerland is over-represented in our dataset, and thus the placement of MRCAs in Central Europe should be treated with some caution.

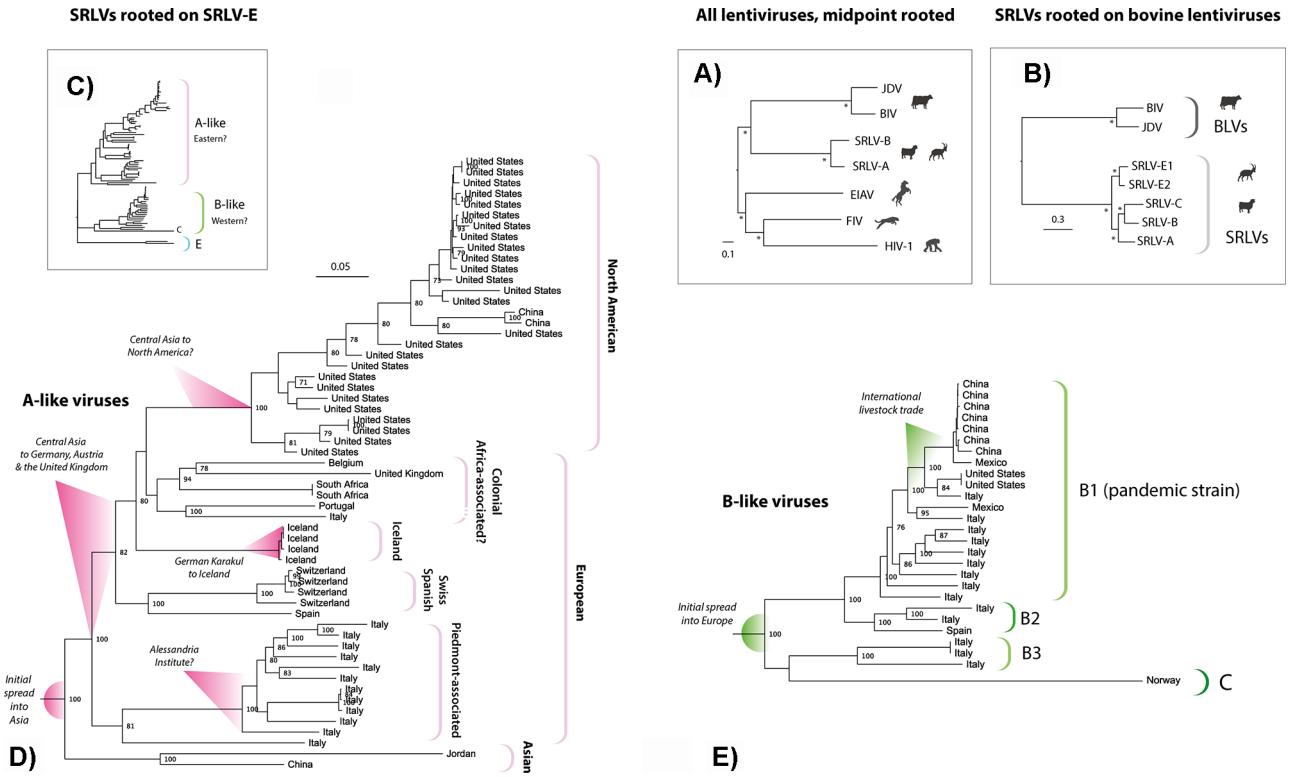

Figure 3. Rooted SRLV phylogenies support an ancient East–West split. (A) A ML phylogeny showing the evolutionary relationships between lentiviruses. The phylogeny is based on an alignment of Gag–Pol polyproteins spanning 1,397 amino acid residues in the Gag–Pol open reading frame and the rtREV substitution model. (B) A ML phylogeny showing the evolutionary relationships between SRLV genotypes and rooted on the BLVs. The phylogeny is based on an alignment of Gag–Pol polyproteins spanning 1,598 amino acid residues in Gag–Pol and the rtREV substitution model. (C) A ML phylogeny showing the evolutionary relationships between eighty-five genome-length SRLV sequences (8,491 nucleotide residues), constructed using the general time reversible (GTR) model of nucleotide substitution, and rooted on SRLV-E. Two of the major clades contained within this tree are shown enlarged, with taxa labels, in panels (D) A-like viruses and (E) B-like viruses. Scale bars indicate evolutionary distance in substitutions per site. Asterisks indicate nodes with ML bootstrap support >85 per cent, based on 1,000 replicates. Abbreviations: BLV = bovine lentivirus; SIV = simian immunodeficiency virus; JDV = Jembrana disease virus; FIV = feline immunodeficiency virus; EIAV = equine infectious anaemia virus; rtREV = retrovirus-specific model.

SRLV diversity in the Fertile Crescent region reflects both ancient and modern spread

To investigate the hypothesis that ancient SRLVs spread out of early centres of agriculture along with domesticated ruminant species, we examined the genetic diversity of SRLV strains circulating in the 'Fertile Crescent' region, where some of the earliest human civilisations are believed to have developed (Fig. 4A). We focussed on three sites in Lebanon and two in Jordan (Bar-Yosef 1998) (Fig. 4B). Serological testing of 886 animals established that the SRLV prevalence was ~21 per cent (Table S9), within the range of estimates obtained in neighbouring countries (Sasani et al. 2013; Tolari et al. 2013). We used polymerase chain reaction (PCR) to obtain *gag* gene amplicons from fifteen positive Lebanese samples. In addition, one complete genome, two long terminal repeat (LTR) sequences, and three *gag* sequences were obtained from Jordanian PCR samples.

Phylogenetic analysis revealed three distinct clades comprising Levantine isolates (Lev-I–III). The Lev-I and II clades group robustly with pandemic SRLV-A strains, whereas the Lev-III clade is separate from all previously sampled SRLV genotypes (Fig. 4C). These findings show that sheep and goat flocks in Jordan and Lebanon are infected with globally prevalent subtype A strains as well as with more divergent 'A-like' strains (e.g. Lev-III), characteristic of the Levant and Iran (Tolari et al. 2013; Molaee et al. 2020) (Fig. S2E). Given that connections to the global livestock industry are now well established in the region, it is likely that overall diversity has been influenced by livestock trade in recent decades, as seems to be the case in other regions (Shah et al. 2004).

A strain-level genotyping algorithm for SRLVs

Our primary objective in constructing the SRLV extension layer in Lentivirus-GLUE was to investigate the history of SRLV spread. However, the data contained within this resource can potentially be used in other kinds of investigations, e.g. efforts to improve SRLV surveillance and diagnostics. GLUE projects follow principles of 'data-oriented programming' that aim to expedite the development of information systems by separating code and data (Singer et al. 2018). This approach simplifies the downstream reuse of datasets, irrespective of analysis context.

To illustrate this principle, we used the phylogenies and genome-length alignments collated in Lentivirus-GLUE to implement a ML-based, phylogenetic genotyping method, applicable to any SRLV sequence. Experience with human immunodeficiency virus type 1 (HIV-1) has shown that, due to extensive recombination, lentivirus classification schemes must be based on full-length (or nearly full-length) sequences (Gifford et al. 2007). However, recombination need not preclude the development of a strain-level nomenclature system or the use of subgenomic sequences to perform strain-level classification, provid-

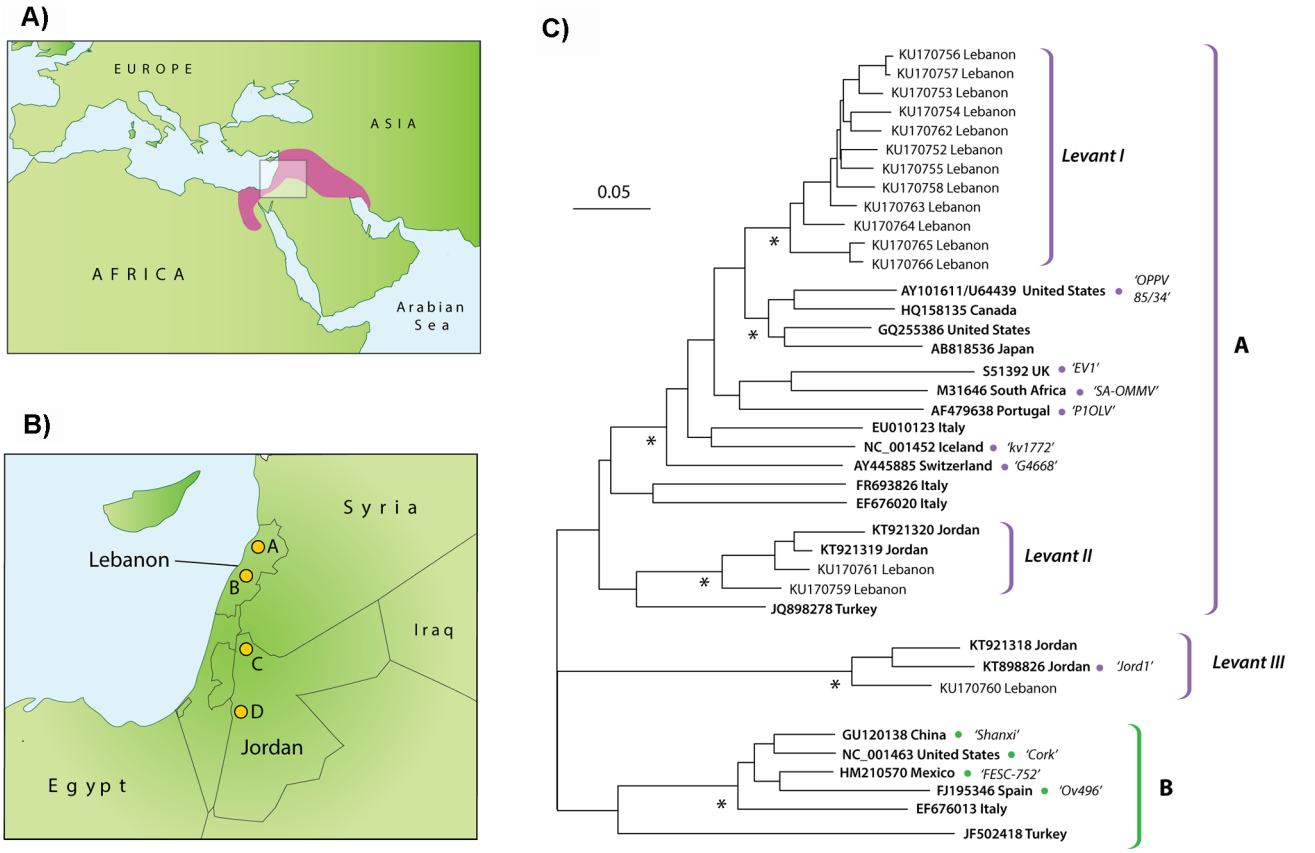

Figure 4. SRLV diversity in the Fertile Crescent region. The map in panel (A) indicates the 'Fertile Crescent' region (shaded). The overlapping square indicates the location of the area shown in panel (B). Circles indicate sampling sites within the Levant region: A = Northern Lebanon (Qornet el Sawda, Arez); B = Bekaa Valley (three sites: North (Aammiq), West (Swairi, Manara, Rashaya), and Eastern (Nahle, Maqne, Knaisse)); D = Northern Jordan and Jordan Valley. (C) ML phylogeny based on an alignment of SRLV capsid(CA) gene nucleotide sequences and showing the evolutionary relationships of Lebanese SRLV isolates to globally sampled SRLVs. The tree was reconstructed using the general time reversible (GTR) model of nucleotide selection as selected using the likelihood ratio test and is midpoint rooted for display purposes. Taxa labels show GenBank accession numbers. Isolate names are shown for complete genome isolates (denoted by coloured circles adjacent taxa labels). Asterisks indicate nodes with ML bootstrap support >85 per cent, based on 1,000 bootstrap replicates. Brackets to the right indicate genotypes and subgroups.

ing that (1) full-length genomes are used to define genotypes/subtypes and (2) genome-scale sequencing is conducted frequently enough to reflect current diversity (Gifford et al. 2007; Rambaut et al. 2020). We therefore used our phylogeny of full-length sequences to define a set of robust SRLV sub-clades. We followed existing nomenclature as far as possible, but only a handful of previously proposed subtype groupings (B1, B2, A1, A3, and A8) are supported when using this more rigorous approach. Sequences in SRLV-GLUE were classified based on these groupings. Genotyping was based on optimal placement of query sequences in our reference phylogeny using RAxML as described by Singer et al. (2018). We compared results obtained using this method with genotype/subtype assignments based on phylogenetic reconstruction (Fig. S2) and found 99 per cent agreement (Fig. S7).

Discussion

Viruses that cause chronic, persistent infections have circulated in animals for millions of years (Gifford 2012; Suh et al. 2014). However, some have only emerged as pandemic pathogens within the far shorter time frame of recorded history (Holmes 2008). Anthropogenic factors are thought to have played a role in enabling the recent emergence of chronic viral infections in humans (Pepin 2011; Faria et al. 2014). Such factors may also have played a role in enabling the emergence of viral infections in domesticated species, particularly livestock, but this has not been investigated in equivalent depth. It is important to understand how anthropogenic factors such as international trade and husbandry practices have impacted livestock disease because this knowledge can underpin the development of more effective interventions. SRLVs present an interesting case because sheep and goats have played an important role in human culture for thousands of years. Thus, disease ecology in domestic small ruminant populations has undoubtedly been shaped by human activities, and the spread of SRLV infection can be examined in this context.

In this study, we used serology, sequencing, and comparative genomics to investigate the origins and history of the SRLV pandemic. Rooted phylogenies reconstructed from analysis of eighty-four complete genome sequences revealed compelling evidence for a division of SRLVs into A-like and B-like lineages, each with clear ancestral geographic associations in Eastern and Western Eurasia, respectively (Fig. 3C–E). These phylogenetic relationships suggest that distinct A-like and B-like SRLV lineages emerged in association with fat-tailed and thin-tailed sheep breeds, respectively. Furthermore, phylogenies are consistent with the proposition that the emergence of SRLV-A in Europe, Africa, and North America was associated with the early 20th century introduction of Central Asian Karakul sheep to these regions (Palsson 1982; Straub 2004) (Fig. 3D, Fig. 5). Phylogenetic trees incorporating

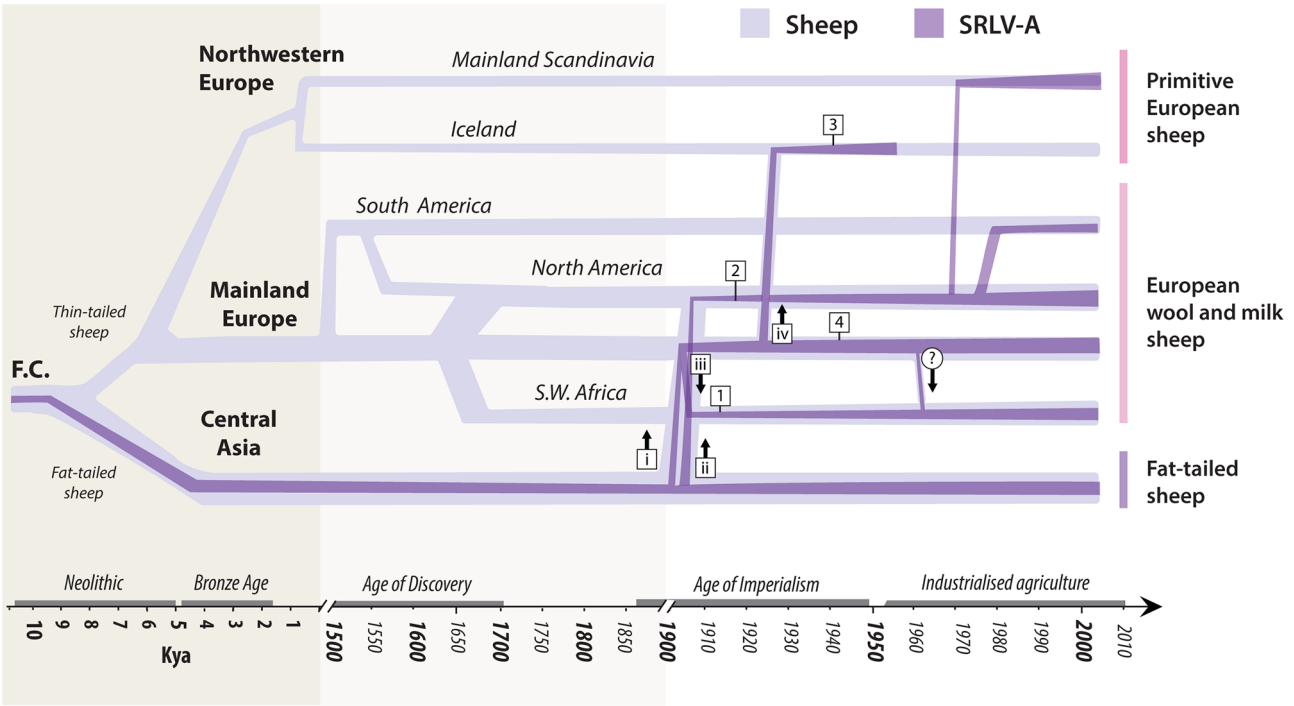

Figure 5. A proposed model of SRLV-A origin and early spread. A schematic representation of our proposed model of SRLV-A evolution. The broken scale bar shows time on three different scales, from left to right as follows: (1) thousands of years (Kya); (2) centuries; and (3) decades. Selected periods of ancient and contemporary human history are indicated. The underlying tree (lighter - see key), represents the historical dissemination of sheep breeds out of the Fertile Crescent (F.C.) region and into other geographic areas as indicated. The overlaid tree (darker) shows the hypothesised pattern of SRLV-A dissemination among global sheep populations. Historical outbreaks of SRLV-associated disease are indicated by numbers in square markers, as follows: (1) Graaff-Reinet disease; (2) Montana sheep disease; (3) Maedi-Visna; and (4) La Bouhite. Squares adjacent arrows containing roman numerals indicate the international export of Karakul stocks between continental regions as follows: (i) Central Asia to Europe; (ii) Central Asia to North America; (iii) Europe to Southwest Africa. (iv) Mainland Europe to The arrowed circle containing the '?' symbol indicates a possible introduction of SRLV-A in Africa in thin-tailed sheep during the 1960s-70s. Please note that, other than this, the figure does not represent the extensive international trade that occurred from ~1960 onwards, only the earlier export events pertinent to our model.

sub-genomic data (Fig. S2A–E) and the Bayesian phylogeographic analysis (Fig. S5) are also consistent with this model.

The geographic isolation of Iceland allowed it to be established beyond reasonable doubt that OMVV, the prototypic SRLV-A isolate, was introduced in the 1930s with Karakul sheep imported from Germany. Despite this, Karakul do not seem to have been considered as the proximal source of European SRLV-A. In fact, a 2003 review article by German veterinarian Otto Straub seems to dismiss the idea based on (1) the lack of reported SRLV infection in the Halle Karakul flock and (2) the fact that SRLV-A infection is not obviously associated with disease in the Karakul breed (Straub 2004). However, current evidence indicates that lentiviruses—including SRLVs (Reina et al. 2009a)—may be relatively apathogenic in populations with which they have longer-term associations (Apetrei et al. 2010). Thus, if Karakul are indeed 'natural' hosts of SRLV-A, asymptomatic infections might be expected. This would account for the lack of obvious disease in farmed Karakul and the failure to detect it during the lengthy quarantine periods that accompanied Karakul importations (Straub 2004). Interestingly, while SRLV infection was never reported in the Halle Karakul flock, a relatively early outbreak (1967) of maedi-visna-like progressive interstitial pneumonia was reported among sheep in Halle (Seffner and Lippmann 1967) (Table S4).

We uncovered evidence of a complex, recombinant history among SRLV strains (Table S6, Fig. S3), indicating that complete virus genome sequences will be required to establish robust links between specific SRLV genotypes and geographic regions. Complete genomes are also required because, due to increasing volumes of international livestock trade from the ~1950 onwards, the overall genetic structure of the SRLV pandemic has become much more heterogeneous over time, confounding efforts to determine the ancestral region and breed associations of SRLV strains. We did not fully appreciate the extent of this issue at the outset of our study because of our use of a strategy based on sub-genomic sequencing when sampling SRLV diversity in the Levant (in 2012).

Molaee et al. proposed a slightly different model of SRLV-A evolution (based on analysis of sub-genomic data) under which the presence of diverse SRLV-A strains in Europe reflects ancient spread along the Danube (Molaee et al. 2020). Currently available data do not allow us to exclude this hypothesis. However, all current data, including the phylogenies presented by Molaee et al. (2020), are compatible with the model of SRLV-A evolution presented in Fig. 5. For example, these authors obtained *gag* and *env* sequences from fat-tailed sheep in Western Asia (Iran) that group robustly with basal, Asian SRLV-A isolates in phylogenetic trees (Fig. S2E) (Molaee et al. 2020; Gifford 2022). This is consistent with their dissemination into Asia during the Neolithic period, along with other SRLV-A ancestors.

It is certainly true that a heterogeneous mix of diverse SRLV-A isolates circulates in Europe—for example, in phylogenies based on genome-length sequences, isolates from Piedmont are fairly distant from those obtained in Iceland (Fig. 3D). However, while this high level of diversity might superficially imply an ancient origin within Europe, it can be accounted for in other ways

that are consistent with our model. Firstly, while most European Karakul are derived from the Halle flock, records indicate that other isolated exports of Karakul from Uzbekistan to Europe occurred around the turn of the century (Anonymous 1900; Holm 1916). Thus, it is possible there have been multiple introductions into Europe, involving SRLV-A strains randomly sampled from a more ancient (and presumably heterogeneous) epidemic in Uzbekistan (Fig. 1). Furthermore, even if we assume that the Halle flock was the source of all European SRLV-A infections, several divergent SRLV-A strains could have been imported at once when the flock was established in 1903. The differential diffusion of these distinct strains during the 20th century could generate the superficial impression of a diverse SRLV-A epidemic within Europe. Finally, it may be significant that the vigorous pursuit of Karakul farming outside of Central Asia coincided with a period of industrialisation in agriculture and experimentation in animal husbandry (Fig. 5). Some activities associated with this—e.g. cross-breeding, artificial insemination, and feeding of young animals with colostrum pooled from multiple ewes/does (Shah et al. 2004)—could have unintentionally amplified transmission in a way that favoured the generation of novel SRLV diversity (e.g. via recombination). To the extent this has occurred, it could account for the relatively high diversity of 'A-like' SRLVs within Europe.

Our dependence on opportunistically sampled, sub-genomic sequence data limited what we could definitively conclude in the present study. Encouragingly, however, our investigation suggests that obtaining further genome-length sequences from SRLVs isolated in different regions of Eurasia may allow for a more definitive assessment of SRLV origin and spread. The resources constructed in our investigation (Table 1, Fig. 2, Fig. S1, Gifford 2022) can expedite further evaluation of our origin/emergence hypothesis (Fig. 5)—as well as alternative hypotheses (Erhouma et al. 2008; Molaee et al. 2020; Olech, Osiński, and Kuźmak 2020)—as new data become available. In addition, it can provide a broad foundation for the development of information systems that utilise comparative genomic approaches to inform SRLV diagnostics and research.

Materials and methods

Sequence data

Previously published SRLV sequences were retrieved from GenBank by using the following search phrases to query the 'Organism' field: 'Maedi-Visna virus', 'Caprine Arthritis Encephalitis Virus', 'Small Ruminant Lentivirus', 'Ovine lentivirus', and 'Caprine lentivirus'. Data were filtered to remove laboratory constructs and sequences <400 bp. Linked data include the year, country and place of sampling, isolate name, host species, and associated publication (if any). Isolation dates were either obtained from GenBank files, from associated publications, or through direct contact with the authors of studies.

Phylogenetic analysis

Sequences were aligned using a BLAST-based codon-aware approach implemented in GLUE (Singer et al. 2018). Nucleotide and amino acid substitution models were selected for MSAs using the likelihood ratio test. ML phylogenies were constructed using RAxML (Stamatakis 2006). Trees were annotated using FigTree (http://tree.bio.ed.ac.uk). We used recombination detection program 4; (Lemey et al. 2009) and Genetic Algorithm for Recombination Detection to test for recombination. Recombination events detected with a significance value lower than 10E–5 and Bonferroni correction were accepted.

Literature search

We documented the breed associations of disease outbreaks involving SRLV-A (Table S4) and SRLV-B (Table S5) via literature search. Electronic searches for studies written in English, German, French, and Portuguese and that documented SRLV outbreaks based on pathology, serology, or nucleic acid amplification were conducted in November 2012. We searched the following databases: PubMed/Medline, JSTOR, Web of Science, WorldCat, and Google Scholar. In addition, searches were conducted for outbreaks reported in ProMED-mail as well as for sequences from unpublished studies reported in the National Center for Biotechnology Information's nucleotide database. A non-redundant list of reports was obtained using the following search terms: 'Maedi Visna Virus', 'Visna', 'Maedi', 'Ovine Progressive Pneumonia Virus', 'Caprine Arthritis Encephalitis Virus', 'MVV', 'CAEV', 'SRLV', 'Small Ruminant Lentiviruses', 'Montana Sheep Disease', 'Montana Progressive Pneumonia', 'Chronic Progressive Pneumonia', and 'Zwoegerziekte'.

Serology

Sampling in Lebanon was conducted during August 2012. Sampling took place in various regions of the 'Beqaa Valley', a local centre of agricultural production. Flocks were chosen at random from the Central, West, and Eastern Beqaa, as well as the northern region of Mount Lebanon. The small ruminant population in Lebanon consists mainly of the local fat-tailed 'Awassi' sheep breed, whereas the goat population is ∼95 per cent 'Baladi' and ∼5 per cent 'Damascus'. Herds are often semi-nomadic, grazing during the day and returning to sheltered structures during the night, and frequently changing location depending on the season. Goats and sheep are often housed in the same areas and will graze together as one flock. Serological testing (enzyme-linked immunosorbent assay (ELISA)) in Lebanon was conducted using plasma samples obtained from 449 animals: 221 goats and 228 sheep. A total of thirty-six goat, sheep, and mixed flocks were sampled, with ten to fifteen animals in each flock being selected at random for plasma sampling. Of the initial 449 animals tested in Lebanon, ∼15.6 per cent (seventy samples in twenty-six flocks: twenty goat and fifty sheep) were seropositive for exposure to SRLVs. Serotype was determined for a representative subset ($n = 45$) of positive samples using P16-25 ELISA. Whole blood was spun at 3500 rpm, and the serum and buffy coat were collected and stored at −20°C for further testing. ELISAs were performed as described by Reina et al. (2009b). Net absorbance was obtained by subtracting the absorbance of negative antigen from the absorbance of each recombinant antigen. The cut-off value was defined as the percentage of reactivity 20 per cent of the absorbance of positive control included in each plate.

The serological reactivity of Jordanian samples to SRLV-A and SRLV-B antigens has been described previously (Tolari et al. 2013). A subset of these sera was further analysed in this study with Eradikit SRLV genotyping kit (In3Diagnostic) to assess the presence of SRLV-E.

DNA sequencing

Genomic DNA of the Lebanese samples was extracted from stored white blood cells (buffy coats) using the DNeasy blood and tissue kit (Qiagen) and quantitated using the NanoDrop. Nested PCR was performed on the Gag region of the genome as described previously (Tolari et al. 2013). PCR products were Thymine and Adenine bases of DNA (TA)-cloned into pCR4.0 sequencing vector (Life Technologies) and sent for sequence analysis (Genewiz). Genomic DNA from the Jordanian samples was

extracted with DNeasy blood and tissue kit (Qiagen). PCR reactions were carried out using HotStarTaq Plus DNA Polymerase (Qiagen). Long-Range PCR Enzyme Mix (Qiagen) was used for the longer amplicons. PCR products of the expected size were gel-purified using MinElute Gel Extraction Kit (Qiagen) and either cloned into pCR2.1 vector (TOPO TA Cloning Kit; Invitrogen) prior to sequencing or directly sequenced on both strands (BMR genomics). A complete proviral sequence was assembled for the Jord1 isolate by combining long-range PCR and primer-walking. In addition, a 1.8-kb region encompassing most of the LTR and *gag* sequences was obtained from three sheep whose sera were reactive to both A and B antigens. Table S10 provides details of PCR primers used to amplify Jordanian SRLV sequences and the co-ordinates of their amplicons. Viral sequences generated in this study have been submitted to GenBank under accession numbers KU170752–KU170766 (Lebanese *gag* sequences), KT898826 (Jordanian full genome), KT921316–KT921317 (Jordanian LTR sequences), and KT921318–KT921320 (Jordanian *gag* sequences).

Bayesian phylogeography

Alignment partitions for phylogeographic analysis were selected from a genome-length MSA containing all published SRLV sequences. We used TempEst (Rambaut et al. 2016) and RogueNaRok (Aberer, Krompass, and Stamatakis 2013) to assess temporal signal in the data and to identify potential rogue taxa. For each MSA partition, a time-calibrated phylogeny was reconstructed using BEAST 1.8.2 with the SDR06 substitution model and an uncorrelated lognormal relaxed clock model. Molecular clock model comparison was investigated using Akaike's information criterion for Markov chain Monte Carlo samples available in Tracer. In all cases, the model comparison indicated support for the relaxed clock. Additionally, the posterior distribution for the coefficient of variation in the uncorrelated lognormal relaxed clock model did not impinge on zero indicating that the relaxed clock model provides a better fit to the data than the strict clock. After testing for the best molecular clock model, we tested two different coalescent models: Bayesian skyline plot and Bayesian skyride. The dispersal between countries was estimated using the discrete state phylogeographic approach described by Lemey et al. (2009) using a symmetric substitution model and implementing the Bayesian stochastic search variable selection model. The Bayesian GMRF skyride coalescent model was selected as a flexible demographic tree prior. Stationarity was assessed based on the effective sampling size (ESS) after a 10 per cent burn-in using Tracer. Two independent runs were generated to confirm convergence. The runs were combined using LogCombiner and summarised using TreeAnnotator. In cases where there was a large amount of uncertainty in the node heights resulting in negative branch length, the common ancestor tree was used to summarise the trees instead of the maximum clade credibility topology. Whole genome analysis did not produce ESS values above 100 for the root height for SRLV-A and SRLV-B, probably due to insufficient data (Table S11).

Genotyping algorithm

We used GLUE to assign genotypes to SRLV sequences. GLUE's genotyping procedure uses the evolutionary placement algorithm implemented in RAxML to place query sequences in an optimal positions within a previously reconstructed reference phylogeny (Stamatakis 2006). Each query sequence is examined separately, and GLUE's ML clade assignment (MLCA) algorithm is used to assign genotypes to queries based on their topological relationships to reference sequences. Assignment of sequences was based on phylogenetic criteria, including bootstrap support values and the topological relationships of query sequences relative to reference sequences, as described previously (Singer et al. 2018). We used Cohen's Kappa statistic (Bangdiwala and Shankar 2013) as implemented in GraphPad (graphpad.com) to compare the results of MLCA genotyping to phylogeny-based assignment of genotypes. The agreement values between the two methods were plotted using RStudio (v2022.07.0-548) (Friendly, David, and Zeileis 2015).

Data Availability

Data available via GitHub: https://github.com/giffordlabcvr/Lentivirus-GLUE.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Acknowledgements

R.J.G. and J.H. were funded by the Medical Research Council of the UK (MC_UU_12014/12). We thank Elie Barbour for contributing materials, lab space, and equipment at the University of Beirut and Francesco Tolari for providing Jordanian sheep and goat blood samples. We also thank Valgerður Andrésdóttir, Katerina Angelopoulou, Beatriz Amorena, Barbara Blacklaws, Miguel Fevereiro, Gordon Harkiss, Isidro Hötzel, Jinhai Huang, Britt Gjerset, Donald P. Knowles, Caroline Leroux, Bandit Nuansirchay, Keisuke Oguma, Monica Olech, Ramsés Reina, Sergio Rosati, Jörg Schüpbach, Hiroshi Sentsui, Hyun-Jin Shin, and Stephen Valas for providing SRLV sequence information.

Conflict of interest: None declared.

References

Aberer, A. J., Krompass, D., and Stamatakis, A. (2013) 'Pruning Rogue Taxa Improves Phylogenetic Accuracy: An Efficient Algorithm and Webservice', *Systematic Biology*, 62: 162–6.

Adams, D. S. et al. (1984) 'Global Survey of Serological Evidence of Caprine Arthritis-Encephalitis Virus Infection', *The Veterinary Record*, 115: 493–5.

Amills, M., Capote, J., and Tosser-Klopp, G. (2017) 'Goat Domestication and Breeding: A Jigsaw of Historical, Biological and Molecular Data with Missing Pieces', *Animal Genetics*, 48: 631–44.

Anonymous (1900) 'Fat-Tailed Sheep in Central Asia and California'. *Pacific Rural Press*, 60: 132–3.

Apetrei, C. et al. (2010) 'Pattern of SIVagm Infection in Patas Monkeys Suggests that Host Adaptation to Simian Immunodeficiency Virus Infection May Result in Resistance to Infection and Virus Extinction', *The Journal of Infectious Diseases*, 202: S371–6.

Ayelet, G. et al. (2001) 'Survey of Maedi-visna (MV) in Ethiopian Highland Sheep', *The Veterinary Journal*, 161: 208–10.

Bangdiwala, S. I., and Shankar, V. (2013) 'The Agreement Chart', *BMC Medical Research Methodology*, 13: 97.

Bar-Yosef, O. (1998) 'The Natufian Culture in the Levant, Threshold to the Origins of Agriculture', *Evolutionary Anthropology: Issues, News, and Reviews*, 6: 159–77.

Bertolotti, L. et al. (2011) 'Characterization of New Small Ruminant Lentivirus Subtype B3 Suggests Animal Trade within the Mediterranean Basin', *The Journal of General Virology*, 92: 1923–9.

Blacklaws, B. A. et al. (2004) 'Transmission of Small Ruminant Lentiviruses', *Veterinary Microbiology*, 101: 199–208.

——— (2012) 'Small Ruminant Lentiviruses: Immunopathogenesis of Visna-Maedi and Caprine Arthritis and Encephalitis Virus', *Comparative Immunology, Microbiology and Infectious Diseases*, 35: 259–69.

Camacho, C. et al. (2009) 'BLAST+: Architecture and Applications', *BMC Bioinformatics*, 10: 421.

Campbell, J. R. et al. (1994) 'The Seroprevalence of Maedi-Visna in Ontario Sheep Flocks and Its Relationship to Flock Demographics and Management Practices', *The Canadian Veterinary Journal*, 35: 39–44.

Capomaccio, S. et al. (2012) 'Geographic Structuring of Global EIAV Isolates: A Single Origin for New World Strains?', *Virus Research*, 163: 656–9.

Chessa, B. et al. (2009) 'Revealing the History of Sheep Domestication Using Retrovirus Integrations', *Science*, 324: 532–6.

Crawford, T. B. et al. (1980) 'Chronic Arthritis in Goats Caused by a Retrovirus', *Science*, 207: 997–9.

De Kock, G. (1929) 'Are the Lesions of Jaagsiekte in Sheep of the Nature of a Neoplasm', *15th Annual Report of the Director of Veterinary Services, Union of South Africa*, 611–41.

Erhouma, E. et al. (2008) 'Small Ruminant Lentivirus Proviral Sequences from Wild Ibexes in Contact with Domestic Goats', *The Journal of General Virology*, 89: 1478–84.

Faria, N. R. et al. (2014) 'HIV Epidemiology. The Early Spread and Epidemic Ignition of HIV-1 in Human Populations', *Science*, 346: 56–61.

Friendly, M., David, M., and Zeileis, A. (2015) *Discrete Data Analysis with R: Visualization and Modeling Techniques for Categorical and Count Data*. 1stedn. New York: Chapman and Hall/CRC.

Gifford, R. J. et al. (2007) 'Phylogenetic Surveillance of Viral Genetic Diversity and the Evolving Molecular Epidemiology of Human Immunodeficiency Virus Type 1', *Journal of Virology*, 81: 13050–6.

——— (2012) 'Viral Evolution in Deep Time: Lentiviruses and Mammals', *Trends in Genetics*, 28: 89–100.

——— (2022). *Lentivirus-GLUE*. <https://giffordlabcvr.github.io/Lentivirus-GLUE> accessed 25 Aug 2022.

Gjerset, B., Storset, A. K., and Rimstad, E. (2006) 'Genetic Diversity of Small-Ruminant Lentiviruses: Characterization of Norwegian Isolates of Caprine Arthritis Encephalitis Virus', *The Journal of General Virology*, 87: 573–80.

Grego, E. et al. (2009) 'Serological Characterization of the New Genotype E of Small Ruminant Lentivirus in Roccaverano Goat Flocks', *Veterinary Research Communications*, 33: 137–40.

Heim, S. (2008) 'Plant Breeding and Agrarian Research in Kaiser-Wilhelm-Institutes 1933-1945', In: *Boston Studies in the Philosophy and History of Science*, Dordrecht: Springer.

Herrmann-Hoesing, L. M. et al. (2008) 'Ovine Progressive Pneumonia Provirus Levels Associate with Breed and Ovar-DRB1', *Immunogenetics*, 60: 749–58.

Highland, M. A. (2017) 'Small Ruminant Lentiviruses: Strain Variation, Viral Tropism, and Host Genetics Influence Pathogenesis', *Veterinary Pathology*, 54: 353–4.

Holm, A. (1916) 'Karakul Sheep', *The Agricultural Journal and Small-Holder of South Africa*, 168–74.

Holmes, E. C. (2008) 'Evolutionary History and Phylogeography of Human Viruses', *Annual Review of Microbiology*, 62: 307–28.

Hunter, D. Y. (2016), *Karakul Sheep: Their American History* (The Shepherd, 47) <http://www.theshepherdmagazine.com/>.

Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772–80.

Keen, J. E. et al. (1997) 'Effect of Ewe Ovine Lentivirus Infection on Ewe and Lamb Productivity', *Preventive Veterinary Medicine*, 30: 155–69.

Krogh, D. C. (1955) 'Economic Aspects of the Karakul Industry in South West Africa', *The South African Journal of Economics*, 23: 99–113.

Lemey, P. et al. (2009) 'Bayesian Phylogeography Finds Its Roots', *PLoS Computational Biology*, 5: e1000520.

L'Homme, Y. et al. (2011) 'Molecular Characterization and Phylogenetic Analysis of Small Ruminant Lentiviruses Isolated from Canadian Sheep and Goats', *Virology Journal*, 8: 271.

Lucam, F. (1942) 'La Bouhite Ou Lymphomatose Pulmonaire Maligne du Mouton', *Recueil de Medecine Veterinaire*, 118: 273–85.

Marsh, H. (1923) 'Progressive Pneumonia in Sheep', *Journal of the American Veterinary Medical Association*, 62: 458–73.

Mekonnen, G. A., Sirak, A., and Chaka, H. (2010) 'Sero-epidemiological Study on Maedi-Visna in Selected Areas of Ethiopia', *Ethiopian Veterinary Journal*, 14: 101–11.

Molaee, V. et al. (2020) 'Phylogenetic Analysis of Small Ruminant Lentiviruses in Germany and Iran Suggests Their Expansion with Domestic Sheep', *Scientific Reports*, 10: 2243.

Olech, M., Osiński, Z., and Kuźmak, J. (2020) 'Seroprevalence of Small Ruminant Lentivirus (SRLV) Infection in Wild Cervids in Poland', *Preventive Veterinary Medicine*, 176: 104905.

Palsson, P. (1982) 'Maedi-Visna of Sheep in Iceland'. In: *The Scottish-Scandinavian Conference on Slow Viruses in Sheep, Goats and Cattle*, Commission of the European Communities: Reykjavik, Edinburgh.

Pepin, J. (2011) *The Origins of AIDS*. Cambridge: Cambridge University Press.

Rambaut, A. et al. (2016) 'Exploring the Temporal Structure of Heterochronous Sequences Using TempEst (Formerly Path-O-Gen)', *Virus Evolution*, 2: vew007.

——— et al. (2020) 'A Dynamic Nomenclature Proposal For SARS-CoV-2 Lineages to Assist Genomic Epidemiology', *Nature Microbiology*, 5: 1403–7.

Reina, R. et al. (2009a) 'Genome Analysis of Small-Ruminant Lentivirus Genotype E: A Caprine Lentivirus with Natural Deletions of the dUTPase Subunit, vpr-Like Accessory Gene, and 70-Base-Pair Repeat of the U3 Region', *Journal of Virology*, 83: 1152–5.

——— et al. (2009b) 'Development of Specific Diagnostic Test for Small Ruminant Lentivirus Genotype E', *Veterinary Microbiology*, 138: 251–7.

Rocha, J. et al. (2011) 'Molecular evidence for fat-tailed sheep domestication', *Trop Anim Health Prod*, 43: 1237–43.

Rwambo, P. M., Brodie, S. J., and DeMartini, J. C. (2001) 'Ovine Lentivirus Is Aetiologically Associated with Chronic Respiratory Disease of Sheep on the Laikipia Plateau in Kenya', *Tropical Animal Health and Production*, 33: 471–87.

Ryder, M. L. (1983) *Sheep and Man*. London: Gerald Duckworh & Co. Ltd.

Saraiva, T. (2013) *The Production and Circulation of Standardized Karakul Sheep and Frontier Settlement in the Empires of Hitler*, Mussolini, *and Salazar* In: Jørgensen, D., Jørgensen, F. A., and Pritchard, S. B. (eds.) *New Natures*, pp. 135–50. University of Pittsburgh Press: Pittsburgh.

Sasani, F. et al. (2013) 'Evaluation of Histopathological on Maedi Disease with Serological Confirmation in North-East of Iran', *Research Journal of Infectious Diseases*, 1.

Schmid, A. (1934) *Breeding of Farm Animals and Domestic Breeds in Switzerland*. Frauenfeld and Leipzig: Huber & Co.

Seffner, W., and Lippmann, R. (1967) 'On the Occurrence of Progressive, Interstitial Pneumonia in Sheep', *Monatshefte Fur Veterinarmedizin*, 22: 901–6.

Seimenis, A. et al. (1982) 'Slow Virus Diseases of the Sheep in Greece' *The Scottish-Scandinavian Conference on Slow Viruses in Sheep, Goats*

and Cattle, Reykjavik, Edinburgh: Commission of the European Communities.

Shah, C. et al. (2004) 'Phylogenetic Analysis and Reclassification of Caprine and Ovine Lentiviruses Based on 104 New Isolates: Evidence for Regular Sheep-to-Goat Transmission and Worldwide Propagation through Livestock Trade', *Virology*, 319: 12–26.

Sigurdsson, B. (1954) 'Observations on Three Slow Infections of Sheep: Maedi, Para Tuberculosis, Rida, a Slow Encephalitis of Sheep with General Remarks on Infections Which Develop Slowly, and Some of Their Special Characteristics', *British Veterinary Journal*, 110: 225-270, 307-322, 341-354.

Singer, J. B. et al. (2018) 'GLUE: A Flexible Software System for Virus Sequence Data', *BMC Bioinformatics*, 19: 532.

Smith, N. H. et al. (2009) 'Myths and Misconceptions: The Origin and Evolution of Mycobacterium Tuberculosis', *Nature Reviews Microbiology*, 7: 537–44.

Stamatakis, A. (2006) 'RAxML-VI-HPC: Maximum Likelihood-Based Phylogenetic Analyses with Thousands of Taxa and Mixed Models', *Bioinformatics*, 22: 2688–90.

Straub, O. C. (2004) 'Maedi-Visna Virus Infection in Sheep. History and Present Knowledge', *Comparative Immunology, Microbiology and Infectious Diseases*, 27: 1–5.

Suh, A. et al. (2014) 'Early Mesozoic Coexistence of Amniotes and Hepadnaviridae', *PLoS Genetics*, 10: e1004559.

Taylor, W. T. T. et al. (2021) 'Evidence for Early Dispersal of Domestic Sheep into Central Asia', *Nature Human Behaviour*, 5: 1169–79.

Thormar, H. (2005) 'Maedi-Visna Virus and Its Relationship to Human Immunodeficiency Virus', *AIDS Reviews*, 7: 233–45.

——— (2013) 'The Origin of Lentivirus Research: Maedi-visna Virus', *Current HIV Research*, 11: 2–9.

Tolari, F. et al. (2013) 'Small Ruminant Lentiviruses in Jordan: Evaluation of Sheep and Goat Serological Response Using Recombinant and Peptide Antigens', *Tropical Animal Health and Production*, 45: 1335–40.

Wandera, J. G. (1970) 'Progressive Interstitial Pneumonia (Maedi) of Sheep in Kenya', *The Veterinary Record*, 86: 434–8.

Woldemeskel, M., and Tibbo, M. (2010) 'Pulmonary Adenomatosis and Maedi-Visna in Ethiopian Central Highland Sheep: A Microscopic Study', *Tropical Animal Health and Production*, 42: 995–9.

Young, C. C. (1914) 'Origin of Karakul Sheep', *Journal of Heredity*, 5: 445–6.

Zeder, M. A. (2008) 'Domestication and Early Agriculture in the Mediterranean Basin: Origins, Diffusion, and Impact', *Proceedings of the National Academy of Sciences of the United States of America*, 105: 11597–604.

Zhao, L. et al. (2021) 'Next-generation Sequencing for the Genetic Characterization of Maedi/Visna Virus Isolated from the Northwest of China', *Journal of Veterinary Science*, 22: e66.