1   **Title: An investigation into the origins and history of pandemic small ruminant**

2   **lentivirus infection.**

3

4   **Running title: Emergence of pandemic lentivirus infection in small ruminants.**

5

6   *Maria-Luisa Carrozza [1*], Anna-Maria Niewiadomska [2*], Maurizio Mazzei [3], Mounir R. Abi-*

7   *Said[4], Stéphane Hué [5], Joshua B. Singer [6], Joseph Hughes [6], Robert J. Gifford [6,†]*

8

9

10   *[1]Scuola Normale Superiore, Pisa, Italy; [2]Aaron Diamond AIDS Research Centre, New York,*

11   *USA; [3]Department of Veterinary Science, Università of Pisa, Pisa, Italy; [4]Faculty of Sciences*

12   *II, Lebanese University, Al Fanar, Lebanon [5]London School of Hygiene and Tropical*

13   *Medicine, London, UK [6]MRC-University of Glasgow Centre for Virus Research, Glasgow, UK*

14

15   * Equal contributions

16   [†]To whom correspondence should be addressed: robert.gifford@glasgow.ac.uk

17

18

Abstract

Small ruminant lentiviruses (SRLVs) cause chronic, persistent infections in populations of domestic sheep and goats throughout the world. In this study, we use genomic data to investigate the origins and history of the SRLV pandemic. To explore the hypothesis that SRLV infection disseminated during Neolithic times, we performed a serology and DNA sequencing-based investigation of SRLVs diversity in the Fertile Crescent region, where domestication of sheep and goats is thought to have originally occurred. While we found an elevated level of viral genetic diversity compared to other regions of the world, we did not find unambiguous evidence that the Fertile Crescent region was the centre of the contemporary SRLV pandemic. We therefore examined historical reports to investigate the relationship between contemporary SRLV distribution and diversity and the emergence of SRLV-associated disease. Historical data suggested that the emergence of SRLV-associated disease might be associated with the long-distance export of exotic small ruminant breeds - in particular, karakul sheep from Central Asia - during the late 19th and early 20th centuries. Phylogeographic analysis could neither confirm nor refute this hypothesis. However, we anticipate that future accumulation of genomic data from SRLV strains found throughout the world may allow for a more definitive assessment. The openly available data and resources assembled in this study will facilitate future investigations in this area.

Importance

Viruses that cause chronic, persistent infections have circulated in animals for millions of years. However, many have only emerged as pathogens within the far shorter timeframe of recorded human history. It is important to understand the history of chronic viral infections in domestic animals, so that more effective control and eradication programs can be developed.

Introduction

Small ruminant lentiviruses (SRLVs) are retroviruses that infect domestic sheep (*Ovis aries*) and goats (*Capra hircus*), causing chronic, persistent infections that ultimately lead to organ failure and death. Infection is usually only apparent following an incubation period of ~3-4 years, and lifelong subclinical infections have been reported (1). General susceptibility and specific disease manifestations are influenced by genetic factors, and vary between species and breeds, but typically include pneumonia, wasting, paralysis, polyarthritis and mastitis.

The prototypic SRLV isolate was obtained from an epidemic that emerged in Icelandic sheep in the late 1930s. This isolate was cultivated *in vitro* in the late 1950s, and was named ovine maedi-visna virus (OMVV) (1-4). A virus closely related to OMVV was isolated from a North American dairy goat in 1974 and named *caprine arthritis-encephalitis virus* (CAEV) (5). It has subsequently become clear that each of these viruses occurs in both sheep and goats. Accordingly, they are now considered to represent two distinct genotypes (A and B respectively) within of a virus species (SRLV) (6). Analysis of complete viral genomes indicates that at least five major SRLV genotypes circulate in domesticated small ruminants (A, B, B3, C and E) (7-9). Genotypes A and B are widespread throughout the world (6, 7, 10, 11) having apparently been disseminated by recent livestock trade (6). By contrast, the divergent genotypes B3, C and E appear to have relatively restricted distributions. Phylogenetic studies have implicated livestock trade in pandemic spread of SRLV genotypes A and B (6). However, the early events that gave rise to pandemic SRLV spread are not well understood. In this study, we examined SRLV genome data using molecular phylogenetic approaches to investigate the origins and early history of the SRLV pandemic.

Materials and methods

Serology and sequencing

Serological testing of sheep and goat flocks in Lebanon was conducted during August 2012. Sampling took place in various regions of the 'Beqaa valley', a local center 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 local fat-tailed 'Awassi' breed sheep, whereas the goat population is ~95% 'Baladi' and ~5% '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 (ELISA) was conducted using plasma

3

83   samples obtained from 449 animals: 221 goats and 228 sheep. A total of 36 goat, sheep and

84   mixed flocks were sampled, with ten to fifteen animals in each flock being selected at random

85   for plasma sampling. Of the initial 449 animals tested in Lebanon, ~15.6 % (70 samples in 26

86   flocks: 20 goat; 50 sheep) were seropositive for exposure to SRLVs. Serotype was

87   determined for a representative subset (n=45) of positive samples using P16-25 ELISA.

88   Whole blood was spun at 3500rpm, the serum and buffy coat collected, and stored at -20C

89   for further testing. ELISA assays were performed as described in (12). Net absorbance was

90   obtained by subtracting the absorbance of negative antigen from the absorbance of each

91   recombinant antigen. Cut-off value was defined as percentage of reactivity 20% of the

92   absorbance of positive control included in each plate. Genomic DNA was extracted from

93   stored white blood cells (buffy coats) using the DNeasy Blood and Tissue kit (QIAGEN), and

94   quantitated using a nanodrop. Nested PCR was performed on the Gag region of the genome

95   as described previously (13). PCR products were TA cloned into pCR4.0 sequencing vector

96   (Life Technologies) and sent for sequence analysis (Genewiz). Viral sequences generated in

97   this study have been submitted to GenBank under accession numbers KU170752-

98   KU170766.

99

100   *Dataset construction*

101   Previously published SRLV sequences were retrieved from GenBank in 2015 using

102   the following search phrases to query the 'Organism' field: 'Maedi Visna virus'; 'Ovine

103   Progressive Pneumonia Virus'; 'Caprine Arthritis Encephalitis Virus'; 'Small Ruminant

104   Lentivirus'; 'Ovine lentivirus'; 'Caprine lentivirus'. We filtered the 5,255 sequences recovered

105   via this means to derive a subset of 585 previously published sequences that each

106   represented a distinct infection. For each determined the pubmed ID of the study it was

107   obtained in, the isolate name, the year of isolation, and the host species (sheep or goat).

108   This information was obtained from GenBank files where possible, or else from scientific

109   papers associated with the published sequences, and failing that via direct contact with study

110   authors. Sequences in this dataset were aligned using a combination of MUSCLE (14) and

111   Pal2Nal (15). We used GLUE – a data-centric bioinformatics framework for working with

112   virus sequence data (16) – to capture the relationships between sequences, alignments and

113   associated tabular data utilised in our investigation. The resultant resource is publicly

114   available via GitHub (https://github.com/giffordlabcvr/SRLV-GLUE).

115

116   *Literature search/historical investigation*

117   To investigate previous reports of SRLV outbreaks, we conducted electronic

118   searches for studies written in English, German, French, Portuguese, and that documented

119   SRLV outbreaks based on either pathology, serology; by nucleic acid amplification. In

120   November 2012, we searched the following databases: PubMed/Medline, JSTOR, Web of

121   Science, WorldCat and Google scholar. In addition searches were conducted for outbreaks

122   reported in ProMED Mail, as well as for sequences from unpublished studies reported in the

123   National Center for Biotechnology Information's nucleotide database. A non-redundant list of

124   reports was obtained using the following search terms; 'Maedi Visna Virus'; 'Visna'; 'Maedi';

125   'Ovine Progressive Pneumonia Virus'; 'Caprine Arthritis Encephalitis Virus'; 'MVV'; 'CAEV';

126   'SRLV'; 'Small Ruminant Lentiviruses'; 'Montana Sheep Disease'; 'Montana Progressive

127   Pneumonia'; 'Chronic Progressive Pneumonia'; 'Zwoegerziekte'.

128

129   *Phylogenetic, phylogeographic and phylodynamic analysis*

130   Phylogenetic analysis of SRLV *gag* sequences obtained from sheep in Lebanon and

131   of complete SRLV genome sequences was performed using RaXML (17), using the general

132   time reversible (GTR) nucleotide substitution model and 1000 bootstrap replicates. Time-

133   calibrated phylogenies were reconstructed using BEAST 1.8.2 with the SDR06 substitution

134   model and an uncorrelated lognormal relaxed clock model. Molecular clock model

135   comparison was investigated using AICM available in Tracer. In all cases the model

136   comparison indicated support for the relaxed clock. Additionally, the posterior distribution for

137   the coefficient of variation in the uncorrelated lognormal relaxed clock model did not impinge

138   on zero indicating that the relaxed clock model provides a better fit to the data than the strict

139   clock. After testing for the best molecular clock model, we tested two different coalescent

140   models: Bayesian skyline plot and Bayesian skyride. Trees were annotated using FigTree

141   (http://tree.bio.ed.ac.uk).

142   The dispersal between countries was estimated using the discrete state

143   phylogeographic approach described in Lemey et al. (18) using a symmetric substitution

144   model and implementing the Bayesian stochastic search variable selection model (BSSVS).

145   The Bayesian GMRF skyride coalescent model was selected as a flexible demographic tree

146   prior. Stationarity was assessed on the basis of the effective sampling size (ESS) after a

147   10% burning using Tracer and two independent runs were generated to confirm

148   convergence. The two independent runs were combined using LogCombiner and

149   summarized with TreeAnnotator. In cases where there was a large amount of uncertainty in

150   the node heights resulting in negative branch length, the common ancestor tree was used to

151   summarise the trees instead of the Maximal Clade Credibility topology. Whole genome

152   analysis did not produce ESS values above 100 for the root height and a number of other

153   parameters for SRLV-A and SRLV-B probably due to insufficient data.

154

155

Results

SRLV diversity in the Fertile Crescent region

Archaeological and genetic evidence indicate that both sheep and goats were domesticated in the 'Fertile Crescent' region of Western Asia ~10.5-11 thousand years ago (19) (**Figure 1a**). The presence of highly divergent SRLV genotypes in remnant populations of primitive small ruminant breeds (genotype C in Norway, genotype E in Sardinia) is consistent with an ancestral radiation of SRLV genotypes out of the Fertile Crescent region in association with the diffusion of early agricultural systems during the early Neolithic (20, 21). However, this pattern might also be accounted for in other ways (e.g. each genotype might represent an independent introduction of virus into the domesticated small ruminant population from a wild ruminant species).

We carried out a study of SRLV diversity in the Fertile Crescent region to investigate whether this area might have been the centre of an ancient SRLV radiation. We focused on Lebanon, a country in the Levant region, where (**Figure 1b**), where some of the earliest human civilizations are believed to have developed (22). Serological testing of 886 animals indicated that diverse SRLVs are present in both countries (**Table 1**). Furthermore, overall prevalence was relatively high (~21%) in comparison with previous global surveys (23), but similar to estimates obtained in other countries in the region (13, 24). We used PCR to obtain *gag* gene amplicons from a randomly selected subset of our SRLV-positive samples (primers are listed in **Table S1**).

We obtained amplicons from fifteen samples and investigated their evolutionary relationships to a representative set of published sequences (**Figure 2**). In phylogenies constructed using this dataset, three distinct clades of Levantine isolates (Lev I-III) were observed. The Lev-I and Lev II clades grouped robustly within the diversity of previously characterized SRLV-A isolates, while Lev-III was relatively distinct from all previously sampled SRLV genotypes. A full-length proviral genome sequence has previously been recovered for one Lev-III isolate sampled in Jordan (Jord1). We used full-length genome sequences to reconstruct the phylogenetic relationships between Jord1 and other SRLV genomes, showing it is approximately as distantly related to established SRLV genotypes as they are to one another (**Figure S1**).

While the Lev III clade is highly divergent, it does not represent an obvious ancestor to all other SRLV genotypes. Furthermore, although the overall level of diversity in the Fertile Crescent region appears to be relatively high, based on observations from our study and others (10), the range of genotypes and novel lineages identified was not as exceptionally high as might be expected if this region were the centre of an ancient SRLV radiation (i.e. similar to the high diversity of human immunodeficiency virus type 1 (HIV-1) found in areas of West Central Africa). In fact, most of the isolates detected in our study grouped within the

6

193 diversity of pandemic SRLV-A strains found in Europe and North America, indicating they are
194 as likely to have been imported into the region as exported from it (**Figure 2**).

195

196 *Recorded history of pandemic SRLV emergence and spread*

197 Previous epidemiological studies have indicated that pandemic spread of SRLVs may
198 only have occurred relatively recently (6, 23, 25, 26). They do not appear to have spread in
199 parallel with the first European colonists, as seems to be the case for many other livestock
200 pathogens (27, 28). For example, SRLV-A has never been reported in Australia or New
201 Zealand, where sheep are largely descended from animals imported in the 19th century from
202 the UK, Germany, South Africa and North America (6, 25). This suggests that SRLVs were
203 not widespread during the 'Age of Discovery' (from ~1500 to ~1700), when the first exports of
204 small ruminants from Europe to Africa, Oceania and the Americas took place.

205 To explore the manner in which SRLVs emerged as globally distributed pathogens,
206 we reviewed the historical evidence for SRLV-associated disease in countries throughout the
207 world. For each country where SRLV infection has been reported, we determined the earliest
208 reliable evidence for the presence of the virus (**Table S1, Table S2**). Previous studies have
209 identified diseases reported in the past that appear likely - in retrospect - to have been
210 caused by SRLVs (1). Cases where the only available evidence for historical SRLV infection
211 is based on a description of disease pathology have to be viewed cautiously, particularly
212 because of potential confusion with sheep pulmonary adenomatosis, caused by Jaagsiekte
213 sheep retrovirus (JSRV). Nonetheless, SRLV-A can confidently be stated to have been
214 present in Europe since at least 1933, because the Icelandic epidemic from which the
215 prototypic SRLV-A strain was derived was definitively traced to an importation of Karakul
216 sheep in this year (29).

217 Intriguingly, the early emergence of SRLV-A infection closely mirrors the early 20th
218 century diffusion of Karakul, which are among the oldest domesticated sheep breeds (30, 31)
219 (**Figure 3**). The breed is native to Central Asia, and arrived in Iceland via the Institute for
220 Animal Breeding, in Halle, Germany. Karakul sheep were originally exported to Halle from
221 Bukhara in Uzbekistan in 1903 (32), and were also introduced to the United States in the
222 same decade (33) (**Table S1**). In the first half of the 20th century, a cluster of similar disease
223 syndromes emerged, all of which are now considered likely to have been caused by SRLV
224 infection (29). '*Graaf-Reinet disease*' was first described in South Africa in 1915, less than a
225 decade after the introduction of Karakul into the region. The same year, '*Montana sheep
226 disease*' was reported for the first time in the United States, seven years after the arrival of
227 Karakul. In Europe, both *maedi-visna* (in Iceland), and '*la bouhite*' (in France) emerged in
228 ~1930-50, in parallel with the initial dissemination of German Karakul in Europe.

229   Similarly, we observed that the pandemic spread SRLV-B (specifically subtype B1)
230   was conspicuously associated with Swiss dairy goat breeds, particularly the Swiss Saanen
231   breed (**Table S2**). Many of the first reported outbreaks of SRLV infection in goats involved
232   this popular breed (34). Indeed, the prototypic SRLV-B isolate (CAEV Cork) derives from an
233   outbreak of chronic, progressive disease that emerged in a North American population of
234   Saanen goats in 1970 (35). Until the late 19th century, small ruminant breeds such as
235   Saanen and Karakul had strictly defined geographic distributions. A role for the livestock
236   trade in SRLV emergence has been proposed previously. However, the role of specific breed
237   movements in driving the early emergence of pandemic strains has not been closely
238   examined.

239
240   *Phylogeographic investigation of pandemic SRLV spread*

241   We used Bayesian phylogeographic approaches to investigate whether available
242   genomic data supported a role for breed dispersal in the emergence of pandemic SRLV
243   infection, as suggested by our investigation of historical records. We compiled a
244   comprehensive set of published SRLV sequences. From an initial set of 5,255 sequences,
245   we derived a subset of 697 that together represented 585 unique field isolates, each
246   obtained from a distinct infected animal (**Table 2, Table S3**). We added the fifteen SRLV *gag*
247   sequences we obtained in Lebanon to this set. Together, this set of sequences represented
248   600 isolates sampled in a total of 30 countries, with dates of sampling ranging from 1949 to
249   2013. A preliminary phylogenetic analysis (data not shown) was performed to determine the
250   genotype of all sequences in this set. We then created individual multiple sequence
251   alignments (MSA) for each of the pandemic SRLV strains (SRLV-A and SRLV-B1). From
252   these alignments, we derived a total of eight alignment partitions (**Table S4**) - two within the
253   group-specific antigen (*gag*) gene, and two partitions within the polymerase (*pol*) gene. We
254   detected temporal structure in five of the eight datasets, despite the presence of outliers
255   (**Table S5**). The results of Bayesian phylogeographic analysis of the four *pol* partitions are
256   shown in **Figure 4** (SRLV-A) and **Figure 5** for (SRLV-B1), while the results obtained for the
257   four *gag* partitions are shown in **Figure S2** (SRLV-A) and **Figure S3** for (SRLV-B1). **Table 4**
258   and **Table 5** provide information about specific nodes of interest in all eight phylogenies. We
259   obtained more robust results when using *pol*, however, we have included the results
260   obtained for *gag* as these provide some potentially useful insights regarding the epidemic in
261   specific parts of the world.

262   In general, phylogenies were only moderately structured as opposed to 'star-like'. We
263   obtained robust support for nodes near the phylogeny tips, particularly within clades of
264   viruses sampled from the same geographic region, but only a proportion of the deeper
265   branching relationships (i.e. between clades of viruses sampled in distinct counries and/or

8

266   geographic regions) had high support. For all four *pol* partitions, and one SRLV-B1 *gag*
267   partition, the location of the most recent common ancestor was placed in Switzerland.
268   Ancestral locations within Asia Minor were obtained for SRLV-A *gag* partitions, but with low
269   support. One SRLV-B1 gag partition placed the most recent common ancestor in the United
270   States (US), with moderately high support.

271   Overall, the results of Bayesian analysis were consistent with a scenario within which
272   both SRLV-A and SRLV-B1 disseminated worldwide from a European epicentre. Where we
273   were able to confidently estimate the location of deeper nodes in the phylogeny, we found
274   these were placed in Central Europe. By contrast, clades sampled in Asia, South America
275   and the around the Northern and Western periphery of Europe (i.e. Ireland, Norway, Finland)
276   were observed to have relatively recent common ancestors (**Table 4**, **Table 5**).

277   Although the lack of resolution in our phylogenies (as well as the limitations of
278   opportunistic sampling) prevented us from inferring the precisely the direction and timing of
279   global SRLV spread, we were able to confidently estimate tMRCAs for most of the MSAs
280   examined. For SRLV-A, we obtained tMRCA estimates in the late 19th century - shortly
281   predating the first Karakul exports from Central Asia - for both reverse transcriptase (RT)
282   (1883, 1820-1923 95%HPD) and integrase (IN) (1887, 1837-1915 95%HPD). The SRLV-A
283   *gag* partitions yielded more conflicting results: for the matrix partition (Gag MSA 1) we
284   obtained much older estimates with wide error margins, whereas the capsid partition (Gag
285   MSA 2) yielded a somewhat more recent tMRCA (**Table 3**). For SRLV-B1 the capsid
286   partition yielded unreliable tMRCA estimates with wide error margins. However, we obtained
287   more reliable and generally consistent results from the other three MSA partitions, which
288   provided tMRCA estimates of 1966 for RT (1958-1974 95%HPD), 1952 for IN (1939-1964
289   95%HPD), and 1972 for matrix (1957-1974 95 HPD) (**Table 4**).

290

Discussion

292   In this study, we used molecular phylogenetic approaches to investigate the origins
293   and history of pandemic SRLV spread. To investigate the hypothesis that SRLV originally
294   disseminated out of Western Asia during the early Neolithic period (i.e. in association with
295   domesticated small ruminants) we sampled SRLV strains circulating in the Fertile Crescent
296   region where domestication of these species is thought to have occurred. We found a
297   relatively high prevalence and diversity of SRLV strains, but we did not find obvious evidence
298   of the region having been the centre of an ancient SRLV radiation. The lack of clear evidence
299   that the various SRLV genotypes derive directly from a virus population that circulated in the
300   ancestral small ruminant populations from which modern breeds are derived, does not rule
301   out the possibility that they did. It could be that our sampling was not entirely representative,

302   or that SRLV diversity in the region has been greatly influenced in recent times by factors
303   such transport and migration (6).

304   To investigate the origins and timing of pandemic SRLV spread, we combined a
305   review of historical records with a phylogenetic investigation of SRLV genome data. Our
306   research led us to consider the possibility that the emergence of pandemic SRLV infection
307   might have its roots in the late 19th and early 20th century, during which time 'exotic' small
308   ruminant breeds were exported to Europe, North America and Africa. In particular, we noted
309   that historical reports suggest the early pattern of emergence of SRLV-associated disease
310   around the world appears – at least superficially – to be correlated with the export of karakul
311   – a breed of fat-tailed sheep, native to Central Asia (**Figure 3**). Even though it was
312   established relatively early on that SRLV-A infection was imported to Iceland along with a
313   flock of karakul sheep in the 1930s, this breed has not previously been implicated in global
314   spread of SRLV. However, a role for karakul in the early emergence of SRLV-A might easily
315   have been missed if the SRLV strains exported along with this breed were not associated
316   with obvious disease in their natural host. Current evidence indicates that Lentiviruses,
317   including some SRLVs, are relatively apathogenic in populations with which they have
318   longer-term associations (36, 37). Thus, in the light of what we now know lentivirus pathology
319   and its relationship to host genetic variation, a plausible hypothesis accounting for the
320   emergence of pandemic SRLV-associated disease is that it was driven by the introduction
321   (via karakul export) of Central Asian SRLV strains into populations of small ruminants in
322   Europe, Africa and America that had previously only been exposed to SRLV strains prevalent
323   in Europe.

324   We propose that the early spread of SRLVs could have been enabled by the creation
325   of new contact networks in livestock populations, leading to particular small ruminant breeds
326   and populations being brought into contact for the first time. Subsequent to this, the
327   development of international livestock trade (6) - as well as the development and uptake of
328   new systems of agriculture - would presumably have helped facilitate the further spread of
329   SRLVs from European and Asian epicenters to Africa, the Americas, Oceania, and East
330   Asia. The capacity of some small ruminant breeds to harbor SRLV infection without
331   exhibiting symptoms, and a lack of understanding about the nature of slow virus infections
332   prior to the 1950s (29), would presumably helped enable this relatively recent wave of global
333   SRLV spread.

334   To investigate whether available genomic data supported this hypothesis, we
335   performed a phylogeographic analysis of the available sequence data for the pandemic
336   SRLV strains A and B1. Overall, results were consistent with both strains having spread
337   worldwide relatively recently (i.e. within the 20th century), emerging from a Eurasian source.
338   For SRLV-A, we obtained tMRCA estimates in the late 19th century - shortly predating the

339 first Karakul exports from Central Asia – and thus consistent with a scenario under which

340 these exports ultimately led to the emergence of an SRLV pandemic.

341   SRLV-B1 appears to have disseminated globally more recently than SRLV-A, from an

342 unknown source in Europe. Global spread of this SRLV lineage is conspicuously associated

343 with Swiss dairy goat breeds in general, and the Saanen breed in particular (**Table S2**).

344 Many of the first reported outbreaks of SRLV infection in goats involved this breed, which

345 was first exported from Switzerland in the late 19th century, and has subsequently been

346 introduced to many countries throughout the world (34). However, given that (i) SRLV

347 infection is known to cause disease in Saanen and other Swiss dairy goat breeds, and; (ii)

348 these breeds have been exported from Switzerland since the late 19th century (34), It seems

349 unlikely that Swiss dairy goats are the ancestral hosts of SRLV-B1. One possibility is that

350 spread of B1 from a Swiss epicenter was preceded by the introduction of novel SRLV

351 diversity into the Swiss small ruminant population during the early 20th century.

352   A major limitation of our analysis is that the majority of sequence data we examined

353 was opportunistically sampled from what was already available in GenBank. Sampling of

354 SRLV diversity so far has been patchy, and in addition, most of the available sequences are

355 sub-genomic (see **Table 2**), which limits what can be inferred. Furthermore, sampling is

356 heavily biased toward certain areas, distorting our phylogeographic inference. For example,

357 Switzerland was probably the most thoroughly sampled country in our dataset, thus the

358 phylogeographic placement of the SRLV-A and SRLV-B MRCAs in Switzerland has to be

359 viewed with some caution. Interestingly, however, Karakul are documented as having been

360 present in Switzerland since at least the early 1930s (38).

361   Test and removal programs are currently used as a means of controlling SRLV

362 infection, with varying levels of effectiveness (39). In fact, recent years have seen the

363 emergence of SRLV infection in countries where it had not previously been documented, as

364 well as the re-emergence of disease in areas where it had been considered to be under

365 control (40, 41). We therefore anticipate that the use of genome data to investigate SRLV

366 breakdowns will increase in future. The increased availability of genome sequences from

367 SRLV strains found throughout the world – particularly in Central Asia – should eventually

368 allow for a more definitive assessment of the hypothesis presented here, and provide a clear

369 picture of the pattern of pandemic SRLV spread. The more ancient spread of SRLV viruses,

370 which is proposed to date back to the Neolithic, is unlikely to be amenable to

371 phylogeographic analysis due to the time-dependence of viral evolutionary rates (42).

372 However, the overall distribution of SRLV strains in Europe and Asia (particularly Central

373 Asia) may well be revealing in this respect. The resources assembled in this study will not

374 only facilitate further investigation of historical SRLV spread, but can also provide a

375 foundation for the development of SRLV surveillance systems that utilise genomic data.

376 **Table 1. ELISA results from small ruminants in Jordan and Lebanon**

Species (Breed)	Sample number	Positive samples		Serotype reactivity		
		#	%	A	B	E
Sheep (Awassi)	457	131	**28.7**	23	20	9
Sheep (Australian)	5	3	**60.0**	2	-	1
Goat (Baladi)	408	48	**11.8**	11	24	4
Goat (Damascus)	14	4	**29.0**	1	1	1
Goat (Baladi x Damascus)	2	0	**0.00**	-	-	-
Totals	**886**	**186**	**21**	**37**	**45**	**15**

377

378 **Legend:** SRLV prevalence and genotypes in Jordan and Lebanon. ELISA assays were performed on
379 sera collected from sheep and goats belonging mostly to local breeds. The serotype reactivity of a
380 subset of sera comprising 45 Lebanese and 112 Jordanian samples was analysed with a genotyping
381 ELISA assay targeted against the p16/matrix (MA) and p25/capsid (CA) proteins of A, B and E
382 genotype.

383

384   **Table 2. Summary of SRLV sequence studies**

385

Study [a]	Accessions (strain) [b]	Species [c]	Date [d]	Genomic region(s) [e]	Country [f]	# Seqs [e,g] Set	# Seqs [e,g] Study
S. America							
(43)	AJ305039-42	Ch, Oa	1989-97	Gag	BR	4	4
DS	AF402664-66, 68	Ch	1999	Gag	BR	4	4
DS	AY101347-8	Ch	*NK*	Gag	BR	2	2
DS	AY081139	Ch	*NK*	Gag	BR	1	1
						11	
N. America							
(44)	HQ158122-36	Ch, Oa	2004	Gag	CA	15	15
(45)	HM210570 (**FESC-752**)	Ch	2004	Genome	MX	1	1
(45)	HM212792-3	Ch	2008	Env	MX	2	2
(46)	K03327	Ch	1985	Pol	US	1	1
(47, 48)	NC_001463 (**Cork**)	Ch	1974	Genome	US	1	1
(49)	*See legend.*	Oa	2004, 08	Gag	US	10	52
(50)	*See legend*	Oa	1976, 03	Env	US	8	17
(5, 51)	M60855	Ch	1980	Env	US	1	1
(52), *DS*	AY101611 (**85/34**)	Oa	1985	Gag-Pol	US	1	1
(52, 53)	U64439 (**85/34**)	Oa	1985	Env	US	1	1
						41	
Africa							
(54)	M31646 (**SA-OMVV**)	Oa	1976	Genome	ZA	1	1
DS	FJ619565-72	Ch	2007	Gag	SU	8	8
						9	
East Asia							
(55)	GU120138 (**Shanxi**)	Ch	1995	Genome	CN	1	1
(56)	AY900630 (**Gansu**)	Ch	1992	Genome	CN	1	1
(41)	AB821356	Oa	2013	Gag	JP	1	1
(40)	N/A	Ch	N/K	Gag	KO	1	1
DS	AB747557, LC002526	Ch	2011	Gag	PH	1	2
DS	GU903321	Ch	2009	Env	TW	1	1
DS	HM237197	Ch	2010	Env	TW	1	1
DS	*See legend*	Ch	2008-11	Gag	TH	14	14
						20	
West Asia							
(57)	JQ898278-85	Oa	2007-08	Gag	TR	8	8
(10)	JF502418 (**TR-DM**)	Oa	2006	Gag	TR	1	1
						9	
South West Europe							
(58)*	*See supp.*	Oa	2008-11	Env	ES	8	350
(59)	JN184351-78	Oa	2008-11	Gag	ES	28	28
(59)	JN184379-94	Oa	2008-11	Env	ES	16	16
(60)	DQ084332-44	Ch, Oa	2005-06	Gag	ES	13	13
(60)	DQ632733-35	Ch, Oa	2005-06	Pol	ES	3	3
(61)	FJ195346 (**Ov496**)	Oa	2004	Genome	ES	1	1
(62)	HQ848062 (**697**)	Oa	2005	Genome	ES	1	1
(63)	AF479638 (**P1OLV**)	Oa	1996	Genome	PT	1	1
						71	
North West Europe							
(64)	AM084203-09	Oa	1994-96, 01	Gag	FI	7	7
(64)	AM084210-15	Ch, Oa	1994-96, 01	Env	FI	6	6
(65)	U35796, 798-801, 804, 811	Ch, Oa	1994	Env	FR	7	7
(66)	DQ149845	Oa	2004	Env	FR	1	1
(67)	AJ400718-21	Ch	1994	Env	FR	4	4
(68)	AF015181	Ch	1994	Gag	FR	1	1
(68)	AF015182	Ch	1994	Pol	FR	1	1
(68)	AF015180	Ch	1994	Env	FR	1	1
DS	AJ969032-37	Ch	1998-02	Pol	IE	6	6
DS	AJ969038-43	Ch	1998-02	Env	IE	6	6
(69)	U51910	Oa	1962	Env	IS	1	1
(70, 71)	NC_001452 (**kv1772**)	Oa	1949	Genome	IS	1	1
DS	X54379	Oa	1970	Gag	NL	1	1
(72)	DQ844911-30	Ch, Oa	1991-03	Pol	NO	20	20
(72)	DQ844931-41	Oa	1976-05	Env	NO	11	11
(20)	DQ015910–16	Ch	1991-02	Env	NO	7	7
(20)	AF322109 (**1GA**)	Ch	1991	Genome	NO	1	1
(73)	S51392 (**EV1**)	Oa	1984	Genome	UK	1	1
						84	
Central, South East and East Europe							
(6)	AY454161-232	Ch, Oa	1999	Gag	CH	72	72

13

(6)	AY454233-296	Ch, Oa	1999	Pol	CH	64	64
(6)	AY445885 (**G4668**)	Ch	1999	Genome	CH	1	1
(7)	AY577031-3	Ch, Oa	2003	Gag-pol	CH	3	3
(7)	AY577034-9	Ch, Oa	2003	Gag	CH	6	6
(74)	AY530289-93	Ch, Oa	2002	Gag	GR	5	5
(75)*	HM449450	Ch	2007	Env	IT	1	234
(76)*	EF685749, 83	Ch	2006	Env	IT	2	194
(11)	*See legend*	Ch, Oa	2009-10	Gag-pol	IT	55	61
(77)*	EF144472, 82, 93, 94	Ch	2006	Env	IT	4	45
(37)	EU293537 (**Roccaverano**)	Ch	2008	Genome	IT	1	1
(37)	EU726488-525	Ch	1999, 06, 07	Gag	IT	38	38
(8)	EF676000-26	Ch, Oa	2006	Gag	IT	24	24
(8)	EU010120-6	Ch	2006	Gag	IT	7	7
(78)	DQ013214-43	Ch, Oa	2005	Pol	IT	30	30
(79)	AY044803-10	Ch	1991	Gag	IT	8	8
(80)	GQ381130 (**Seui**)	Ch	2003	Genome	IT	1	1
(80)	GQ428519-36	Ch	2008-09	Gag	IT	18	18
(10)	JF502416 (**Fonni**)	Oa	2003	Genome	IT	1	1
(10)	JF502417 (**Volterra**)	Oa	2004	Genome	IT	1	1
(10)	JF520393-406	Ch, Oa	2008-09	Gag	IT	5	5
(81)	EU709743, EU702487	Oa	1991	Env	IT	2	3
(82)	FJ623120-25	Ch, Oa	2008	Gag	PL	6	6
(82)	FJ623110-19	Ch, Oa	2008	Env	PL	10	10
DS	JN084103	Ch	2008	Gag	RU	1	1
DS	JN008914	Ch	2008	Env	RU	1	1
(83)	JQ610956–JQ611030	Oa	2011	Pol	SL	75	75
(83)	JX469600-10	Ch, Oa	2008-11	Gag-pol	SL	11	11
					Total	**453** **697**	

386
387 **Legend.** Previously sequenced SRLV isolates, grouped by study. [a] *DS* = direct submission to
388 GenBank. Asterisks indicate studies of intra-host diversity from which a subset of sequences (one per
389 animal) was selected. [b] Accession numbers are shown only for sequences included in the reference
390 set. For complete genome sequences the isolate name is shown in bold and in parenthesis. *NA* = not
391 applicable [c] Oa=*Ovis aires*; Ch=*Capra hircus*. [d] Year of collection is shown where it could be
392 ascertained, *NK* = not known. [e] Only sequences that encoded >500 base pairs of *gag*, *pol*,; *env* were
393 included in the reference set [f] ISO country code (state). [g] Reference set was restricted to one
394 sequence per infected animal (number shown as 'in set'). Where multiple sequences were available
395 the total is indicated under 'in study'. Accession numbers not listed in the table are as follows; Thai
396 sequences (GQ161209-15, JF714253-5, EU983108-9, EU919141, FJ167525)); North American
397 sequences from reference (49) (GQ255391, GQ255386, GQ255401, GQ255411, GQ255385,
398 GQ255403, GQ255417, GQ255423, GQ255428, GQ255433); North American sequences from
399 reference (50) (AY362022, AY362024, AY362026, AY362029, AY362030, AY362032, AY362035,
400 AY362038); Italian sequences from reference (11) (FR687200, FR693808-32, FR694686-96,
401 FR694908-21, FR695719, FR828809-16.
402

14

Table 3. tMRCAs of selected clades within SRLV-A

Partition [a]	Node # [b]	Clade description [c]	Posterior [d]	Region(s) [e]	Location [f]	p.loc [g]	Age	95% HPD
Gag MSA 1	1	Polish	0.55	C. Europe	Poland	0.86	1884	54-296
Gag MSA 1	2	US	0.98	N. America	USA	0.93	1891	56-272
Gag MSA 1	3	PL-US-JP-IT-IS	0.94	Global	South Africa	0.16	1798	105-509
Gag MSA 1	4	Finnish	1.00	N. Europe	Finland	1.00	1970	24-72
Gag MSA 1	5	Spanish	1.00	W. Europe	Spanish	1.00	1858	63-361
Gag MSA 1	6	Root	1.00	Global	Jordan	0.12	1760	112-596
Gag MSA 2	1	Italy	0.95	S. Europe	Italy	0.93	1959	9-18
Gag MSA 2	2	Greece	0.96	S. Europe	Greece	0.98	1975	24-51
Gag MSA 2	3	Jordan-Lebanon-Turkey	0.98	Asia Minor	Turkey	0.68	1972	24-56
Gag MSA 2	4	S Europe-Asia Minor	0.77	S. Europe-Asia Minor	Turkey	0.55	1951	43-81
Gag MSA 2	5	Lev-II	1.00	Asia Minor	Lebanon	0.99	1990	10-34
Gag MSA 2	6	US-CA	1.00	N. America	USA	0.64	1978	24-48
Gag MSA 2	7	Swiss	0.99	C. Europe	Switzerland	0.97	1976	25-48
Gag MSA 2	8	Turkish	0.96	Asia Minor	Turkey	0.96	1968	25-61
Gag MSA 2	9	ES-UK	0.99	W. Europe	Spain	0.47	1978	30-40
Gag MSA 2	10	US-CA	0.99	N. America	USA	0.66	1964	35-61
Gag MSA 2	11	US-CA-FI-ES	0.76	Europe-N. America	Turkey	0.53	1942	50-88
Gag MSA 2	12	Finland	1.00	N. Europe	Finland	1.00	1987	21-29
Gag MSA 2	13	Finnish-BR	0.95	N. Europe	Finland	0.29	1955	38-75
Gag MSA 2	14	Swiss	1.00	C. Europe	Switzerland	1.00	1985	22-34
Gag MSA 2	15	ES	1.00	W. Europe	Spain	0.99	1958	38-70
Gag MSA 2	16	Root	1.00	Asia Minor	Turkey	0.38	1930	67-100
Pol MSA 1	1	Swiss	1.00	C. Europe	Switzerland	0.99	1977	24-49
Pol MSA 1	2	Swiss-Spanish-Slovenian	0.76	C. Europe	Switzerland	0.80	1911	63-157
Pol MSA 1	3	Swiss-US	0.99	C. Europe	Switzerland	0.75	1916	58-148
Pol MSA 1	4	Irish	1.00	W. Europe	Ireland	0.99	1993	14-24
Pol MSA 1	5	Irish-ZA	0.94	Europe-Africa	South Africa	0.28	1932	51-117
Pol MSA 1	6	Irish-ZA-UK-SL	0.91	C. Europe	Switzerland	0.38	1918	60-139
Pol MSA 1	7	Norwegian	1.00	C. Europe	Norway	0.99	1984	15-43
Pol MSA 1	8	Norway-Swiss	0.76	C. Europe	Switzerland	0.61	1920	54-139
Pol MSA 1	9	Norway-Iceland-Portugal	0.93	Europe	Switzerland	0.34	1920	77-155
Pol MSA 1	10	Root	1.00	C. Europe	Switzerland	0.60	1883	88-191
Pol MSA 2	1	Slovenian	1.00	C. Europe	Slovenia	1.00	2003	9-15
Pol MSA 2	2	Slovenia-UK	0.98	C. Europe	Slovenia	0.43	1927	54-113
Pol MSA 2	3	Slovenia-UK-IS-ZA	0.91	C. Europe	Switzerland	0.72	1905	84-159
Pol MSA 2	4	Swiss	1.00	C. Europe	Switzerland	0.87	1926	55-116
Pol MSA 2	5	ES	1.00	W. Europe	Spain	0.98	1951	32-86
Pol MSA 2	6	Swiss-ES-US-SL	1.00	C. Europe	Switzerland	0.83	1907	70-137
Pol MSA 2	7	Root	1.00	C. Europe	Switzerland	0.76	1875	97-174

Footnote: [a] Alignment partition name (see Table S4 for partition information); [b] Node number as shown in corresponding tree figures (Figures 4-5, S2-S3); [c] Internal nodes are described in terms of the countries sampled among all daughter nodes. For some nodes ISO country codes are shown instead of country names to preserve space; [d] Posterior probability of node; [e] Regions included in daughter nodes; [f] Estimated country location of node; [g] Posterior probability of location.

407 **Table 4. tMRCAs of selected clades within SRLV-B**

Partition	Node #	Clade description	Posterior	Region(s)	Location	Loc P	Age	95% HPD
Gag MSA 1	1	Thailand-Poland	1.0	C. Europe / S.E. Asia	Thailand	0.77	1993	10-33
Gag MSA 1	2	Spanish	1.0	W. Europe	Spain	0.97	1992	13-29
Gag MSA 1	3	Italian	0.84	W. Europe	Italy	0.94	1990	11-36
Gag MSA 1	4	TH-PL-ES-IT	0.75	Europe / S.E. Asia	Italy	0.71	1981	20-45
Gag MSA 1	5	Canada-China-Mexico	0.62	N. America	Canada	0.71	1982	21-42
Gag MSA 1	6	Global	0.84	Global	Canada	0.37	1978	25-42
Gag MSA 1	7	Root	1.0	Global	USA	0.73	1972	37-54
Gag MSA 2	1	Slovenia-Italy	0.83	S. Europe	Switzerland	0.79	1888	51-206
Gag MSA 2	2	Brazil	1.0	S. America	Brazil	1.0	1974	19-60
Gag MSA 2	3	Canada-Mexico	0.92	N. America	Canada	1.0	1877	61-228
Gag MSA 2	4	USA-Canada-Mexico-CN	0.85	Global	Canada	0.5	1856	70-263
Gag MSA 2	5	Switzerland-Italy	0.83	Europe	Switzerland	1.0	1862	66-250
Gag MSA 2	6	Root	1.0	Global	Switzerland	0.92	1825	82-322
Pol MSA 1	1	Italy-Switzerland	1.0	C. Europe	Switzerland	0.98	1986	16-34
Pol MSA 1	2	Slovenia	1.0	S. Europe	Slovenia	1.0	2001	6-12
Pol MSA 1	3	Italy-Switzerland	0.96	Europe	Switzerland	1.0	1966	25-47
Pol MSA 1	4	Root	1.0	Global	Switzerland	0.63	1973	37-53
Pol MSA 2	1	Slovenia	1.0	S. Europe	Slovenia	1.0	1989	11-35
Pol MSA 2	2	Switzerland-Slovenia	0.89	Europe	Switzerland	1.0	1963	35-61
Pol MSA 2	3	China-US-Switzerland	0.89	Global	Switzerland	1.0	1960	44-59
Pol MSA 2	4	Root	1.0	C. Europe	Switzerland	1.0	1952	47-72

408
409
410
411
412
413

Footnote: [a] Alignment partition name (see Table S4 for partition information); [b] Node number as shown in corresponding tree figures (Figures 4-5, S2-S3); [c] Internal nodes are described in terms of the countries sampled among all daughter nodes. For some nodes ISO country codes are shown instead of country names to preserve space; [d] Posterior probability of node; [e] Regions included in daughter nodes; [f] Estimated country location of node; [g] Posterior probability of location.

414 **<u>Figure Legends</u>**

415 **Figure 1. Sampling in the Fertile Crescent region.** The map in panel **(a)** indicates the
416 'Fertile Crescent' region in pink. The overlapping square indicates the location of the area
417 shown in panel **(b)**, which includes the countries sampled in this study. Yellow circles
418 indicate sites in Lebanon where sampling was performed in this study; A=Northern Lebanon
419 (Qornet el Sawda, Arez); B= Beqaa Valley. Beqaa Valley comprised three sites; North
420 (Aammiq), West (Swairi, Manara, Rashaya); Eastern (Nahle, Maqne, Knaisse). Pink circles
421 indicate sites previously sampled in Jordan: C=Northern Jordan; D=Jordan Valley.

422

423 **Figure 2.** Maximum likelihood (ML) phylogeny of 219 SRLV sequences based on capsid
424 (CA) gene nucleotide sequences**.** The tree is midpoint rooted, and has been split across two
425 panels. Taxa labels include isolate details separated by underscores as follows; accession
426 number/isolate; species (Ovis=*Ovis aries*; Capra=*Capra hircus*); country of sampling (ISO
427 two-letter country code); year of sampling. Circles next to taxa names indicate complete
428 genome sequences, strain is shown in bold for these isolates. Brackets on the right indicate
429 clades with bootstrap support that either derive from a single country or geographic region, or
430 that have previously been assigned genotype subgroup status. For these clades, text
431 associated with each bracket indicated whether the clade has been found in sheep only,
432 goats only, or both sheep and goats (mixed). Inverted commas indicate that subgroup status
433 is based on analysis of sub-genomic sequences. Asterisks indicate nodes with ML bootstrap
434 support >85%, based on 100 bootstrap replicates. The scale bar indicates evolutionary
435 distance in substitutions per site.

436

437 **Figure 3. Map showing the geographic spread of Karakul in relation to early outbreaks**
438 **of infection.** Dashed lines show export of Karakul sheep in the period 1900-1950. Early
439 outbreaks of disease associated with SRLV infection are indicated (see key), as recorded in
440 **Table S2**.

441

442 **Figure 4. Time-calibrated maximum likelihood Bayesian phylogenies of SRLV-A.**
443 Panel **(a)** shows analysis of an MSA partition in reverse transcriptase (RT), panel **(b)** shows
444 analysis of an MSA partition in integrase (IN). The scale bar indicates year dates. The key to
445 the upper left shows colour codes for the predicted country location of each node. Nodes for
446 which a country estimate could not be derived with posterior probabilities <0.5 are coloured
447 black. Error bars illustrating mean coalescence time estimates with 95% highest posterior
448 density (HPD) are shown on internal nodes.

449

17

450 **Figure 5. Time-calibrated maximum likelihood Bayesian phylogenies of SRLV-B1.**

451 Panel **(a)** shows analysis of an MSA partition in reverse transcriptase (RT), panel **(b)** shows

452 analysis of an MSA partition in integrase (IN). The scale bar indicates year dates. The key to

453 the upper left shows colour codes for the predicted country location of each node. Nodes for

454 which a country estimate could not be derived with posterior probabilities <0.5 are coloured

455 black. Error bars illustrating mean coalescence time estimates with 95% highest posterior

456 density (HPD) are shown on internal nodes.

457

458

Uncategorized References

1. **Thormar H.** 2005. Maedi-visna virus and its relationship to human immunodeficiency virus. AIDS reviews **7:**233-245.
2. **Sigurdsson B, Palsson PA.** 1958. Visna of sheep; a slow, demyelinating infection. British journal of experimental pathology **39:**519-528.
3. **Sigurdsson B, Thormar H, Pálsson PA.** 1960. Cultivation of visna virus in tissue culture. . Arch Gesamte Virusforsch **10:**328.
4. **Gudnadottir M, Palsson PA.** 1967. Transmission of maedi by inoculation of a virus grown in tissue culture from maedi-affected lungs. The Journal of infectious diseases **117:**1-6.
5. **Crawford TB, Adams DS, Cheevers WP, Cork LC.** 1980. Chronic arthritis in goats caused by a retrovirus. Science **207:**997-999.
6. **Shah C, Boni J, Huder JB, Vogt HR, Muhlherr J, Zanoni R, Miserez R, Lutz H, Schupbach J.** 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.
7. **Shah C, Huder JB, Boni J, Schonmann M, Muhlherr J, Lutz H, Schupbach J.** 2004. Direct evidence for natural transmission of small-ruminant lentiviruses of subtype A4 from goats to sheep and vice versa. Journal of virology **78:**7518-7522.
8. **Grego E, Bertolotti L, Quasso A, Profiti M, Lacerenza D, Muz D, Rosati S.** 2007. Genetic characterization of small ruminant lentivirus in Italian mixed flocks: evidence for a novel genotype circulating in a local goat population. The Journal of general virology **88:**3423-3427.
9. **Leroux C, Cruz JC, Mornex JF.** 2010. SRLVs: a genetic continuum of lentiviral species in sheep and goats with cumulative evidence of cross species transmission. Current HIV research **8:**94-100.
10. **Bertolotti L, Mazzei M, Puggioni G, Carrozza ML, Dei Giudici S, Muz D, Juganaru M, Patta C, Tolari F, Rosati S.** 2011. Characterization of new small ruminant lentivirus subtype B3 suggests animal trade within the Mediterranean Basin. The Journal of general virology **92:**1923-1929.
11. **Giammarioli M, Bazzucchi M, Puggioni G, Brajon G, Dei Giudici S, Taccori F, Feliziani F, De Mia GM.** 2011. Phylogenetic analysis of small ruminant lentivirus (SRLV) in Italian flocks reveals the existence of novel genetic subtypes. Virus genes **43:**380-384.
12. **Reina R, Grego E, Profiti M, Glaria I, Robino P, Quasso A, Amorena B, Rosati S.** 2009. Development of specific diagnostic test for small ruminant lentivirus genotype E. Veterinary microbiology **138:**251-257.
13. **Tolari F, Al-Ramadneh W, Mazzei M, Carrozza ML, Forzan M, Bandecchi P, Grego E, Rosati S.** 2013. Small ruminant lentiviruses in Jordan: evaluation of sheep and goat serological response using recombinant and peptide antigens. Tropical animal health and production **45:**1335-1340.
14. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792-1797.

513   15.   **Suyama M, Torrents D, Bork P.** 2006. PAL2NAL: robust conversion of protein
514   sequence alignments into the corresponding codon alignments. Nucleic Acids Res
515   **34:**W609-612.

516   16.   **Singer JB, Thomson EC, McLauchlan J, Hughes J, Gifford RJ.** 2018. GLUE: A
517   flexible software system for virus sequence data. bioRxiv doi:10.1101/269274.

518   17.   **Stamatakis A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic
519   analyses with thousands of taxa and mixed models. Bioinformatics **22:**2688-2690.

520   18.   **Lemey P, Rambaut A, Drummond AJ, Suchard MA.** 2009. Bayesian
521   phylogeography finds its roots. PLoS Comput Biol **5:**e1000520.

522   19.   **Zeder MA.** 2008. Domestication and early agriculture in the Mediterranean Basin:
523   Origins, diffusion, and impact. Proceedings of the National Academy of Sciences of
524   the United States of America **105:**11597-11604.

525   20.   **Gjerset B, Storset AK, Rimstad E.** 2006. Genetic diversity of small-ruminant
526   lentiviruses: characterization of Norwegian isolates of Caprine arthritis encephalitis
527   virus. The Journal of general virology **87:**573-580.

528   21.   **Grego E, Lacerenza D, Arias RR, Profiti M, Rosati S.** 2009. Serological
529   characterization of the new genotype E of small ruminant lentivirus in Roccaverano
530   goat flocks. Veterinary research communications **33 Suppl 1:**137-140.

531   22.   **Bar-Yosef O.** 1998. The Natufian culture in the Levant, threshold to the origins of
532   agriculture. Evolutionary Anthropology **6:**159-177.

533   23.   **Adams DS, Oliver RE, Ameghino E, DeMartini JC, Verwoerd DW, Houwers DJ,**
534   **Waghela S, Gorham JR, Hyllseth B, Dawson M, et al.** 1984. Global survey of
535   serological evidence of caprine arthritis-encephalitis virus infection. The Veterinary
536   record **115:**493-495.

537   24.   **Sasani F, Javanbakht J, Hemmatzadeh F, Rajabi-Moghadam M,**
538   **Aghamohammad Hassan M.** 2013. Evaluation of histopathological on maedi
539   disease with serological confirmation in North-East of Iran. Research Journal of
540   Infectious Diseases **1**.

541   25.   **Houwers DJ.** 1990. Economic Importance, Epidemiology and Control, p 83-117. *In*
542   Pétursson G, Hoff-Jørgensen R (ed), Maedi-Visna and Related Diseases, vol 10.
543   Springer US.

544   26.   **MacDiarmid SC.** 1983. Survey suggests low prevalence of caprine arthritis
545   encephalitis. Surveillance **10:**4-8.

546   27.   **Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV.** 2009. Myths and
547   misconceptions: the origin and evolution of Mycobacterium tuberculosis. Nat Rev
548   Microbiol **7:**537-544.

549   28.   **Capomaccio S, Cappelli K, Cook RF, Nardi F, Gifford R, Marenzoni ML,**
550   **Passamonti F.** 2012. Geographic structuring of global EIAV isolates: a single origin
551   for New World strains? Virus research **163:**656-659.

552   29.   **Thormar H.** 2013. The origin of lentivirus research: Maedi-visna virus. Curr HIV Res
553   **11:**2-9.

554   30.   **Ryder ML.** 1983. Sheep and Man. Gerald Duckworh & Co. Ltd, London.

555   31.   **Rocha J, Chen S Fau - Beja-Pereira A, Beja-Pereira A.** 2011. Molecular evidence
556   for fat-tailed sheep domestication.

557   32.   **Saraiva T.** 2013. The Production and Circulation of Standardized Karakul Sheep and
558   Frontier Settlement in the Empires of Hitler, Mussolini, and Salazar, p 135-150. *In*
559   Jørgensen D, Jørgensen FA, Pritchard SB (ed), New Natures. University of
560   Pittsburgh Press, Pittsburgh.

561   33.   **Young CC.** 1914. Origin of Karakul Sheep. Journal of Heredity **5:**445-446.

562   34.   **Rogers AL.** 1981. Saanen Roots: A History of the Saanen Dairy Goat Breed in the
563   United States. Dairy Goat Journal Publishing Company.

564   35.   **Cork LC, Hadlow WJ, Crawford TB, Gorham JR, Piper RC.** 1974. Infectious
565   leukoencephalomyelitis of young goats. The Journal of infectious diseases **129:**134-
566   141.

567 36. **Apetrei C, Gaufin T, Gautam R, Vinton C, Hirsch V, Lewis M, Brenchley J,**
568 **Pandrea I.** 2010. Pattern of SIVagm infection in patas monkeys suggests that host
569 adaptation to simian immunodeficiency virus infection may result in resistance to
570 infection and virus extinction. J Infect Dis **202 Suppl 3:**S371-376.

571 37. **Reina R, Grego E, Bertolotti L, De Meneghi D, Rosati S.** 2009. Genome analysis
572 of small-ruminant lentivirus genotype E: a caprine lentivirus with natural deletions of
573 the dUTPase subunit, vpr-like accessory gene, and 70-base-pair repeat of the U3
574 region. J Virol **83:**1152-1155.

575 38. **Schmid A.** 1934. Breeding of Farm Animals and Domestic Breeds in Switzerland. .
576 Frauenfeld and Leipzig : Huber &Co.

577 39. **Reina R, Berriatua E, Lujan L, Juste R, Sanchez A, de Andres D, Amorena B.**
578 2009. Prevention strategies against small ruminant lentiviruses: an update. Veterinary
579 journal **182:**31-37.

580 40. **Park JE, Son SY, Shin HJ.** 2010. Sequence comparison on gag gene of caprine
581 arthritis encephalitis virus from Korea. Virus Genes **41:**99-101.

582 41. **Oguma K, Tanaka C, Harasawa R, Kimura A, Sasaki J, Goryo M, Sentsui H.**
583 2013. Isolation of Maedi/visna Virus from a Sheep in Japan. The Journal of veterinary
584 medical science / the Japanese Society of Veterinary Science
585 doi:DN/JST.JSTAGE/jvms/13-0269 [pii].

586 42. **Gifford RJ.** 2012. Viral evolution in deep time: lentiviruses and mammals. Trends
587 Genet **28:**89-100.

588 43. **Ravazzolo AP, Reischak D, Peterhans E, Zanoni R.** 2001. Phylogenetic analysis of
589 small ruminant lentiviruses from Southern Brazil. Virus research **79:**117-123.

590 44. **L'Homme Y, Ouardani M, Levesque V, Bertoni G, Simard C, Pisoni G.** 2011.
591 Molecular characterization and phylogenetic analysis of small ruminant lentiviruses
592 isolated from Canadian sheep and goats. Virology journal **8:**271.

593 45. **Ramirez H, Glaria I, de Andres X, Martinez HA, Hernandez MM, Reina R, Iraizoz**
594 **E, Crespo H, Berriatua E, Vazquez J, Amorena B, de Andres D.** 2011.
595 Recombinant small ruminant lentivirus subtype B1 in goats and sheep of imported
596 breeds in Mexico. Veterinary journal **190:**169-172.

597 46. **Chiu IM, Yaniv A, Dahlberg JE, Gazit A, Skuntz SF, Tronick SR, Aaronson SA.**
598 1985. Nucleotide sequence evidence for relationship of AIDS retrovirus to
599 lentiviruses. Nature **317:**366-368.

600 47. **Saltarelli M, Querat G, Konings DA, Vigne R, Clements JE.** 1990. Nucleotide
601 sequence and transcriptional analysis of molecular clones of CAEV which generate
602 infectious virus. Virology **179:**347-364.

603 48. **Cork LC, Hadlow WJ, Gorham JR, Piper RC, Crawford TB.** 1974. Pathology of
604 viral leukoencephalomyelitis of goats. Acta neuropathologica **29:**281-292.

605 49. **Herrmann-Hoesing LM, Broughton-Neiswanger LE, Gouine KC, White SN,**
606 **Mousel MR, Lewis GS, Marshall KL, Knowles DP.** 2010. Evaluation of a caprine
607 arthritis-encephalitis virus/maedi-visna virus indirect enzyme-linked immunosorbent
608 assay in the serological diagnosis of ovine progressive pneumonia virus in U.S.
609 sheep. Clinical and vaccine immunology : CVI **17:**307-310.

610 50. **Herrmann LM, Hotzel I, Cheevers WP, On Top KP, Lewis GS, Knowles DP.** 2004.
611 Seven new ovine progressive pneumonia virus (OPPV) field isolates from Dubois
612 Idaho sheep comprise part of OPPV clade II based on surface envelope glycoprotein
613 (SU) sequences. Virus research **102:**215-220.

614 51. **Knowles DP, Jr., Cheevers WP, McGuire TC, Brassfield AL, Harwood WG, Stem**
615 **TA.** 1991. Structure and genetic variability of envelope glycoproteins of two antigenic
616 variants of caprine arthritis-encephalitis lentivirus. Journal of virology **65:**5744-5750.

617 52. **Kennedy RC, Eklund CM, Lopez C, Hadlow WJ.** 1968. Isolation of a virus from the
618 lungs of Montana sheep affected with progressive pneumonia. Virology **35:**483-484.

619 53. **Karr BM, Chebloune Y, Leung K, Narayan O.** 1996. Genetic characterization of two
620 phenotypically distinct North American ovine lentiviruses and their possible origin
621 from caprine arthritis-encephalitis virus. Virology **225:**1-10.

54. **Querat G, Audoly G, Sonigo P, Vigne R.** 1990. Nucleotide sequence analysis of SA-OMVV, a visna-related ovine lentivirus: phylogenetic history of lentiviruses. Virology **175:**434-447.

55. **Huang J, Sun Y, Liu Y, Xiao H, Zhuang S.** 2012. Development of a loop-mediated isothermal amplification method for rapid detection of caprine arthritis-encephalitis virus proviral DNA. Archives of virology **157:**1463-1469.

56. **Juan-juan Q, Wei G, Li-ping Z, Rong-xian S, Wen-hua X.** 2005. The Gene Cloning and Sequence Analysis of the Whole Genome of Caprine Arthritis Encephalitis Virus(CAEV)-GANSU Strain Chinese Journal of Virology **5.**

57. **Muz D, Oguzoglu TC, Rosati S, Reina R, Bertolotti L, Burgu I.** 2012. First molecular characterization of visna/maedi viruses from naturally infected sheep in Turkey. Archives of virology **158:**559-570.

58. **Ramirez H, Reina R, Bertolotti L, Cenoz A, Hernandez MM, San Roman B, Glaria I, de Andres X, Crespo H, Jauregui P, Benavides J, Polledo L, Perez V, Garcia-Marin JF, Rosati S, Amorena B, de Andres D.** 2012. Study of compartmentalization in the visna clinical form of small ruminant lentivirus infection in sheep. BMC veterinary research **8:**8.

59. **de Andres X, Ramirez H, Bertolotti L, San Roman B, Glaria I, Crespo H, Jauregui P, Minguijon E, Juste R, Leginagoikoa I, Perez M, Lujan L, Badiola JJ, Polledo L, Garcia-Marin JF, Riezu JI, Borras-Cuesta F, de Andres D, Rosati S, Reina R, Amorena B.** 2013. An insight into a combination of ELISA strategies to diagnose small ruminant lentivirus infections. Veterinary immunology and immunopathology **152:**277-288.

60. **Reina R, Mora MI, Glaria I, Garcia I, Solano C, Lujan L, Badiola JJ, Contreras A, Berriatua E, Juste R, Mamoun RZ, Rolland M, Amorena B, de Andres D.** 2006. Molecular characterization and phylogenetic study of Maedi Visna and Caprine Arthritis Encephalitis viral sequences in sheep and goats from Spain. Virus research **121:**189-198.

61. **Glaria I, Reina R, Crespo H, de Andres X, Ramirez H, Biescas E, Perez MM, Badiola J, Lujan L, Amorena B, de Andres D.** 2009. Phylogenetic analysis of SRLV sequences from an arthritic sheep outbreak demonstrates the introduction of CAEV-like viruses among Spanish sheep. Veterinary microbiology **138:**156-162.

62. **Glaria I, Reina R, Ramirez H, de Andres X, Crespo H, Jauregui P, Salazar E, Lujan L, Perez MM, Benavides J, Perez V, Polledo L, Garcia-Marin JF, Riezu JI, Borras F, Amorena B, de Andres D.** 2012. Visna/Maedi virus genetic characterization and serological diagnosis of infection in sheep from a neurological outbreak. Veterinary microbiology **155:**137-146.

63. **Barros SC, Ramos F, Duarte M, Fagulha T, Cruz B, Fevereiro M.** 2004. Genomic characterization of a slow/low maedi visna virus. Virus genes **29:**199-210.

64. **Laamanen I, Jakava-Viljanen M, Sihvonen L.** 2007. Genetic characterization of maedi-visna virus (MVV) detected in Finland. Veterinary microbiology **122:**357-365.

65. **Leroux C, Chastang J, Greenland T, Mornex JF.** 1997. Genomic heterogeneity of small ruminant lentiviruses: existence of heterogeneous populations in sheep and of the same lentiviral genotypes in sheep and goats. Archives of virology **142:**1125-1137.

66. **Germain K, Valas S.** 2006. Distribution and heterogeneity of small ruminant lentivirus envelope subtypes in naturally infected French sheep. Virus research **120:**156-162.

67. **Valas S, Benoit C, Baudry C, Perrin G, Mamoun RZ.** 2000. Variability and immunogenicity of caprine arthritis-encephalitis virus surface glycoprotein. Journal of virology **74:**6178-6185.

68. **Valas S, Benoit C, Guionaud C, Perrin G, Mamoun RZ.** 1997. North American and French caprine arthritis-encephalitis viruses emerge from ovine maedi-visna viruses. Virology **237:**307-318.

69. **Andresdottir V, Tang X, Agnarsdottir G, Andresson OS, Georgsson G, Skraban R, Torsteinsdottir S, Rafnar B, Benediktsdottir E, Matthiasdottir S, Arnadottir S,**

677   **Hognadottir S, Palsson PA, Petursson G.** 1998. Biological and genetic differences
678   between lung- and brain-derived isolates of maedi-visna virus. Virus genes **16:**281-
679   293.
680 70. **Sigurdsson B, Grimsson H, Palsson PA.** 1952. Maedi, a chronic, progressive
681   infection of sheep's lungs. The Journal of infectious diseases **90:**233-241.
682 71. **Braun MJ, Clements JE, Gonda MA.** 1987. The visna virus genome: evidence for a
683   hypervariable site in the env gene and sequence homology among lentivirus
684   envelope proteins. Journal of virology **61:**4046-4054.
685 72. **Gjerset B, Jonassen CM, Rimstad E.** 2007. Natural transmission and comparative
686   analysis of small ruminant lentiviruses in the Norwegian sheep and goat populations.
687   Virus research **125:**153-161.
688 73. **Sargan DR, Bennet ID, Cousens C, Roy DJ, Blacklaws BA, Dalziel RG, Watt NJ,
689   McConnell I.** 1991. Nucleotide sequence of EV1, a British isolate of maedi-visna
690   virus. The Journal of general virology **72 (Pt 8):**1893-1903.
691 74. **Angelopoulou K, Karanikolaou K, Papanastasopoulou M, Koumpati-Artopiou M,
692   Vlemmas I, Papadopoulos O, Koptopoulos G.** 2005. First partial characterisation
693   of small ruminant lentiviruses from Greece. Veterinary microbiology **109:**1-9.
694 75. **Pisoni G, Bertoni G, Manarolla G, Vogt HR, Scaccabarozzi L, Locatelli C, Moroni
695   P.** 2010. Genetic analysis of small ruminant lentiviruses following lactogenic
696   transmission. Virology **407:**91-99.
697 76. **Pisoni G, Moroni P, Turin L, Bertoni G.** 2007. Compartmentalization of small
698   ruminant lentivirus between blood and colostrum in infected goats. Virology **369:**119-
699   130.
700 77. **Pisoni G, Bertoni G, Puricelli M, Maccalli M, Moroni P.** 2007. Demonstration of
701   coinfection with and recombination by caprine arthritis-encephalitis virus and maedi-
702   visna virus in naturally infected goats. Journal of virology **81:**4948-4955.
703 78. **Pisoni G, Quasso A, Moroni P.** 2005. Phylogenetic analysis of small-ruminant
704   lentivirus subtype B1 in mixed flocks: evidence for natural transmission from goats to
705   sheep. Virology **339:**147-152.
706 79. **Grego E, Profiti M, Giammarioli M, Giannino L, Rutili D, Woodall C, Rosati S.**
707   2002. Genetic heterogeneity of small ruminant lentiviruses involves immunodominant
708   epitope of capsid antigen and affects sensitivity of single-strain-based immunoassay.
709   Clinical and diagnostic laboratory immunology **9:**828-832.
710 80. **Reina R, Bertolotti L, Dei Giudici S, Puggioni G, Ponti N, Profiti M, Patta C,
711   Rosati S.** 2010. Small ruminant lentivirus genotype E is widespread in Sarda goat.
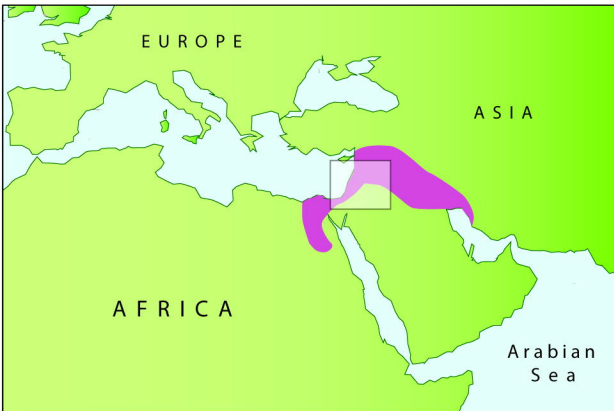712   Veterinary microbiology **144:**24-31.
713 81. **Carrozza ML, Mazzei M, Lacerenza D, Del Chiaro L, Giammarioli M, Marini C,
714   Rutili D, Rosati S, Tolari F.** 2009. Seroconversion against SU5 derived synthetic
715   peptides in sheep experimentally infected with different SRLV genotypes. Veterinary
716   microbiology **137:**369-374.
717 82. **Olech M, Rachid A, Croise B, Kuzmak J, Valas S.** 2012. Genetic and antigenic
718   characterization of small ruminant lentiviruses circulating in Poland. Virus research
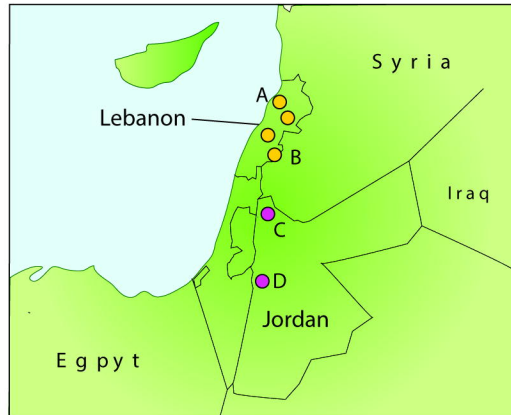719   **163:**528-536.
720 83. **Kuhar U, Barlic-Maganja D, Grom J.** 2012. Phylogenetic analysis of small ruminant
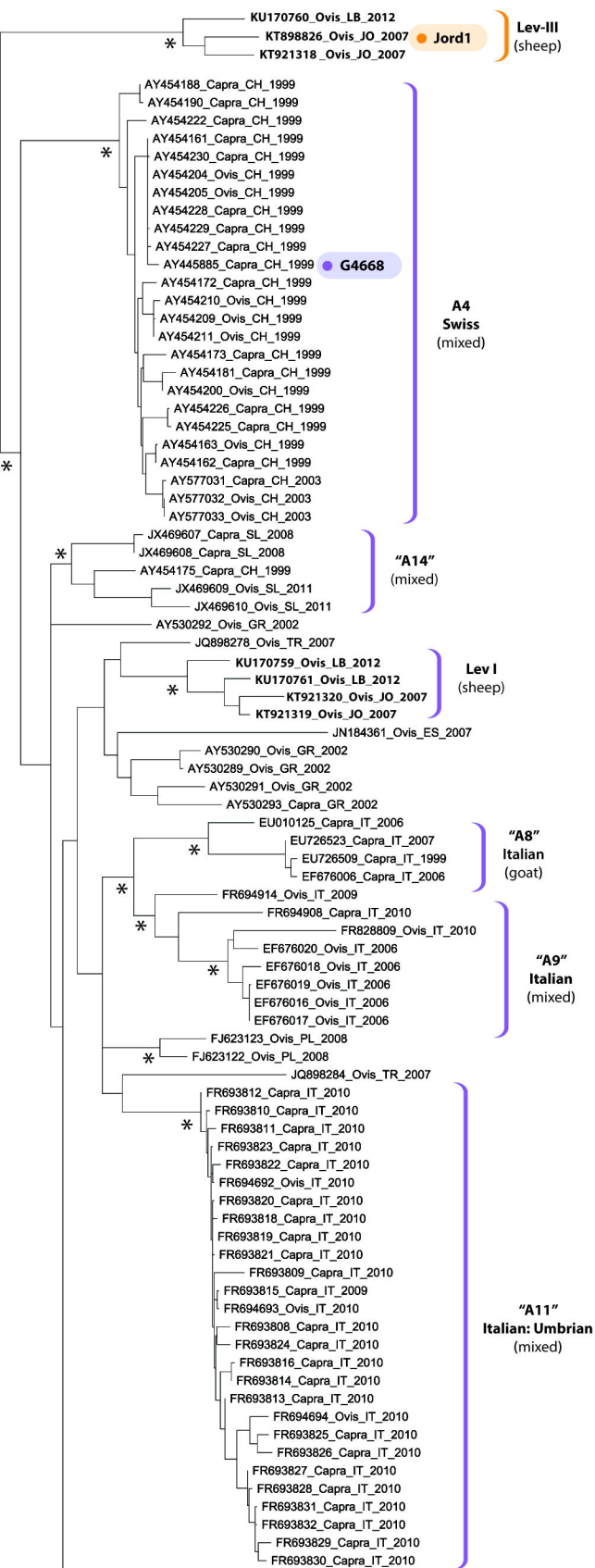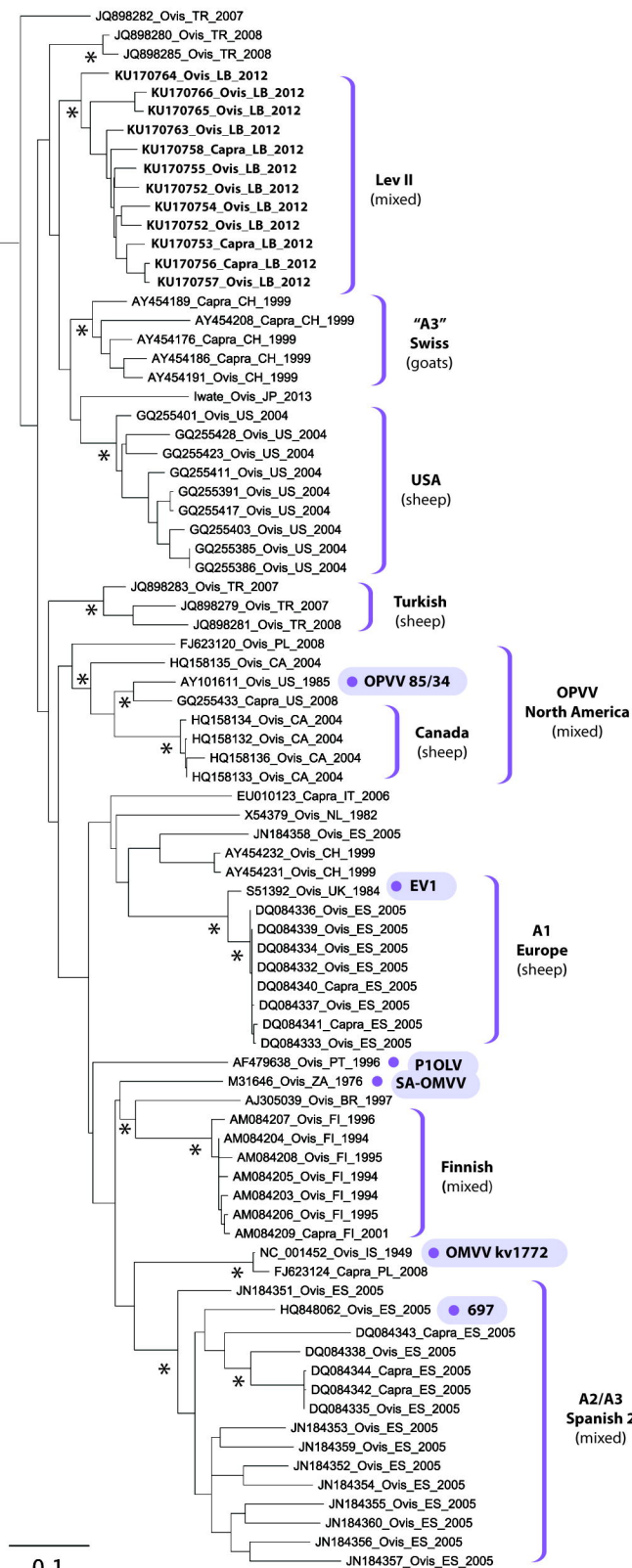721   lentiviruses detected in Slovenia. Veterinary microbiology **162:**201-206.
722

(i)

Lev-III (sheep) — Jord1

A4 Swiss (mixed) — G4668

"A14" (mixed)

Lev I (sheep)

"A8" Italian (goat)

"A9" Italian (mixed)

"A11" Italian: Umbrian (mixed)

(ii)

from (i)

Lev II (mixed)

"A3" Swiss (goats)

USA (sheep)

Turkish (sheep)

OPVV North America (mixed) — OPVV 85/34, Canada (sheep)

A1 Europe (sheep) — EV1

P1OLV, SA-OMVV

Finnish (mixed)

OMVV kv1772

A2/A3 Spanish 2 (mixed) — 697

0.1

■	Montana sheep disease (1915)
⊜	Graaf-Reinet disease (1915)
◉	Maedi-Visna (1939)
☆	La Bouhite (1942)
⊜	Maedi-Visna (1960s)
□	Bukhara, Uzbekistan
○	Halle, Germany

SRLV-B RT

location
- CH
- CN
- FR
- IT
- MX
- SL
- US

SRLV-B IN

location
- CH
- CN
- MX
- SL
- US

a

b

1950   1970   1990   2010

1940   1950   1960   1970   1980   1990   2000   2010