Molecular epidemiology of peste des petits ruminants virus emergence in critically endangered Mongolian saiga antelope and other wild ungulates

Authors

Camilla T. O. Benfield[1*], Sarah Hill[1], Munkduuren Shatar[2], Enkhtuvshin Shiilegdamba[3], Batchuluun Damdinjav[4], Amanda Fine[5], Brian Willett[6], Richard Kock[1], Arnaud Bataille[7,8]

[1]Department of Pathobiology and Population Sciences, The Royal Veterinary College, UK

[2]Department of Veterinary Services of Dundgobi province, General Authority for Veterinary Services of Mongolia (GAVS), Mongolia

[3]Wildlife Conservation Society, Ulaanbaatar, Mongolia

[4]State Central Veterinary Laboratory, Mongolia

[5]Wildlife Conservation Society, Health Program, Bronx, NY, USA

[6]MRC-University of Glasgow Centre for Virus Research, Glasgow, UK

[7]CIRAD, UMR ASTRE, F-34398 Montpellier, France

[8]ASTRE, University of Montpellier, CIRAD, INRA, Montpellier, France

[*]Corresponding author

Email: cbenfield@rvc.ac.uk (CTOB)

Abstract

Peste des petits ruminants virus (PPRV) causes disease in domestic and wild ungulates, is the target of a global eradication programme and threatens biodiversity. Understanding the epidemiology and evolution of PPRV in wildlife is important, but hampered by the paucity of wildlife-origin PPRV genomes. In this study, full PPRV genomes were generated from three Mongolian saiga antelope, one Siberian ibex and one goitered gazelle from the 2016-2017 PPRV outbreak. Phylogenetic analysis showed that for Mongolian and Chinese PPRV since 2013, the wildlife and livestock-origin genomes were closely related and interspersed. There was strong phylogenetic support for a monophyletic group of PPRV from Mongolian wildlife and livestock, belonging to clade of lineage IV PPRV from livestock and wildlife from China since 2013. Discrete diffusion analysis found strong support for PPRV spread into Mongolia from China and phylogeographic analysis indicated Xinjiang Province as the most likely origin, although genomic surveillance for PPRV is poor and lack of sampling from other regions could bias this result. Times of most recent common ancestor (TMRCA) were June 2015 (95% HPD: August 2014 – March 2016) for all Mongolian PPRV genomes and May 2016 (95% HPD: October 2015 – October 2016) for Mongolian wildlife-origin PPRV. This suggests that PPRV was circulating undetected in Mongolia for at least six months before the first reported outbreak in August 2016, and that wildlife were likely infected before livestock vaccination began in October 2016. Finally, genetic variation and positively-selected sites were identified that might be related to PPRV emergence in Mongolian wildlife. This study is the first to sequence multiple PPRV genomes from a wildlife outbreak, across several host species. Additional full PPRV genomes and associated metadata from the livestock-wildlife interface are needed to enhance the power of molecular epidemiology, support PPRV eradication and safeguard the health of the whole ungulate community.

Author Summary

Recent mass mortality of critically endangered Mongolian saiga antelope due to peste des petits ruminants virus (PPRV) has dramatically highlighted the threat this viral disease represents for biodiversity. The genome of viruses such as PPRV evolve fast, so virus genetic data gathered from infected animals can be used to trace disease spread between livestock and wildlife, and to determine if the virus is adapting to infect wildlife more efficiently. Here we obtained PPRV virus genomes from Mongolian wildlife and compared them with other published PPRV genomes. Using a molecular clock, we estimated that the disease was circulating in Mongolia well before it was first reported. Genetic analyses support the hypothesis of virus spread from livestock to wildlife, with genetic changes potentially helping infection in Asian wild ungulates. However, more PPR virus genomes and epidemiology data are needed from disease outbreaks in areas shared between livestock and wildlife to confirm these results and take efficient actions to safeguard the health of the whole ungulate community.

Introduction

Peste des petits ruminants (PPR) is a contagious viral disease of sheep and goats with high morbidity and mortality rates, which is a major barrier to sustainable small ruminant production, dependent livelihoods and economies. Consequently, PPR is the only livestock disease currently targeted by a Global Eradication Programme (GEP), which aims to rid the world of PPR by 2030 through vaccination of livestock, and thereby contribute to achieving the Sustainable Development Goals. The etiological agent, peste des petits ruminants virus (PPRV), has a broad host range, with serological or virological evidence of natural infection in a growing list of wild species within the order Artiodactyla [1-4]. PPRV infection of both captive and free-ranging wildlife may result in severe outbreaks and mortality, threatening species' survival and ecosystem integrity. PPRV has caused mass mortality of mountain caprine species categorised as vulnerable by the IUCN [5, 6], with > 1000 deaths of wild goats (*Capra aegagrus*) and sheep (*Ovis orientalis*) in Iran [7] and > 750 wild goats in Iraq [8]. Fatal PPR outbreaks have also been reported in free-ranging Sindh ibex (*Capra aegagrus blythi*) in Pakistan [9] and in ibex (*Capra ibex*) [10-12], bharal (*Pseudois nayaur*) [10-14], argali sheep (*Ovis ammon*) [10], goitered gazelle (*Gazella subgutturosa*) [10] and Przewalski's gazelle (*Procapra przewalskii*) [15] in China. To date, the most devastating impact of PPRV on biodiversity was its emergence in the critically endangered Mongolian saiga antelope (*Saiga tatarica mongolica)* in 2016-2017, which caused a mass mortality event and contributed to loss of ~80% of the population [16, 17]. In contrast, clinical disease has not been confirmed in free-ranging wildlife in Africa, despite high apparent PPRV seropositivity in wildlife populations in East Africa [18, 19]. The only published disease outbreak in free-ranging African wildlife in Africa occurred in Dorcas gazelles (*Gazella dorcas*) in Dinder National Park, Sudan [20]. However, this was not supported by field data to confirm the nature of the epidemic or event, and so whether this represents true wildlife disease remains equivocal, whilst African species in captivity have been shown to express PPR disease in zoological collections in the Middle East [21-23]. Therefore, while it is now clear that PPRV poses a threat to biodiversity, the determinants of differential disease expression among

wildlife hosts are not understood. There are also significant knowledge gaps regarding the role of wildlife in the epidemiology and evolution of PPRV. It remains unclear whether wildlife can maintain or transmit the virus to livestock, and thereby pose a threat to the PPR GEP.

It is important to assess the genetic diversity of PPRV to understand whether host range plasticity and viral virulence are linked to genetic changes in the virus. PPRV is a morbillivirus with a negative sense single-stranded RNA genome of approximately 16 kilobases, which encodes six structural proteins, the nucleocapsid (N), phosphoprotein (P), matrix (M), fusion (F), hemagglutinin (H), and polymerase (L) proteins, and two non-structural proteins, V and C. The infectivity of PPRV is mediated by its envelope glycoproteins, H and F, which are therefore key viral determinants of cellular and host tropism. H binds the morbillivirus receptors SLAM and nectin-4 on immune and epithelial host cells, respectively, while F mediates the subsequent membrane fusion events to enable cell entry. The efficiency of receptor usage and entry into target cells are likely to be critical barriers to the emergence of morbilliviruses in atypical hosts. A recent study showed that a single amino acid substitution in PPRV H enabled it to use human SLAM as an entry receptor [24]. Studies on the related morbillivirus canine distemper virus have also shown that only one or two amino acid changes in H are associated with host range expansion in nature [25, 26] or via *in vitro* adaptation [27]. The crystal structure of measles virus (MeV) H protein in complex with marmoset SLAM has been solved [28] and shows that the receptor binding domain (RBD) comprises four sites on MeV H which interact with SLAM and which are well conserved in PPRV H [24]. Several recent mutagenesis studies have also identified amino acid residues in PPRV H important for its ability to bind SLAM [29] and induce cell fusion [30]. In addition to cell entry, PPRV evidently requires efficient replicative and immune-evasive abilities for successful infection of atypical hosts, but the role of other viral genes in host range remains obscure.

PPRV is classified into 4 genetically distinct lineages, which can be discriminated based on phylogenetic analysis of short gene regions, often a few hundred nucleotides of the N gene [31, 32]. Lineage IV viruses have dominated both the host range and geographic expansion of PPRV

seen in recent years and are now replacing other lineages in many African countries [33-36]. Understanding this expansion is critical to mitigate challenges to the PPR GEP and to understand the threat of PPRV to biodiversity. To do so necessitates the phylogenetic resolution provided by full genome sequencing using high coverage high throughput sequencing technologies [37], which is particularly important since such limited molecular epidemiological data on PPRV in wildlife exists at the global level. Earlier molecular evolutionary studies of PPRV based on full genomes have included a few wildlife-origin sequences [38, 39]. However, no studies have hitherto used phylogenomic approaches to address inter-species transmission patterns of PPRV.

In Mongolia, PPR was first confirmed in August 2016 (https://wahis.oie.int/#/report-info?reportId=8043), and a full PPRV genome was generated from livestock sampled in September 2016 [40]. The outbreak in Mongolian wildlife was laboratory-confirmed in December 2016 and led to mortalities of Mongolian saiga antelope (*Saiga tatarica mongolica),* goitered gazelle *(Gazella subgutturosa)*, Siberian ibex *(Capra ibex sibirica*) and Argali (*Ovis ammon*), thought to have been caused by spillover of the virus from livestock and subsequent spread among wild ungulates (https://wahis.oie.int/#/report-info?reportId=10463, [17]). Previously, the only molecular data for PPRV from Mongolian wildlife were partial N gene sequences from two saiga antelope [17]. Here, we generated five novel full genome sequences for the PPRV which emerged in three species of Mongolian wildlife: saiga antelope, goitered gazelle and Siberian ibex. Using these sequences and all other PPRV genomes available in GenBank from both wildlife and livestock hosts, we performed phylogenetic and molecular evolutionary analyses to address PPRV emergence in Mongolian wildlife and dynamics at the livestock-wildlife interface.

136

137

138

139

140

141

Results

Tissue distribution of PPRV in Mongolian wildlife

Total RNA was extracted from tissue samples collected at necropsy from four saiga antelope, one goitered gazelle and one Siberian ibex (Table 1). To determine the tissue distribution of PPRV replication in the wildlife hosts, and select samples for whole genome sequencing, RT-PCR for a 350 nucleotide region of the N gene was performed on all available samples. Every tissue tested was RT PCR-positive (S1 Fig), with an amplicon of the expected size, namely liver and ocular swab from Siberian ibex; tongue, soft palate and lung samples from goitered gazelle, and tongue, soft palate, ocular and nasal swabs, gum scurf, mesenteric lymph node, spleen, liver, lung, heart and blood from saiga antelope. Nucleic acid sequencing showed that this N gene region was identical in all six wildlife hosts, and to the two published partial N gene sequences from saiga [17], and differed from the Mongolian livestock PPRV (KY888168.1) by two nucleotides (data not shown).

Table 1. Sampling locations and sample types for PPRV-infected wildlife.

Host ID	Species Common Name	Province	Soum or subdistrict	Name of sampling location	GPS coordinates		Sample Type (RNA)	Sample ID
					Latitude	Longitude		
Saiga_2	Mongolian saiga antelope	Khovd	Chandmani	Takhilt	47.35.55.5	93.13.51.5	lung	1
							liver	2
							spleen	3
							heart	4
							mesenteric lymph node	5
Saiga_1	Mongolian saiga antelope	Khovd	Chandmani	Nuramt	48.03.51.3	92.46.15.3	eye swab	6
							gum scurf	7
							palate scurf	8
							nasal swab	9
							tongue scurf	10
Saiga_3	Mongolian saiga antelope	Khovd	Chandmani	Suudal khuruu	47.28.09.1	93.30.48.8	spleen	11
							liver	13
							mesenteric lymph node	14
							nasal swab	15
							eye swab	16
							blood	17
Saiga_4	Mongolian saiga antelope	Gobi-Altai	Khukhmorit	Sain ust	47.16.03.6	94.07.22.4	spleen	18

							liver	19
							heart	20
							lung	21
Siberian ibex	Siberian ibex	Gobi-Altai	Tugrug	Khurengoliin ekh	45.45.21.1	95.11.57.1	liver	22
							eye swab	23
Goitered gazelle	Goitered gazelle	Khovd	Darvi	Tungalagiin us	46.48.04.9	93.40.08.0	lung	24
							tongue tissue	25
							soft palate	26

Sample ID refers to lane labels in S1 Fig.

PPRV genome sequences from Mongolian wildlife hosts

Using the Illumina NextSeq sequencing platform, five new PPRV genomes were obtained for three wildlife species: three individuals of the Mongolian saiga antelope, one goitered gazelle, and one Siberian ibex (S1 Table). The genomes were 15,954 nucleotides in length, and contained a 6-nucleotide insertion within the 5′ UTR of F gene (at position 5216 in the alignment), shared by PPRV from Mongolian livestock (KY888168.1) and Chinese lineage IV strains after 2013 [41], but not observed in other published PPRV genomes. Two of the PPRV genomes from saiga had complete nucleotide coverage across the entire genome (saiga3 and saiga4) whereas another sequence from saiga contained sequence gaps totalling 1133 nucleotides, the goitered gazelle sequence had a 33 nucleotide gap (in the M-F intergenic region) and the ibex sequence a 759 nucleotide gap (in the M-F intergenic region, plus 39nt of the F gene which was later confirmed by F gene RT-PCR). The two complete PPRV sequences from different saiga individuals differed from each other at only three nucleotide sites. Aligning the PPRV sequences from saiga with the only full PPRV genome available for Mongolian livestock (KY888168.1), showed 99.7% nucleotide identity, i.e. out of 15,954 nt sites in the genome, there were 42 (saiga 4) or 45 (saiga 3) nucleotide differences.

Evolutionary rates and lineage divergence of PPRV

Following sequence curation as described above, 76 PPRV genomes from Genbank were added to the five novel PPRV genomes generated in this study, yielding a total of 81 sequences for

phylogenetic analysis. These spanned 49 years from 1969 to December 2018, and included isolates from 24 countries. Previous phylogenomic analyses have included sequences found to be unreliable by our recombination analysis [38, 39, 41, 42], which we excluded (see methods section and S2 Table). We therefore first analysed the evolutionary rates and global lineage diversification of PPRV using our Bayesian time-scaled phylogeny of 81 genomes.

The mean evolutionary rate across the phylogeny, under an uncorrelated relaxed clock model found to be the best fit for the data, was 9.22E-4 nucleotide substitutions/site/year (95% highest posterior density (HPD) interval: 6.78E-4 - 1.17E-3) (Fig 1A). Table 2 gives the countries of origins and divergence times of PPRV lineages inferred from the Bayesian phylogenetic analysis. For lineage IV, which is expanding its geographic and host range, the median TMRCA was estimated to be 1975 (95% HPD: 1961-1985) and its country of origin was inferred as Nigeria, with moderate support (67% root state posterior probability).

Table 2. Time to the most recent common ancestor (TMRCA) and country of origin of PPRV lineages.

PPRV Lineage	TMRCA		Country of origin	
	Median	95% HPD interval	Country	Root state posterior probability
All (root)	1919	1884-1945	Benin	46.0%
I	1930	1902-1949	Benin	52.4%
II	1945	1926-1959	Benin	55.5%
III	1919	1884-1945	Benin	46.0%
IV	1975	1961-1985	Nigeria	67.0%

TMRCA and countries of origin were inferred from the Bayesian phylogeny shown in Fig 1A using FigTree. TMRCA is given to the nearest complete year. HPD= Highest Posterior Density.

Phylogenetic analysis of PPRV at the livestock-wildlife interface

Including the five novel genomes from this study, twelve wildlife-origin PPRV genomes are currently available (Table 3), although one of these (KT633939.1/ibex/China/2015-01-20) was these

196 was excluded from our phylogenetic analysis. Two cases occurred in zoological collections, and the
197 other ten PPRV genomes were from infections of free-ranging wildlife. Eleven of the wildlife PPRV
198 genomes belong to lineage IV and one to lineage III. The only countries having both wildlife and
199 livestock sequences were Mongolia and China. In contrast, three Middle Eastern countries, Israel,
200 Iraq and UAE, had full PPRV genomes available in GenBank from wildlife hosts, but no genomes
201 from infected livestock were available for these countries. Table 3 summarises the key
202 epidemiological data for the 2016-2017 outbreak in Mongolian wildlife [17, 43] and the disease
203 events associated with the other available wildlife-origin PPRV genomes.

204 **Table 3. Twelve PPRV full genomes available from wildlife host species and associated**
205 **metadata.**

Sequence ID/ GenBank Accession No.	PPRV Lineage	Date (YYYY-MM)	Country	Host species (*Latin name*)	Free-ranging/ Captive	Disease event	Key epidemiological & livestock interface data for disease events associated with wildlife-origin PPRV genomes
Saiga_1	IV	2017-01	Mongolia	Mongolian saiga antelope (*Saiga tatarica mongolica*)	Free-ranging	Mass mortality >80% population decline estimated	The first suspected (not laboratory confirmed) PPR deaths in saiga were reported by herders in December 2016 before official confirmation of PPRV infection in saiga on 27th December 2016 [17]. The first PPR outbreak in livestock in Mongolia occurred in August 2016 (OIE report 20934). Mass livestock vaccination was undertaken in Western Mongolia in October 2016 [17, 43]. The last saiga PPR case was reported in May 2017 [17].
Saiga_3 (MZ061719)	IV	2017-01	Mongolia	Mongolian saiga antelope	Free-ranging	As above	As above
Saiga_4 (MZ061720)	IV	2017-01	Mongolia	Mongolian saiga antelope	Free-ranging	As above	As above
Siberian ibex (MZ061721)	IV	2017-01	Mongolia	Siberian ibex (*Capra sibirica*)	Free-ranging	Mortality (clusters)	Clusters of cases; 24 ibex carcasses disposed by government January-June 2017 [17]. Suspected (not laboratory confirmed) cases reported in ibex in July/August 2016 in South-western part of Khovd province. The latest confirmed ibex case occurred in January 2018.
Goitered gazelle (MZ061722)	IV	2017-01	Mongolia	Goitered gazelle (*Gazella subgutturosa*)	Free-ranging	Mortality (sporadic)	Sporadic cases; 41 Goitered gazelle carcasses disposed by government January-June 2017 [17]. Post-mortem histological findings in PPRV-infected goitered gazelle given in [17].
MN121838.1	IV	2018-12	China	Przewalski's gazelle (*Procapra*	Free-ranging	Mortality (single case)	Full genome reported by [15]. Outbreak of PPRV in sheep occurred in the same area of

				przewalskii)			northwest Gansu province in October 2018 [15].
KT633939.1*	IV	2015-01	China	Ibex (*Capra ibex*)	Free-ranging	Mortality (38 ibex mortalities)	Full genome reported by [12]. 38 ibex mortalities reported in January/February 2015 [11]. Mucopurulent oculo-nasal discharge, diarrhoea and pulmonary congestion at post-mortem was reported [12].
JX217850.1	IV	2008-01	China (Tibet)	Bharal (*Pseudois nayaur*)	Free-ranging	Mortality	Full genome reported by [13]. Sequenced case showed PPR-compatible clinical signs. 19 dead bharal and 6 dead Mongolian gazelles (*Procapra gutturosa*) were found in the same location but PPRV was not confirmed in these [14]. Epidemiological evidence of PPR outbreaks in livestock nearby in September 2007 [14].
MN369542.1	IV	2018-08	UAE	Mountain gazelle (*Gazella gazella*)	Free-ranging	Mass mortality (hundreds of deaths)	Gazelles share desert grazing with free-roaming domestic small ruminants in this area. No concurrent disease was seen in domestic small ruminants.
KJ867545.1	III	1986	UAE	Dorcas Gazelle (*Gazella dorcas*)	Captive	Multi-species disease outbreak	Full genome reported by [38]. Disease outbreak in a zoological collection at Al Ain which clinically affected gazelles (Gazellinae), ibex and sheep (Caprinae) and gemsbok (Hippotraginae) [22].
MF678816.1	IV	2017-01	Israel	Nubian ibex (*Capra nubiana*)	Captive	Mortality	Over 2/3 of a captive herd of 32 Nubian ibex died following peracute/acute presentation [21]. Pathological findings differed from those typical in sheep and goats: abomasitis was seen commonly but oral and pulmonary lesions were rare.
MK408669.1	IV	2011-02	Iraq (Kurdistan)	Wild goat (*Capra aegagrus*)	Free-ranging	Mass mortality	>750 deaths in *Capra aegagrus* were reported August 2010 and February 2011 [8]. This was the first report of PPRV in Kurdistan but the affected area is close to the Turkish border where PPRV is endemic [8]. No concurrent disease in domestic ruminants was reported, likely due to annual vaccination [8]. Due to its close relationship to AJ849636.2/sheep/Turkey/2000, this strain was inferred to have been circulating in the region for more than 10 years (Hoffmann et al., 2012). In experimental challenge studies, the MK408669.1 strain was virulent and transmissible among sheep and goats [44] and caused clinical signs and onwards transmission in pigs [45].

The genomes generated in this study are shown in grey. (*): this sequence showed a significant signature of recombination, most likely associated with laboratory contamination, and was not included in this study.

209

210 Assessing the host-traited MCC tree shows that the wildlife-origin PPRV from China after
211 2013, i.e. MN121838.1 (Przewalski's gazelle), lies within the clade of livestock PPRV (Fig 1B). In
212 contrast, several other wildlife PPRV genomes lie on branches that are basal to clades circulating in
213 nearby locations. For example, PPRV from a bharal (Tibet, 2008) was basal to Chinese livestock
214 sequences from 2007, PPRV from a mountain gazelle (UAE, 2018) was basal to a clade of six
215 livestock sequences from India in 2014-2016, and PPRV from a Nubian ibex (Israel, 2017) was on a
216 long branch basal to isolates from Turkey and Iraqi Kurdistan close to the Turkish border (Fig 1B).
217 Similar phylogenetic relationships for wildlife-origin PPRV genomes were seen using ML
218 phylogenetic reconstruction (Fig 2).

219

220 **Phylogenetic analysis and TMRCA for PPRV emergence in Mongolian wildlife**

221 The five novel genomes from wild Mongolian ungulates were most closely related to PPRV
222 from a Mongolian goat (KY888168.1) sampled in September 2016, the only PPRV genome sequence
223 available from Mongolian livestock. There was strong support for monophyletic grouping of the
224 Mongolian wildlife and livestock sequences using both ML and Bayesian inference methods (Figs 1
225 and 2). There was also strong support for the grouping of all five Mongolian wildlife sequences and,
226 although there was poor support for the clade structure within the Mongolian wildlife clade, in
227 every analysis the Siberian ibex formed a sister branch to the four other wildlife sequences from
228 saiga and the goitered gazelle (Figs 1, 2, S3). The Mongolian sequences lie within a strongly
229 supported clade of lineage IV sequences from China, that includes livestock sequences from 2013-
230 2015, and a Przewalski's gazelle sequence from 2018.

231 The dates of PPRV emergence in Mongolia and its wildlife were inferred using TMRCA
232 analysis of the Bayesian time-scaled MCC tree shown in Fig 1A. The median TMRCA of the six
233 Mongolian PPRV genomes was June 2015 (95% HPD: August 2014 – March 2016). The median date
234 for the MRCA for the five Mongolian wildlife PPRV sequences was May 2016 (95% HPD: October

2015 – October 2016). The ancestor at the node linking all the Mongolian PPRV sequences with the most closely related Chinese sequences was dated to July 2013 (95% HPD: March 2013 – November 2013). To check that the data partitioning for the traited analysis did not substantially alter the TMCRA analysis, an untraited phylogeny was also analysed, and gave very similar results (S2 Fig).

Phylogeographic analysis of PPRV emergence in Mongolian wildlife

SpreaD3 was used to visualise geographic spread inferred through discrete phylogeographic analysis and identify well supported rates using Bayes Factor tests. This found that PPRV spread from China into Mongolia was very strongly supported with a Bayes Factor of 494 and associated posterior probability of 0.96 (S3 Table).

To further explore and visualise phylogeographic patterns in the data, samples were geocoded and analysed using Microreact (www.microreact.org). Samples were geocoded at the highest resolution possible. GPS sampling locations were available for all Mongolian wildlife samples (Table 1), the Mongolian livestock sample and the 2008 bharal sequence from Tibet. Of the other 33 PPRV genomes within the Chinese 2013-2018 clade, 28 samples had province-level location data, hence region centroids were used for geocoding, while country-level location only was available for five sequences, in which case the China centroid was used. An open-access interactive dynamic visualisation of our global PPRV dataset, integrating phylogenetic, spatial, temporal and host (wildlife/livestock host) data is available at the permanent link https://microreact.org/project/5WNeX14MRFvwe8YLhn5a1S/e2d5dafd (and will be updated as further genomes become publicly-available).

The map highlights the proximity of the sampling locations for the wildlife and livestock PPRV genomes in the Western Mongolian provinces of Khovd and Gobi Altai, with <40km between the livestock sample and one of the saiga antelope sampled in Khovd near the Khar-Us lake four months later (Fig 3). The Siberian ibex sample, from Tugrug soum of Gobi-Altai (Table 1), was

260 furthest from the sampling location of the Mongolian livestock (~355 km) and marginally closer to
261 the Chinese border than other detected cases.

262 The PPRV genomes KX421386.1 and KX421384.1, from December 2013, are closely
263 phylogenetically related to the Mongolian sequences, and also the geographically closest
264 sequences, from Xinjiang province of Northwest China, which borders Mongolia (Fig 3).

265

266 **Amino acid polymorphisms associated with host range expansion**

267 The five new PPRV sequences were compared with the 76 other genomes in our dataset in
268 order to identify polymorphisms of interest that might be associated with PPRV emergence in
269 Mongolian wildlife (Table 4).

270 **Table 4. Amino acid polymorphisms.**

PPRV gene	AA position	Residue present					
		Mongolian wildlife species			Mongolian livestock (KY888168.1)	China clade since 2013	Other PPRV genomes
		Saiga	Ibex	Gazelle			
H	112	S	**G**	S	S	S	S in all
	157	**R**	**R**	**R**	**R**	K	K in all
	244	**A**	**A**	**A**	**A**	T	T in all
	263	L	L	L	**F**	L	F in LIII; L in all other genomes
	315	**R**	**R**	**R**	**R**	**R**	K in all
	450	**K**	**K**	**K**	**K**	**K**	G in two LIV sequences; S in two LI sequences; R in all others
	506	D	D	D	**N**	D	D in all
	546	**S**	**S**	**S**	**S**	**S**	G in all
F	66	K	K	K	**R**	K	R in MG581412.1 (LIV); T in KR828813.1 (LIV); all others K
	518	K	K	**N**	K	K	K in LIV; R in LI, LII & LIII (except K in LII EU267274.1)
P	137	**A**	**A**	**A**	**A**	V	V is the consensus; a few sequences encode L, F or I
	285	**P**	**P**	**P**	**P**	S	S in other LIV from Asia; L in LIV from Africa and Middle East & LI, LII & LIII
	509	**L**	**L**	**L**	**L**	P	P in all; except L in two Middle Eastern LIV sequences

V	137	**A**	**A**	**A**	**A**	V	V is the consensus; a few sequences encode L, F or I
	285	**A**	**A**	**A**	**A**	V	V in all
N	484	N	N	N	**K**	N	T in four sequences (from LI and LIII); N in all other

The amino acid sequences for each PPRV gene were compared between Mongolian, closely related Chinese and other PPRV genomes to identify polymorphic sites. The amino acid positions and encoded residues for identified polymorphic sites are shown. Under 'Other PPRV genomes', lineages are referred to as LI, LII, LIII or LIV.

Among the polymorphisms observed, the Mongolian livestock PPRV (KY888168.1) encodes asparagine (N) at position 506 of PPRV H, whereas every other sequence encodes aspartate (D) as part of the completely conserved ^{505}DDD^{507} motif. The PPRV H from Siberian ibex has a glycine (G) at amino acid 112, whereas the other 80 sequences in the dataset encode serine (S). In addition, two residues in H are unique to the Mongolian PPRV sequences, including the five wildlife and one livestock sequence: arginine (R) instead of lysine (K) at amino acid 157, and hydrophobic alanine (A) instead of hydrophilic threonine (T) at position 244. We also noted several sites within the H gene where the monophyletic clade of 39 Chinese/Mongolian sequences encode common signature residues compared other PPRV sequences, namely amino acids 315 (R), 450 (K) and 546 (S) (Table 4). For the F protein, PPRV from goitered gazelle was the only genome to encode asparagine (N) at position 518, a significantly different residue from lysine (K) (in all other lineage IV genomes) or R (lineages I, II and III).

Mongolian PPRV sequences encode alanine (A) at position 137 of their P and V proteins, a residue not seen in any other PPRV isolate in the database (otherwise valine (V) is the consensus, with a few sequences encoding L, F and I). Another polymorphism was seen at amino acid 285, where proline (P) is seen only in the P proteins of Mongolian sequences instead of either Serine (S) in other lineage IV PPRV from East and South Asia or leucine (L) in lineage IV viruses from Africa and the Middle East as well as lineages I, II and III. This mutation lies after the RNA editing site, leading to a different substitution in the V protein, with Mongolian sequences encoding alanine (A)

and all other sequences encoding valine (V) at amino acid 285, which lies close to the zinc-binding domain comprising amino acids 237-280.

Molecular modelling of polymorphic amino acids in Mongolian PPRV H

To determine the location of the polymorphic sites identified in PPRV H from Mongolian wildlife and livestock, structural homology modelling was performed, based the solved crystal structure of the head domain of MeV H in complex with SLAM, the host cell entry receptor for morbilliviruses [28]. Of the polymorphic sites identified between H sequences, sites 112 and 157 were not captured by the crystal structure. Amino acid residues 244 and 263 were predicted to be distant from the SLAM binding interface and surface exposed (Fig 4A,B). Residues 506 and 546 were predicted to lie within the region that forms the SLAM binding interface (Fig 4A,B). While amino acid site 546 of H is not thought to form a direct contact with SLAM, amino acid 506 lies between two residues (D505 and D507) which in MeV H form salt bridges to K77 and R90 of marmoset SLAM, comprising site 1 of the RBD (Fig 4C) [28]. Modelling caprine SLAM in place of marmoset SLAM reveals K78 in place of K77, whereas R90 in marmoset SLAM is replaced by R91 in caprine SLAM, with the preceding P90 facing away from PPRV H (Fig 4D). Replacing G506 of MeV H with D506 (i.e. the consensus residue in PPRV H) or N506 (i.e. the substitution seen in PPRV from Mongolian livestock) could affect the interaction with SLAM due to its proximity (Fig 4D,E,F).

Selection pressure analysis

To test for positive selection, methods were used that assess the numbers of non-synonymous (dN) to synonymous (dS) nucleotide substitutions per site, with dN > dS indicative of positive selection (i.e. adaptive evolution). To assess positive selection acting on individual codons, the CDS of each PPRV gene was analysed using MEME, FUBAR, FEL and CodeML. Sites under positive selection were identified in all genes with all four methods, except for the F gene (three methods), the C gene (two methods) and the M gene (one method) (Table 5, S4 Table). The amino

acid positions identified as evolving under positive selection by all four methods were amino acid 246 of the H protein, amino acid 616 of the L protein and amino acids 52 and 101 of the P and V proteins. Of note, all methods except CodeML also identified positive selection acting on amino acid 137 of the P and V proteins.

Table 5. Amino acid sites under positive selection

Protein	MEME	FUBAR	FEL	CodeML
N	*46*, *431*	*456*	*211, 431, 467*	424, 456
P	10, 20, *28*, *52*, *79*, *83*, *92*, 101, *102*, *106*, *111*, 137, *161*, *170, 171, 176, 269, 277*, *284*, *293*, 295, *382*, 425, *509*	*51, 52*, *79*, 101, *111, 137,* 161, *171, 176,* 295, *425*	*28, 51,* 52, *71, 79,* 101, *102, 111, 137,* 161, *170, 171, 176, 284,* 295, 425	52, 101
C	3,13, *36,150,154*, 176	-	*150, 154,* 176	-
V	10, 20, *28*, *51, 52*, *83*, 101, *102*, *111,* 137, 161 *171*, *176, 235, 269, 284,* 295	*28, 51, 52, 79, 101, 102, 111, 137,* 161, *171, 176, 284,* 295	*28, 51,* 52, *79, 101, 102, 111, 137,* 161, *170, 171, 173,176, 221, 284,* 295	52, 100, 101, 152, 275, 285
M	311	-	-	-
F	3, 8, *11*, 46, 145, 368, *530*	*8*	8	-
H	21, 210, *211,* 245, 246, *288, 305*, 330, *339*, *476, 574*	*246*	246, *476, 574*	246
L	54, 68, *113, 124, 230,* 336, 349, *614*, 616, 719, 720, 899, 1343, 1708, 1901, *2038,* 2080, 2115	*614*, 616, *1257*	*614,* 616, *623, 1257*	616

Each PPRV gene was tested for site-specific selection across the phylogeny using the four analysis methods shown: Mixed Effects Model of Evolution (MEME), Fast Unconstrained Bayesian Approximation (FUBAR), Fixed Effects Likelihood (FEL) and CodeML. Amino acid positions identified as evolving under positive selection by each method are shown, using their default threshold of significance ($p \leq 0.05$ for CodeML, $p \leq 0.1$ for MEME, FUBAR and FEL). Sites with significance $0.05 \leq p < 0.1$ are shown in italics. '-' indicates no positively selected sites were identified for that gene using that method.

Methods for detecting lineage specific selection were also used to test whether the clade of Mongolian and Chinese PPRV sequences since 2013 showed evidence of positive selection. Both BUSTED and aBSREL found no evidence of selection for N, P, C, V, M, F or H genes but did find evidence of positive selection acting on the L gene of the Mongolian/Chinese PPRV clade ($p<0.05$)

(Table 6). The BUSTED result indicates L gene-wide episodic diversifying selection, i.e. evidence that at least one site on at least one test branch within the clade has experienced diversifying selection. aBSREL identified episodic diversifying selection acting on two branches, namely the single Mongolian livestock PPRV sequence (KY888168.1) and the PPRV L gene from Chinese goat (KP260624) (Table S5). FEL identified positively selected sites within each gene in the Chinese/Mongolian clade (Table 6). Some sites were identified both by lineage specific FEL and earlier by MEME in all but the C proteins, including position 137 in the P and V proteins (Table 5 and 6).

Table 6. Lineage specific selection tests.

Protein	BUSTED	FEL	aBSREL
N	-	**46**, 423, 426, 484	-
P	-	45, **83**, *98*,103, **106**, **137**, 163, **176**, 233, *258*, 263, **269**, **295**, **382**, **509**	-
C	-	*91*,129	-
V	-	45, **83**,103, 106, **137**,163, **176**, 233, 263, **269**, **295**	-
M	-	**311**	-
F	-	*187*, 518, **530**	-
H	-	**21**, 112, 162, *244*, **305**, **330**, **339**, 394, 410, 438, 440, *506*, 575, 590	-
L	+ (LRT=18.64; p= 4.5 x10^{-5})	**68**, *87*, **113**, 121, **124**, *152*, *277*, 299, **336**, 485, 617, 647, 707, ***719***, **720**, 723,1022, 1031, 1198, 1264, 1272, **1343**, *1375*, 1452, *1622*, 1656, *1715*, **1901**, 1990, 2010, **2038**, 2089	+ (see Table 7)

The clade of China/Mongolia sequences was selected as foreground branches on which to test for positive selection using the three analysis methods shown: Branch-Site Unrestricted Statistical Test for Episodic Diversification (BUSTED), Fixed Effects Likelihood (FEL) and adaptive Branch-Site Random Effects Likelihood (aBSREL). '-' denotes no evidence of positive selection, as assessed using the Likelihood Ratio Test (LRT) using the default threshold of significance for BUSTED and aBSREL ($p \leq 0.05$). For FEL, amino acid sites identified as under positive selection are indicated (using the default threshold of significance of $p \leq 0.1$). Sites with significance $0.05 \leq p < 0.1$ are shown in italics. Sites in bold were also identified by MEME in phylogeny-wide testing.

356 Eight sites at which Mongolian PPRV-specific amino acid polymorphisms had been noted in
357 either wildlife or livestock (Table 4) were also identified as positively selected by the lineage-specific
358 FEL analysis (Table 6), suggesting functional significance of these sites, namely H_112, H_244,
359 H_506, F_518, P_137, P_509, V_137 and N_484.
360
361
362
363
364

Discussion

Despite the broad host tropism and impact of PPRV across wild ungulate species, there is a paucity of wildlife-origin PPRV genomes which, along with the lack of field epidemiological data, has hindered understanding of viral evolution and dynamics at the wildlife-livestock interface. This understanding is critical in order to design effective disease control and eradication strategies and thereby support the success of the PPR Global Eradication Programme (GEP) and mitigate the threat of PPRV to both domestic and wild ungulates. In this study, full PPRV genomes were generated from three Mongolian saiga antelope, one goitered gazelle and one Siberian ibex that were part of a major mortality event in Mongolian wildlife in 2016-2017. These were analysed together with a curated set of all other genomes available in GenBank, to examine PPRV evolution and cross-species transmission. This showed strong support for monophyletic grouping of genomes from Mongolian wildlife and livestock, and for incursion of PPRV into Mongolian from China. Our TMRCA analysis also indicated that PPRV emerged in Mongolia's endangered wildlife populations before livestock vaccination was initiated, and sequence polymorphisms and signatures of adaptive evolution were identified.

Prior to our phylogenetic analysis, recombination detection algorithms were applied to our dataset and identified several potential recombinant PPRV genomes. Recombination has never been reported for PPRV and is rare among negative sense RNA viruses [46, 47]. However, bioinformatics errors or laboratory contamination can lead to false signals of recombination in PPRV [39, 48]. Previous reports have also highlighted the issue of contamination in published PPRV sequences [33, 39, 48, 49], so the most probable explanation is that the recombinant PPRV genomes we identified are the result of laboratory contamination during the process of genome sequencing. However, PPRV recombination cannot be totally ruled out and should be further explored in dedicated infection experiments. The inclusion of dubious and/or recombinant genomes in earlier phylogenomic studies may have influenced their results [38, 39, 41, 42], and we therefore first analysed PPRV molecular evolution globally.

391 The genome-wide mean evolutionary rate inferred from the Bayesian MCC phylogeny of 81
392 PPRV genomes was 9.22E-4 nucleotide substitutions/site/year (95% highest posterior density (HPD)
393 interval: 6.78E-4 - 1.17E-3), which is comparable to previous estimates [38, 39]. The TMRCA and
394 country of origin of each of the PPRV lineages were inferred from our Bayesian MCC phylogeny
395 (Table 2). In the case of lineage IV, which is becoming the predominant lineage globally, the median
396 TMRCA was estimated as 1975 (95% HPD: 1961-1985). This median date is slightly earlier than that
397 reported by Muniraju et al. using a more limited dataset of twelve genomes (1987), although the
398 HPD intervals overlap, or from partial N or F gene datasets [38, 50], and broadly equivalent to
399 estimates obtained from analysis of 53 full genomes (1968) [42].
400 All phylogeographic analyses are affected by sampling bias. Our analysis indicates with
401 moderate support that the country of origin of lineage IV was Nigeria, with the second most
402 supported origin in Benin. Previous analyses identified India as the most likely origin of lineage IV,
403 but recent sequences from Nigeria were not included in that analysis [38]. The extremely poor
404 sampling prior to the 1990s means that whilst an origin of lineage IV in Africa seems plausible, we
405 do not believe that there is sufficient resolution of genomic sequences to robustly establish the
406 country of origin. However, the detection of lineage IV PPRV in livestock in Cameroon in 1997, in
407 the Central African Republic in 2004 and in several other Northern and Eastern African countries in
408 the early 2000s, in the absence of known animal movement from Asia [33, 34, 51, 52], also supports
409 the hypothesis that lineage IV may have emerged in Africa before spreading to the Middle East and
410 Asia and then re-emerging in African countries. Additional full genomes for lineage IV viruses from
411 Africa, both historical and contemporary, would improve the phylogenetic power for more robust
412 testing of this hypothesis.
413 Phylogenetic relationships between wildlife and livestock origin PPRV were assessed across
414 the global phylogeny. With the exception of the Mongolian wildlife clade, all wildlife sequences
415 occur as isolated sequences within the broader livestock diversity. This is consistent with repeated
416 dead-end spillovers of PPRV from livestock into wildlife, but the lack of sequences from wildlife,

long branches and geographic distances between the most related livestock and wildlife sequences makes it impossible to rule out transmission from largely undetected outbreaks in wildlife into livestock. No additional PPRV genomes were available from UAE, Iraq or Israel, although 18 full genomes from livestock in Israel between 1997-2004 were sequenced previously, but unfortunately not made publicly available [42]. More dense sampling of PPRV genomes globally would likely correct sampling bias and help reveal that viruses from wild and domestic hosts cluster together according to geographic location, as seen for China and Mongolia (Figs 1 and 2), and indicated by partial N gene analysis [53].

The novel PPRV genomes from Mongolian wildlife were most closely phylogenetically related to the PPRV genome from Mongolian livestock, strongly supported in both Bayesian (Fig 1) and ML (Fig 2) phylogenies. Although this might be expected, it was not evident in earlier analysis, which showed closer phylogenetic proximity of saiga-origin PPRV to Chinese livestock as opposed to Mongolian livestock sequences, owing to insufficient phylogenetic resolution of partial N gene sequence data used (255 nucleotides) [17]. China is the only other country for which both wildlife and livestock genomes are available, with phylogenetic analysis showing PPRV from Przewalski's gazelle embedded, albeit with poor node resolution, within a well-supported clade of 32 livestock viruses sampled since 2013.

Our study is the first to sequence multiple PPRV genomes from an outbreak in wildlife. Taking advantage of this, the full PPRV genome available for Mongolian livestock [40] and the dense sampling of PPRV genomes in China since 2013 [41] we examined PPRV emergence in Mongolian wildlife. The six Mongolian PPRV genomes formed a monophyletic group within a large clade (n=39) that includes lineage IV PPRV from China sampled since 2013 (Figs 1 and 2). In China, the first PPR outbreak was in Tibet in 2007-2008 and there were subsequently no reports of PPR until its re-emergence in Xinjiang province in November 2013 [54]. PPRV from China in 2007/2008 (n=4 genomes) was more closely phylogenetically related to viruses from India (spanning 2014-2016) than to the PPRV in China since 2013, consistent with earlier reports [41, 54]. Statistical

testing of the Bayesian analysis of full genomes provided very strong phylogenetic support that PPRV spread into Mongolia from China (Bayes Factor: 494; posterior probability: 0.96), although the lack of surveillance or genomic sequencing from other neighbouring countries (e.g., Kazakhstan, Russia) means that we cannot exclude introduction via an intermediary location. Phylogeographic visualisation using Microreact showed that PPRV genomes closely related to Mongolian PPRV were from livestock in Xinjiang province. Incursion of PPRV from Xinjiang province would be consistent with epidemiological reports that initial livestock cases and the earliest suspected wildlife cases clustered at the southwestern Mongolia-China border [17, 43]. Between 2013 and 2016, PPRV cases in Argali sheep, *Capra ibex* and goitered gazelle have been confirmed across six different counties within Xinjiang province, where shared grazing provides opportunity for cross-species transmission [10, 12]. Owing to the presence of multiple PPRV-susceptible wild ungulate species in Xinjiang [55] and its extensive national borders (with Mongolia, Kazakhstan, Kyrgyzstan, India, Pakistan, Russia, Afghanistan and Tajikistan and Afghanistan), Xinjiang should be a focus for increased surveillance and sampling of PPRV across the wild and domestic ungulate community.

The TMRCA of the clade of Mongolian PPRV genomes (one livestock and five wildlife) in our analysis was June 2015 (95% HPD: August 2014 – March 2016). The TMRCA linking the Mongolian to the Chinese genomes was inferred as July 2013 (95% HPD: March 2013 – November 2013). The introduction of PPRV into Mongolia likely occurred between the TMRCA estimates for these two nodes. Therefore, the very latest date of emergence in the country that is consistent with the phylogenetic analysis is March 2016 (with 95% probability). This suggests that PPRV was circulating undetected for minimum of six months before the first reported PPR outbreak in Mongolia, which was reported in livestock in Khovd province in August 2016 (notified to OIE in September 2016, Notification report REF OIE 20934). The TMRCA analysis indicates a median date for PPRV emergence in Mongolian wildlife of May 2016 (95% HPD: October 2015 – October 2016). This suggests that PPRV infections in wildlife may have pre-dated the first confirmed wildlife case in late

December 2016, as proposed earlier based on interview data and undiagnosed wildlife mortalities [17, 43]. However, more genomic sequencing of PPRV detections in livestock during 2015 to 2016 in this region is required to confirm this, as it is possible that 'missing' livestock sequences would intersperse with the (currently monophyletic) wildlife clade, thereby changing this interpretation. Pruvot et al. reported that earliest suspected unconfirmed cases in Mongolian wildlife occurred in Siberian ibex in July/August 2016 [17]. All wildlife PPRV genomes in our analysis date from January 2017 and the clade branching structure is poorly supported in our analyses, thus no inference can be made regarding the relative timing of emergence among different wildlife species. Interestingly, however, phylogenetic reconstruction consistently showed PPRV from Siberian ibex located on a sister branch to the saiga and goitered gazelle viruses (Figs 1, 2, S2). If the phylogenetic separation observed correctly captures true patterns of ancestry, phylogenetic separation could be related to (i) some structuring of the virus between wild species, potentially related to a longer period of transmission and evolution in ibex consistent with earlier emergence in this species, and/or (ii) the phylogenetic structure seen could be related to geographical separation of the ibex sample, which was further southeast than the other sampling sites.

Based on epidemiological and ecological evidence, multiple spillover events from livestock to different wildlife populations during the Mongolian outbreak have been proposed [17]. The lack of additional livestock-origin genomes means that this cannot currently be confirmed using phylogenetic approaches, although our data are compatible with the hypothesis of spillover from domestic into wild ungulates in Mongolia. Our TMRCA analyses suggest that transmission of PPRV at the livestock-wildlife interface occurred prior to, or at the latest contemporaneous with, the livestock vaccination campaign which began in October 2016. This could explain emergence in wildlife despite vaccination of ~10.4 million small ruminant livestock in Mongolia's Western provinces, without inferring inadequacies in vaccination coverage or seroconversion [43].

At the time of the disease outbreak, it is clear that the PPRV infecting wildlife was not only virulent but already well-adapted for infection and transmission in these hosts, at least in saiga

antelope. It is possible that mutations present in the wildlife-infective PPRV strains, or their ancestral viruses, contributed to this, and therefore genomic sequences were analysed for notable sequence features and signatures of adaptive evolution. The genomes from Mongolian wildlife, as well as from Mongolian livestock and all PPRV from China since 2013, have a six nucleotide insertion in the 5'UTR of the F gene, which renders the genome 15,954 nucleotides long, instead of 15,948 nucleotides in all other PPRV strains [12, 40, 41, 56]. This insertion maintains the 'rule of six', i.e. that genome length is a multiple of six, which is necessary for efficient replication of members of the *Paramyxoviridae* family [57]. Comparing the nucleotide sequence at the insertion site shows that this is much more cytosine rich in lineage IV viruses and that a tract of six cytosines is present directly upstream of the insertion site in PPRV from sister clades to the Mongolia/China clade (S1 file alignment). Insertions in measles virus genomes are thought to arise via polymerase slippage errors when transcribing homopolymeric regions [58], suggesting a similar mechanism may have occurred in an ancestor to the 2013 China PPRV clade, and the resulting insertion has been maintained in the Mongolian PPRV. Interestingly, six nucleotide insertions in the 5'UTR of the F gene were also reported in goat-adapted rinderpest virus vaccine strains [59]. Similarly, several variants of measles virus have been identified with a net gain of six nucleotides within the M-F UTR and possibly associated with lineage emergence [60, 61]. In light of these reports, and the ability of the PPRV-F 5'UTR to enhance F gene translation [62], experiments to address the functional consequences, and potential selective advantage, of the observed insertion in the Mongolian and Chinese PPRV should be prioritised.

The CDS of each PPRV gene was assessed for non-synonymous nucleotide changes and these were correlated with the results of selection pressure analyses using multiple methods. Positive selection in viral genomes indicates adaptive evolution in response to changing fitness or functional requirements [63, 64], including infection and replication in novel hosts and evasion of host innate and adaptive immune responses, and is most likely at interaction interfaces between viral and host cell molecules.

For the H and F envelope entry proteins, polymorphisms specific to Mongolian wildlife PPRV were present at amino acid 112 in H from Siberian ibex and at amino acid 518 in F from goitered gazelle, with both sites under lineage-specific positive selection (Table 4, Table 6). More PPRV genomes would clearly be needed to assess the possibility that these are host species-specific mutations. Of note, however, are two polymorphisms in H which are specific to the six PPRV genomes from Mongolian livestock and wildlife, namely R157 (instead of K157 in all 76 other genomes) and A244 (instead of T244 in all 76 other genomes), raising the possibility that these substitutions may have helped PPRV spillover to wildlife. Whereas K157R is not a major change in physiochemical or steric properties of the amino acid, the change from the polar, uncharged amino acid threonine (T) to hydrophobic alanine (A) at position 244 of the Mongolian PPRV H protein could have more significant effects on inter-molecular interactions. The adjacent residues 234-243 form a predicted T cell epitope [65] and therefore one possibility is that T244A influences interactions with MHC I. This region of H is also subject to adaptive evolution: site 244 is detected as under positive selection in the China/Mongolia linage using FEL methods, site 245 is detected by MEME as under episodic positive selection and site 246 is identified by all four methods in phylogeny-wide testing (Table 5), and was also reported previously as a positively selected site [66]. The single genome from Mongolian livestock has two substitutions in H, L263F and D506N, which are not shared by the five wildlife-origin PPRV and should be assessed in other livestock samples. If consistently seen, D506N is particularly noteworthy since it lies at the SLAM binding interface (Fig 4) and all other PPRV strains encode aspartate (D) at this site [24]. Neighbouring residues 505N and 507N are completely conserved among morbilliviruses [24], and comprise site 1 of the RBD (Fig 4), forming salt bridges to SLAM in the crystal structure for MeV H [28] and in homology models for PPRV H [4, 66]. Therefore this residue could be functionally important for infectivity and viral tropism. Substitutions already present in the Chinese lineage of PPRV since 2013 may also have contributed to emergence in Mongolian wildlife, including unique residues at sites 315, 450 and 546 of the H protein. At position 546, all sequences from the Mongolia/China clade encode serine, as

547 does MeV at the equivalent position, whereas every other PPRV strain in the phylogeny encodes
548 glycine. This residue is located adjacent to hydrophobic site 4 of the RBD (Fig 4) but is not thought
549 to form a direct contact with SLAM [28].

550 The P and V proteins antagonise the interferon (IFN) response via intermolecular
551 interactions with host STAT1, STAT2, Tyk2 and Jak1 proteins [67-69]. Potentially, Mongolian PPRV-
552 specific residues in the P (137, 285, 509) and V proteins (137, 285), particularly those under positive
553 selection (i.e. 137 in P and V; 509 in P) might affect evasion of host innate immunity and emergence.
554 In addition to its role in genome encapsidation and replication, the N protein was shown recently
555 also to inhibit type I IFN by binding IRF3 [70]. We identified residues 46 and 426 in N as positively
556 selected, with both MEME and FEL (Tables 5 and 6), as well as a Mongolian livestock-specific
557 residue and lineage-specific positive selection at site 484 (Tables 4 and 6). The L gene, although
558 lacking any Mongolian-PPRV specific polymorphisms, was the only gene identified as positively
559 selected by all three methods when the China/Mongolia clade was tested (Table 6), suggesting this
560 catalytic polymerase protein has undergone adaptive evolution in this clade. Functional studies
561 including pseudotyping (to assess receptor usage), minigenome (to assess replication), immune
562 reporter and intermolecular binding assays are beyond the scope of the current study but should be
563 planned to define the role of the mutations identified here in lineage IV PPRV expansion.

564 In summary, our study has provided insights into PPRV transmission at the livestock-wildlife
565 interface. More PPRV full genomes are required in order to strengthen molecular epidemiological
566 studies. The occurrence of cases in wildlife should serve as a trigger to initiate local
567 surveillance/sampling in livestock and the surveillance and opportunistic sampling for wildlife
568 disease events should be increased. However, quantifying the direction and extent of pathogen
569 transmission in multi-host systems is challenging, even in the case of extremely detailed
570 longitudinal study systems with pathogen genomic and host life history data [71]. There is thus a
571 need for other types of data, including interspecific contact rates and pathways, and enhanced
572 epidemiological and serological data, to enable approaches which integrate these with genomics

573 [71-73] and can lead to improved understanding of interspecific PPRV transmission dynamics and
574 the epidemiological roles of wildlife species.
575

576 ## Materials and Methods

577 ### Sample collection

578 Samples from five wild Mongolian ungulates suspected of being infected with PPRV were all
579 collected in January 2017 during an emergency field mission to urgently respond, assess and advise
580 the Mongolian Government through the National Emergency Committee. The mission involved
581 FAO's Crisis Management Centre-Animal Health (CMC-AH), the Wildlife Conservation Society, the
582 Royal Veterinary College (RVC), the Veterinary and Animal Breeding Agency (VABA) and the State
583 Central Veterinary Laboratory (SCVL) [43]. Tissues were collected from fresh carcass necropsies
584 from three saiga antelope, one goitered gazelle and one Siberian ibex. GPS locations of sampled
585 animals are provided in Table 1. Total RNA was extracted from tissues using a NucleoSpin RNA
586 Virus mini kit (Macherey-Nagel, 740956) at SCVL, Ulaanbaatar. Samples were then imported to
587 RVC under APHA and CITES import permits.
588

589 ### RT-PCR

590 One-step RT-PCR was performed on 2ul total RNA using the SuperScript IV One-Step PCR kit
591 (ThermoFisher, Catalog No. 12594025) either to amplify nucleotides 1232-1583 of the PPRV N gene
592 using published primers NP3 and NP4 (Couacy-Hymann et al 2002) or to amplify the full length H
593 gene (with primers 5'-CTCCACGCTCCACCACAC-3' and 5'-CTCGGTGGCGACTCAAGG-3') or the full
594 length F gene (with primers 5'- GCTATGCGGCCGCACCATGACGCGGGTCGCAATYTT-3' and 5'-
595 GGTGAGGATCCCTACAGTGATCTTACGTACGAC-3').
596

597 ### Whole genome next generation sequencing

598 Total RNA from tissues was converted to cDNA with SuperScript IV and sequencing libraries
599 prepared with the Nextera XT DNA Library Preparation Kit (Illumina, FC-131-1024). Sequencing was
600 performed using the Illumina NextSeq platform with 150bp paired-end reads. Sequencing data was
601 mapped to a reference PPRV genome using BWA [74] and then consensus calling was performed
602 using SAMtools [75]. This was corroborated by removing host reads and then undertaking *de novo*
603 assembly with SPAdes [76], and then both outputs were aligned to confirm PPRV genome
604 sequences.
605
606 **PPRV full genome dataset curation and recombination analysis**
607 In addition to the five novel PPRV genomes, all full PPRV genomes available in Genbank were
608 downloaded (last accessed 22/11/2020). Vaccine sequences were removed from the dataset, as well
609 as one sequence noted in GenBank as multiply passaged in cell culture (MN369543.1). Sequence
610 alignment was performed using MAFFT [77] in the Geneious software package, followed by manual
611 editing. TempEst was run to assess temporal signal in the data and evidenced clock-like evolution
612 (S3 Fig) [78]. TempEst also identified an outlier sequence (KJ867543.1, lineage III) that was removed
613 prior to phylogenetic analysis. The alignment of 84 PPRV genomes was then analysed using the
614 Recombination Detection Program (RDP) version 4.101 [79], using seven different recombination
615 detection methods (RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan and 3Seq) and default
616 settings. Signatures of recombination were detected in three sequences
617 (KR828814.1/goat/Nigeria/2012-05-09, KJ867541.1/goat/Ethiopia/2010 and
618 KT633939.1/ibex/China/2015-01-20) using all seven detection algorithms ($p<0.01$) (S2 Table). These
619 are most likely the result of laboratory contamination [47] and these sequences were considered
620 unreliable and removed from our dataset, leaving a total of 81 PPRV genomes (alignment provided
621 in S1 File).
622
623 **Phylogenetic analysis**

The General Time Reversible (GTR) nucleotide substitution model with gamma-distributed variable rates (G) and some invariable sites (+I) best fitted our dataset, according to the Baysian Information Criterion (BIC) values calculated using MEGA7 [80]. Maximum likelihood (ML) phylogenetic reconstruction was performed using PhyML with a GTR nucleotide substitution model and 100 bootstrap replicates. Phylogenetic analysis was also performed using a Bayesian Markov Chain Monte Carlo (MCMC) framework using BEAUti and BEAST v1.10.4 [81] and run via the CIPRES server. Prior to traiting the sequences, marginal likelihood estimation was performed using path sampling/stepping-stone sampling to choose the most appropriate combination of tree model (coalescent constant size or coalescent GMRF Bayesian skyride) and clock models (strict or uncorrelated relaxed) using a GTR nucleotide substitution model (4 gamma categories, estimated base frequencies and no codon partitioning). MCMC outputs from different runs were evaluated and convergence confirmed using Tracer v1.7.1 [82]. A model with a coalescent constant size tree prior and uncorrelated relaxed clock (lognormal distribution) [83] was determined to be the best fit for the data, based on the log marginal likelihood estimates from path sampling/stepping-stone sampling. This model of nucleotide evolution was used for subsequent analysis of the discrete traits, 'host category (livestock or wildlife)' and 'country'. Asymmetric substitution models were selected for the discrete traits since these are the biologically more plausible scenario of virus transmission. Bayesian Stochastic Search Variable Selection (BSSVS) was also used, which limits the number of rates to those which adequately explain the phylogenetic diffusion process. At least two independent MCMC chains, of 40 million steps each, were performed for each analysis, and Tracer v1.7.1 was used to confirm that the MCMC chains converged at the same level and assess effective sample sizes (ESS). LogCombiner was used to combine the output of the independent BEAST runs to generate tree and log files for analysis. Maximum clade credibility (MCC) trees were generated using TreeAnnotator v1.10.4. and Figtree v1.4.4 was used to visualise and interpret MCC trees and derive TMRCA estimates. Evolutionary rates (ucld.mean) were taken from the combined log files analysed in Tracer v1.7.1. MCC trees and geocoded metadata were imported into Microreact to

visualise temporal and geographic spread [84]. The MCC tree and log files from the BEAST analysis were uploaded to the SpreaD3 software (Spatial Phylogenetic Reconstruction of EvolutionAry Dynamics using Data-Driven Documents (D3) [85], in order to visualize the output from the BSSVS procedure and compute Bayes Factors for transitions. For country transitions, each country was assigned one latitude and longitude coordinate, either the precise sampling location for those countries with a single PPRV sequence in the dataset (Kenya, Tibet, Ghana), the GPS location of the saiga3 sample for Mongolia, or the country centroid coordinates (worldmap.harvard.edu) for other countries with >1 sequence.

Detection of selection pressures

Multiple analysis methods were implemented to detect positive selection in our phylogeny. Fast Unconstrained Bayesian Approximation (FUBAR, [86]) and Fixed Effects Likelihood (FEL, [87]) both available in datamonkey.org [88] were used to detect positive selection at individual sites across the whole PPRV phylogeny. Mixed Effects Model of Evolution (MEME, [89]) analysis was also performed on datamonkey.org, with the capability to identify sites under episodic selection (i.e. in a subset of branches) as well as under pervasive selection [64, 89]. In addition, the ratio of non-synonymous to synonymous nucleotide substitutions (ω = dN/dS) was estimated for different selection models (in which the ω ratio varies among codons) using CodeML as implemented in EasyCodeML [90]. The model M7 (beta; no positive selection) was compared to the model M8 (beta&ω; positive selection) for each gene using likelihood ratio tests (LRTs) [91, 92] (S4 Table). If model M8 was more likely than M7, the Bayes empirical Bayes (BEB) method [93] was used to calculate the posterior probabilities for site classes and identify sites under positive selection. Finally, three methods were implemented in datamonkey.org to detect linage-specific selection, using user-defined PhyML maximum likelihood trees and alignments for the CDS of each gene. The monophyletic clade of Mongolian/Chinese sequences was selected *a priori* to test for selection acting on these branches. Adaptive Branch-Site Random Effects Likelihood (aBSREL, [94]) was

676 used to detect branches under positive selection. Branch-Site Unrestricted Statistical Test for
677 Episodic Diversification (BUSTED, [95]) was also used to test for gene wide lineage-specific positive
678 selection. Specific sites under selection in the selected clade were identified using Fixed Effects
679 Likelihood (FEL, [87]).
680
681 **Protein homology modelling**
682 The predicted structures of PPRV Hs were modelled by submitting the amino acid sequences of
683 each H to the SWISS_MODEL automated protein structure homology modelling server in
684 "alignment" mode [96]. Structures were visualised from the .pdb files using Swiss-pdb viewer.
685
686 ## Figure captions
687 **Fig 1. Bayesian time-scaled Maximum Clade Credibility Trees using country (A) or host category**
688 **(B) partitions.**
689 Bayesian phylogenetic analysis (n=81 genomes) was run using BEAST v1.10.4 and trees are the
690 combined output of three (A) or two (B) independent MCMC chains, visualized in FigTree. x-axis
691 shows date. Branches are colour-coded by country (A) or host (B) as inferred using discrete trait
692 analyses. Lineages, referred to as LI, LII, LIII or LIV, are shown. Arrows to the x-axis in (A) show
693 ancestral nodes and corresponding TMRCAs for different lineages. * indicates posterior probability
694 > 0.9 at the node opposite.
695
696 **Fig 2. Maximum likelihood phylogeny of PPRV genomes.**
697 81 PPRV genomes were analysed using PhyML with a GTR nucleotide substitution model and 100
698 bootstrap replicates. The novel genomes from this study are shown in blue. Lineages, referred to as
699 LI, LII, LIII or LIV, are shown. Scale bar shows nucleotide substitutions per site. * indicates bootstrap
700 proportion > 0.9 at the node opposite.

Fig 3. Phylogeographic visualization.

An MCC phylogenetic tree (81 sequences) was uploaded to Microreact together with geocoded locations of PPRV genomes and metadata to produce the dynamic visualization of phylogenetic, spatial and temporal relationships of the global PPRV dataset. The figure shows one view available at the Microreact project link (https://microreact.org/project/5WNeX14MRFvwe8YLhn5a1S/e2d5dafd), showing the location (A), phylogenetic relationship (B) and timeline (C) for the clade of PPRV genomes from Mongolia and China since 2013. Symbol colour denotes country of origin (Mongolia in green and China in blue). The symbol shape denotes different resolutions of geocoding for genomes: stars denotes samples with GPS coordinates; triangles denote region centroids; squares denote country centroids and a circle on the map view indicates multiple genomes from the same location. The ringed samples in panels (A)-(C) show samples from Xinjiang province, including KX421386.1 and KX421384.1. The map shown in (A) uses base map and data from OpenStreetMap and OpenStreetMap Foundation, available under the Open Database License, with tiles from the Mapbox mapping platform, used via the freely-available Microreact application.

Fig 4. Homology modelling of the PPRV H-SLAM complex.

Modelling was performed using SWISS_MODEL using the published crystal structure of MeV H bound to marmoset SLAM to show positions of Mongolian-specific amino acid polymorphisms in PPRV H. Side (A) and top (B) orthogonal views of PPRV H bound to caprine SLAM (capSLAM) showing residues for PPRV H from saiga antelope (244A, 263L, 506D, 546S). Residues at site 1 of the RBD for (C) MeV H and marmoset SLAM (maSLAM), (D) PPRV H 506D and capSLAM, (E) PPRV H 506N and capSLAM, or (F) overlayed image of panels C and D. H proteins are shown in purple and SLAM proteins in turquoise.

Acknowledgements

Support for field sample collection and multi-sectoral coordination was provided by the Wildlife Conservation Society, the Morris Animal Foundation's Betty White Wildlife Rapid Response Fund, the FAO Crisis Management Centre-Animal Health and through the Science for Nature and People Partnership (SNAPP), a partnership of The Nature Conservancy, the Wildlife Conservation Society (WCS), and the National Center for Ecological Analysis and Synthesis (NCEAS) at University of California, Santa Barbara.

Special thanks to Khovd and Gobi-Altai Province Veterinary Agencies for their generous support for wildlife field investigations and response, and to Dr. Munkhbaatar Batsukh and Dr. Burenbayasakh Ravdan for their support.

References

1. Aziz Ul R, Wensman JJ, Abubakar M, Shabbir MZ, Rossiter P. Peste des petits ruminants in wild ungulates. Trop Anim Health Prod. 2018;50(8):1815-9. Epub 2018/06/09. doi: 10.1007/s11250-018-1623-6. PubMed PMID: 29881925.

2. Fine AE, Pruvot M, Benfield CTO, Caron A, Cattoli G, Chardonnet P, et al. Eradication of Peste des Petits Ruminants Virus and the Wildlife-Livestock Interface. Front Vet Sci. 2020;7:50. Epub 2020/04/02. doi: 10.3389/fvets.2020.00050. PubMed PMID: 32232059; PubMed Central PMCID: PMCPMC7082352.

3. Munir M. Role of wild small ruminants in the epidemiology of peste des petits ruminants. Transbound Emerg Dis. 2014;61(5):411-24. Epub 2013/01/12. doi: 10.1111/tbed.12052. PubMed PMID: 23305511.

4. Dou Y, Liang Z, Prajapati M, Zhang R, Li Y, Zhang Z. Expanding Diversity of Susceptible Hosts in Peste Des Petits Ruminants Virus Infection and Its Potential Mechanism Beyond. Front Vet Sci. 2020;7:66. Epub 2020/03/18. doi: 10.3389/fvets.2020.00066. PubMed PMID: 32181263; PubMed Central PMCID: PMCPMC7059747.

5. Weinberg P, Ambarli H. Capra aegagrus. The IUCN Red List of Threatened Species 2020: e.T3786A22145942. https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T3786A22145942.en. Downloaded on 22 April 2021. 2020.

6. Michel S, Ghoddousi A. Ovis vignei (errata version published in 2021). The IUCN Red List of Threatened Species 2020: e.T54940655A195296049. https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T54940655A195296049.en. Downloaded on 22 April 2021. 2020.

7. Marashi M, Masoudi S, Moghadam MK, Modirrousta H, Marashi M, Parvizifar M, et al. Peste des Petits Ruminants Virus in Vulnerable Wild Small Ruminants, Iran, 2014-2016. Emerg Infect Dis. 2017;23(4):704-6. Epub 2017/03/23. doi: 10.3201/eid2304.161218. PubMed PMID: 28322692; PubMed Central PMCID: PMCPMC5367395.

765 8. Hoffmann B, Wiesner H, Maltzan J, Mustefa R, Eschbaumer M, Arif FA, et al. Fatalities in
766 wild goats in Kurdistan associated with Peste des Petits Ruminants virus. Transbound Emerg Dis.
767 2012;59(2):173-6. Epub 2011/11/15. doi: 10.1111/j.1865-1682.2011.01270.x. PubMed PMID: 22074184.
768 9. Abubakar M, Rajput ZI, Arshed MJ, Sarwar G, Ali Q. Evidence of peste des petits ruminants
769 virus (PPRV) infection in Sindh Ibex (Capra aegagrus blythi) in Pakistan as confirmed by detection of
770 antigen and antibody. Trop Anim Health Prod. 2011;43(4):745-7. Epub 2011/01/12. doi:
771 10.1007/s11250-010-9776-y. PubMed PMID: 21221782.
772 10. Li J, Li L, Wu X, Liu F, Zou Y, Wang Q, et al. Diagnosis of Peste des Petits Ruminants in Wild
773 and Domestic Animals in Xinjiang, China, 2013-2016. Transbound Emerg Dis. 2017;64(6):e43-e7.
774 Epub 2017/01/20. doi: 10.1111/tbed.12600. PubMed PMID: 28101989.
775 11. Xia J, Zheng XG, Adili GZ, Wei YR, Ma WG, Xue XM, et al. Sequence analysis of peste des
776 petits ruminants virus from ibexes in Xinjiang, China. Genet Mol Res. 2016;15(2). Epub 2016/06/21.
777 doi: 10.4238/gmr.15027783. PubMed PMID: 27323119.
778 12. Zhu Z, Zhang X, Adili G, Huang J, Du X, Zhang X, et al. Genetic Characterization of a Novel
779 Mutant of Peste Des Petits Ruminants Virus Isolated from Capra ibex in China during 2015. Biomed
780 Res Int. 2016;2016:7632769. Epub 2016/03/22. doi: 10.1155/2016/7632769. PubMed PMID:
781 26998489; PubMed Central PMCID: PMCPMC4779526.
782 13. Bao J, Wang Q, Parida S, Liu C, Zhang L, Zhao W, et al. Complete genome sequence of a
783 Peste des petits ruminants virus recovered from wild bharal in Tibet, China. J Virol.
784 2012;86(19):10885-6. Epub 2012/09/12. doi: 10.1128/JVI.01503-12. PubMed PMID: 22966182;
785 PubMed Central PMCID: PMCPMC3457324.
786 14. Bao J, Wang Z, Li L, Wu X, Sang P, Wu G, et al. Detection and genetic characterization of
787 peste des petits ruminants virus in free-living bharals (Pseudois nayaur) in Tibet, China. Res Vet Sci.
788 2011;90(2):238-40. Epub 2010/07/02. doi: 10.1016/j.rvsc.2010.05.031. PubMed PMID: 20591454.

789 15. Li L, Cao X, Wu J, Dou Y, Meng X, Liu D, et al. Epidemic and evolutionary characteristics of
790 peste des petits ruminants virus infecting Procapra przewalskii in Western China. Infect Genet Evol.
791 2019;75:104004. Epub 2019/08/16. doi: 10.1016/j.meegid.2019.104004. PubMed PMID: 31415822.
792 16. Aguilar XF, Fine AE, Pruvot M, Njeumi F, Walzer C, Kock R, et al. PPR virus threatens wildlife
793 conservation. Science. 2018;362(6411):165-6. Epub 2018/10/13. doi: 10.1126/science.aav4096.
794 PubMed PMID: 30309937.
795 17. Pruvot M, Fine AE, Hollinger C, Strindberg S, Damdinjav B, Buuveibaatar B, et al. Outbreak
796 of Peste des Petits Ruminants among Critically Endangered Mongolian Saiga and Other Wild
797 Ungulates, Mongolia, 2016-2017. Emerg Infect Dis. 2020;26(1):51-62. Epub 2019/12/20. doi:
798 10.3201/eid2601.181998. PubMed PMID: 31855146; PubMed Central PMCID: PMCPMC6924898.
799 18. Mahapatra M, Sayalel K, Muniraju M, Eblate E, Fyumagwa R, Shilinde L, et al. Spillover of
800 Peste des Petits Ruminants Virus from Domestic to Wild Ruminants in the Serengeti Ecosystem,
801 Tanzania. Emerg Infect Dis. 2015;21(12):2230-4. Epub 2015/11/20. doi: 10.3201/eid2112.150223.
802 PubMed PMID: 26583961; PubMed Central PMCID: PMCPMC4672450.
803 19. Fernandez Aguilar X, Mahapatra M, Begovoeva M, Kalema-Zikusoka G, Driciru M,
804 Ayebazibwe C, et al. Peste des Petits Ruminants at the Wildlife-Livestock Interface in the Northern
805 Albertine Rift and Nile Basin, East Africa. Viruses. 2020;12(3). Epub 2020/03/12. doi:
806 10.3390/v12030293. PubMed PMID: 32156067; PubMed Central PMCID: PMCPMC7150925.
807 20. Asil RM, Ludlow M, Ballal A, Alsarraj S, Ali WH, Mohamed BA, et al. First detection and
808 genetic characterization of peste des petits ruminants virus from dorcas gazelles "Gazella dorcas" in
809 the Sudan, 2016-2017. Arch Virol. 2019;164(10):2537-43. Epub 2019/07/17. doi: 10.1007/s00705-019-
810 04330-w. PubMed PMID: 31309291.
811 21. Berkowitz A, Avni-Magen N, Bouznach A, Waner T, Litvak A, Friedgut O, et al. Pathological
812 and molecular characterisation of peste des petits ruminants in Nubian ibex (Capra nubiana) in
813 Israel. Arch Virol. 2019;164(8):1997-2003. Epub 2019/05/16. doi: 10.1007/s00705-019-04269-y.
814 PubMed PMID: 31089959.

22. Furley CW, Taylor WP, Obi TU. An outbreak of peste des petits ruminants in a zoological collection. Vet Rec. 1987;121(19):443-7. Epub 1987/11/07. doi: 10.1136/vr.121.19.443. PubMed PMID: 3424615.

23. Kinne J, Kreutzer R, Kreutzer M, Wernery U, Wohlsein P. Peste des petits ruminants in Arabian wildlife. Epidemiol Infect. 2010;138(8):1211-4. Epub 2010/01/14. doi: 10.1017/S0950268809991592. PubMed PMID: 20067659.

24. Abdullah N, Kelly JT, Graham SC, Birch J, Goncalves-Carneiro D, Mitchell T, et al. Structure-Guided Identification of a Nonhuman Morbillivirus with Zoonotic Potential. J Virol. 2018;92(23). Epub 2018/09/21. doi: 10.1128/JVI.01248-18. PubMed PMID: 30232185; PubMed Central PMCID: PMCPMC6232486.

25. Nikolin VM, Olarte-Castillo XA, Osterrieder N, Hofer H, Dubovi E, Mazzoni CJ, et al. Canine distemper virus in the Serengeti ecosystem: molecular adaptation to different carnivore species. Mol Ecol. 2017;26(7):2111-30. Epub 2016/12/09. doi: 10.1111/mec.13902. PubMed PMID: 27928865; PubMed Central PMCID: PMCPMC7168383.

26. Ohishi K, Suzuki R, Maeda T, Tsuda M, Abe E, Yoshida T, et al. Recent host range expansion of canine distemper virus and variation in its receptor, the signaling lymphocyte activation molecule, in carnivores. J Wildl Dis. 2014;50(3):596-606. Epub 2014/05/09. doi: 10.7589/2013-09-228. PubMed PMID: 24807184.

27. Bieringer M, Han JW, Kendl S, Khosravi M, Plattet P, Schneider-Schaulies J. Experimental adaptation of wild-type canine distemper virus (CDV) to the human entry receptor CD150. PLoS One. 2013;8(3):e57488. Epub 2013/04/05. doi: 10.1371/journal.pone.0057488. PubMed PMID: 23554862; PubMed Central PMCID: PMCPMC3595274.

28. Hashiguchi T, Ose T, Kubota M, Maita N, Kamishikiryo J, Maenaka K, et al. Structure of the measles virus hemagglutinin bound to its cellular receptor SLAM. Nat Struct Mol Biol. 2011;18(2):135-41. Epub 2011/01/11. doi: 10.1038/nsmb.1969. PubMed PMID: 21217702.

840 29. Meng X, Zhu X, Alfred N, Zhang Z. Identification of amino acid residues involved in the
841 interaction between peste-des-petits-ruminants virus haemagglutinin protein and cellular
842 receptors. J Gen Virol. 2020;101(3):242-51. Epub 2019/12/21. doi: 10.1099/jgv.0.001368. PubMed
843 PMID: 31859612; PubMed Central PMCID: PMCPMC7416607.
844 30. Gallo G, Conceicao C, Tsirigoti C, Willett B, Graham SC, Bailey D. Application of error-prone
845 PCR to functionally probe the morbillivirus Haemagglutinin protein. J Gen Virol. 2021. Epub
846 2021/03/20. doi: 10.1099/jgv.0.001580. PubMed PMID: 33739251.
847 31. Couacy-Hymann E, Roger F, Hurard C, Guillou JP, Libeau G, Diallo A. Rapid and sensitive
848 detection of peste des petits ruminants virus by a polymerase chain reaction assay. J Virol Methods.
849 2002;100(1-2):17-25. Epub 2001/12/18. doi: 10.1016/s0166-0934(01)00386-x. PubMed PMID:
850 11742649.
851 32. Kumar KS, Babu A, Sundarapandian G, Roy P, Thangavelu A, Kumar KS, et al. Molecular
852 characterisation of lineage IV peste des petits ruminants virus using multi gene sequence data. Vet
853 Microbiol. 2014;174(1-2):39-49. Epub 2014/09/25. doi: 10.1016/j.vetmic.2014.08.031. PubMed PMID:
854 25248690.
855 33. Dundon WG, Diallo A, Cattoli G. Peste des petits ruminants in Africa: a review of currently
856 available molecular epidemiological data, 2020. Arch Virol. 2020;165(10):2147-63. Epub 2020/07/13.
857 doi: 10.1007/s00705-020-04732-1. PubMed PMID: 32653984; PubMed Central PMCID:
858 PMCPMC7497342.
859 34. Libeau G, Diallo A, Parida S. Evolutionary genetics underlying the spread of peste des petits
860 ruminants virus. Animal Frontiers. 2014;4(1):14-20. doi: 10.2527/af.2014-0003.
861 35. Tounkara K, Bataille A, Adombi CM, Maikano I, Djibo G, Settypalli TBK, et al. First genetic
862 characterization of Peste des Petits Ruminants from Niger: On the advancing front of the Asian
863 virus lineage. Transbound Emerg Dis. 2018;65(5):1145-51. Epub 2018/07/26. doi: 10.1111/tbed.12901.
864 PubMed PMID: 30043436.

865 36. Tounkara K, Kwiatek O, Sidibe CAK, Sery A, Dakouo M, Salami H, et al. Persistence of the
866 historical lineage I of West Africa against the ongoing spread of the Asian lineage of peste des petits
867 ruminants virus. Transbound Emerg Dis. 2021. Epub 2021/03/12. doi: 10.1111/tbed.14066. PubMed
868 PMID: 33704888.
869 37. Baron MD, Diop B, Njeumi F, Willett BJ, Bailey D. Future research to underpin successful
870 peste des petits ruminants virus (PPRV) eradication. J Gen Virol. 2017;98(11):2635-44. Epub
871 2017/10/13. doi: 10.1099/jgv.0.000944. PubMed PMID: 29022862; PubMed Central PMCID:
872 PMCPMC5845661.
873 38. Muniraju M, Munir M, Parthiban AR, Banyard AC, Bao J, Wang Z, et al. Molecular evolution
874 of peste des petits ruminants virus. Emerg Infect Dis. 2014;20(12):2023-33. Epub 2014/11/25. doi:
875 10.3201/eid2012.140684. PubMed PMID: 25418782; PubMed Central PMCID: PMCPMC4257836.
876 39. Shabbir MZ, Rahman AU, Munir M. A comprehensive global perspective on phylogenomics
877 and evolutionary dynamics of Small ruminant morbillivirus. Sci Rep. 2020;10(1):17. Epub
878 2020/01/09. doi: 10.1038/s41598-019-54714-w. PubMed PMID: 31913305; PubMed Central PMCID:
879 PMCPMC6949297.
880 40. Shatar M, Khanui B, Purevtseren D, Khishgee B, Loitsch A, Unger H, et al. First genetic
881 characterization of peste des petits ruminants virus from Mongolia. Arch Virol. 2017;162(10):3157-
882 60. Epub 2017/07/02. doi: 10.1007/s00705-017-3456-4. PubMed PMID: 28667443.
883 41. Bao J, Wang Q, Li L, Liu C, Zhang Z, Li J, et al. Evolutionary dynamics of recent peste des
884 petits ruminants virus epidemic in China during 2013-2014. Virology. 2017;510:156-64. Epub
885 2017/07/25. doi: 10.1016/j.virol.2017.07.018. PubMed PMID: 28734191; PubMed Central PMCID:
886 PMCPMC7111700.
887 42. Clarke B, Mahapatra M, Friedgut O, Bumbarov V, Parida S. Persistence of Lineage IV Peste-
888 des-petits ruminants virus within Israel since 1993: An evolutionary perspective. PLoS One.
889 2017;12(5):e0177028. Epub 2017/05/26. doi: 10.1371/journal.pone.0177028. PubMed PMID:
890 28545149; PubMed Central PMCID: PMCPMC5436660.

43. Kock R. Food and Agriculture Organisation of the United Nations, Crisis Management Centre-Animal Health Mission Report: Investigation of Peste des Petits Ruminants (PPR) among wild animals and its potential impact on the current PPR situation in livestock. 2017.

44. Wernike K, Eschbaumer M, Breithaupt A, Maltzan J, Wiesner H, Beer M, et al. Experimental infection of sheep and goats with a recent isolate of peste des petits ruminants virus from Kurdistan. Vet Microbiol. 2014;172(1-2):140-5. Epub 2014/06/09. doi: 10.1016/j.vetmic.2014.05.010. PubMed PMID: 24908276.

45. Schulz C, Fast C, Schlottau K, Hoffmann B, Beer M. Neglected Hosts of Small Ruminant Morbillivirus. Emerg Infect Dis. 2018;24(12):2334-7. Epub 2018/11/21. doi: 10.3201/eid2412.180507. PubMed PMID: 30457523; PubMed Central PMCID: PMCPMC6256395.

46. Beaty SM, Lee B. Constraints on the Genetic and Antigenic Variability of Measles Virus. Viruses. 2016;8(4):109. Epub 2016/04/26. doi: 10.3390/v8040109. PubMed PMID: 27110809; PubMed Central PMCID: PMCPMC4848602.

47. Han GZ, Worobey M. Homologous recombination in negative sense RNA viruses. Viruses. 2011;3(8):1358-73. Epub 2011/10/14. doi: 10.3390/v3081358. PubMed PMID: 21994784; PubMed Central PMCID: PMCPMC3185808.

48. Shabbir MZ, Rahman AU, Munir M. Author Correction: A comprehensive global perspective on phylogenomics and evolutionary dynamics of Small ruminant morbillivirus. Sci Rep. 2020;10(1):20997. Epub 2020/11/28. doi: 10.1038/s41598-020-78219-z. PubMed PMID: 33244107; PubMed Central PMCID: PMCPMC7691326.

49. Liu F. Letter to the editor concerning "First report of peste des petits ruminants virus lineage II in Hydropotes inermis, China" by Zhou et al. (Transbound Emerg Dis; 2017: https://doi.org/10.1111/tbed.12683). Transbound Emerg Dis. 2018;65(4):1125. Epub 2018/02/25. doi: 10.1111/tbed.12840. PubMed PMID: 29476605.

915 50. Padhi A, Ma L. Genetic and epidemiological insights into the emergence of peste des petits
916 ruminants virus (PPRV) across Asia and Africa. Sci Rep. 2014;4:7040. Epub 2014/11/14. doi:
917 10.1038/srep07040. PubMed PMID: 25391314; PubMed Central PMCID: PMCPMC4229660.
918 51. Banyard AC, Parida S, Batten C, Oura C, Kwiatek O, Libeau G. Global distribution of peste
919 des petits ruminants virus and prospects for improved diagnosis and control. J Gen Virol. 2010;91(Pt
920 12):2885-97. Epub 2010/09/17. doi: 10.1099/vir.0.025841-0. PubMed PMID: 20844089.
921 52. Kwiatek O, Ali YH, Saeed IK, Khalafalla AI, Mohamed OI, Obeida AA, et al. Asian lineage of
922 peste des petits ruminants virus, Africa. Emerg Infect Dis. 2011;17(7):1223-31. Epub 2011/07/19. doi:
923 10.3201/eid1707.101216. PubMed PMID: 21762576; PubMed Central PMCID: PMCPMC3381390.
924 53. Rahman AU, Munir M, Shabbir MZ. A comparative phylogenomic analysis of peste des
925 petits ruminants virus isolated from wild and unusual hosts. Mol Biol Rep. 2019;46(5):5587-93. Epub
926 2019/07/19. doi: 10.1007/s11033-019-04973-7. PubMed PMID: 31317455.
927 54. Wu X, Li L, Li J, Liu C, Wang Q, Bao JY, et al. Peste des Petits Ruminants Viruses Re-
928 emerging in China, 2013-2014. Transbound Emerg Dis. 2016;63(5):e441-6. Epub 2015/01/27. doi:
929 10.1111/tbed.12308. PubMed PMID: 25620455.
930 55. Gao X, Wen-xuan X, Yang W, Blank D, Qiao J, Xu K. Status and distribution of ungulates in
931 Xinjiang, China. Journal of Arid Land. 2011;3:49-60.
932 56. Li XP, Zhai SL, He DS, Guo PJ, Lv DH, Wen XH, et al. Genome characterization and
933 phylogenetic analysis of a lineage IV peste des petits ruminants virus in southern China. Virus
934 Genes. 2015;51(3):361-6. Epub 2015/11/18. doi: 10.1007/s11262-015-1249-y. PubMed PMID:
935 26573282.
936 57. Calain P, Roux L. The rule of six, a basic feature for efficient replication of Sendai virus
937 defective interfering RNA. J Virol. 1993;67(8):4822-30. Epub 1993/08/01. doi: 10.1128/JVI.67.8.4822-
938 4830.1993. PubMed PMID: 8392616; PubMed Central PMCID: PMCPMC237869.
939 58. Ivancic-Jelecki J, Slovic A, Santak M, Tesovic G, Forcic D. Common position of indels that
940 cause deviations from canonical genome organization in different measles virus strains. Virol J.

2016;13:134. Epub 2016/07/31. doi: 10.1186/s12985-016-0587-2. PubMed PMID: 27473517; PubMed Central PMCID: PMCPMC4966754.

59. King S, Rajko-Nenow P, Ropiak HM, Ribeca P, Batten C, Baron MD. Full genome sequencing of archived wild type and vaccine rinderpest virus isolates prior to their destruction. Sci Rep. 2020;10(1):6563. Epub 2020/04/18. doi: 10.1038/s41598-020-63707-z. PubMed PMID: 32300201; PubMed Central PMCID: PMCPMC7162898.

60. Bankamp B, Liu C, Rivailler P, Bera J, Shrivastava S, Kirkness EF, et al. Wild-type measles viruses with non-standard genome lengths. PLoS One. 2014;9(4):e95470. Epub 2014/04/22. doi: 10.1371/journal.pone.0095470. PubMed PMID: 24748123; PubMed Central PMCID: PMCPMC3991672.

61. Gil H, Fernandez-Garcia A, Mosquera MM, Hubschen JM, Castellanos AM, de Ory F, et al. Measles virus genotype D4 strains with non-standard length M-F non-coding region circulated during the major outbreaks of 2011-2012 in Spain. PLoS One. 2018;13(7):e0199975. Epub 2018/07/17. doi: 10.1371/journal.pone.0199975. PubMed PMID: 30011283; PubMed Central PMCID: PMCPMC6047782.

62. Chulakasian S, Chang TJ, Tsai CH, Wong ML, Hsu WL. Translational enhancing activity in 5' UTR of peste des petits ruminants virus fusion gene. FEBS J. 2013;280(5):1237-48. Epub 2013/01/08. doi: 10.1111/febs.12115. PubMed PMID: 23289829.

63. Frost SDW, Magalis BR, Kosakovsky Pond SL. Neutral Theory and Rapidly Evolving Viral Pathogens. Mol Biol Evol. 2018;35(6):1348-54. Epub 2018/04/25. doi: 10.1093/molbev/msy088. PubMed PMID: 29688481; PubMed Central PMCID: PMCPMC6279309.

64. Spielman SJ, Weaver S, Shank SD, Magalis BR, Li M, Kosakovsky Pond SL. Evolution of Viral Genomes: Interplay Between Selection, Recombination, and Other Forces. Methods Mol Biol. 2019;1910:427-68. Epub 2019/07/07. doi: 10.1007/978-1-4939-9074-0_14. PubMed PMID: 31278673.

65. Agrawal A, Gupta R, Gattani A, Kumar Patel S, Hira Khan M, Singh P. Novel T Cell Epitope Designing from PPRV HN Protein for Peptide based Subunit Vaccine: An Immune Informatics

967 Approach. IntJCurrMicrobiolAppSci. 2020;9(3):2432-9. doi:
968 https://doi.org/10.20546/ijcmas.2020.903.278.
969 66. Liang Z, Yuan R, Chen L, Zhu X, Dou Y. Molecular Evolution and Characterization of
970 Hemagglutinin (H) in Peste des Petits Ruminants Virus. PLoS One. 2016;11(4):e0152587. Epub
971 2016/04/02. doi: 10.1371/journal.pone.0152587. PubMed PMID: 27035347; PubMed Central PMCID:
972 PMCPMC4818033.
973 67. Chinnakannan SK, Nanda SK, Baron MD. Morbillivirus v proteins exhibit multiple
974 mechanisms to block type 1 and type 2 interferon signalling pathways. PLoS One. 2013;8(2):e57063.
975 Epub 2013/02/23. doi: 10.1371/journal.pone.0057063. PubMed PMID: 23431397; PubMed Central
976 PMCID: PMCPMC3576338.
977 68. Li L, Shi X, Zhang D, Cao X, Ali A, Bai J. Peste des petits ruminants virus V protein inhibits
978 cellular antiviral response by blocking the secretion of endogenous and exogenous interferons. J
979 Vet Med Sci. 2020. Epub 2020/11/17. doi: 10.1292/jvms.20-0487. PubMed PMID: 33191335.
980 69. Li P, Zhu Z, Zhang X, Dang W, Li L, Du X, et al. The Nucleoprotein and Phosphoprotein of
981 Peste des Petits Ruminants Virus Inhibit Interferons Signaling by Blocking the JAK-STAT Pathway.
982 Viruses. 2019;11(7). Epub 2019/07/11. doi: 10.3390/v11070629. PubMed PMID: 31288481; PubMed
983 Central PMCID: PMCPMC6669484.
984 70. Zhu Z, Li P, Yang F, Cao W, Zhang X, Dang W, et al. Peste des Petits Ruminants Virus
985 Nucleocapsid Protein Inhibits Beta Interferon Production by Interacting with IRF3 To Block Its
986 Activation. J Virol. 2019;93(16). Epub 2019/06/07. doi: 10.1128/JVI.00362-19. PubMed PMID:
987 31167907; PubMed Central PMCID: PMCPMC6675899.
988 71. Crispell J, Benton CH, Balaz D, De Maio N, Ahkmetova A, Allen A, et al. Combining
989 genomics and epidemiology to analyse bi-directional transmission of Mycobacterium bovis in a
990 multi-host system. Elife. 2019;8. Epub 2019/12/18. doi: 10.7554/eLife.45833. PubMed PMID:
991 31843054; PubMed Central PMCID: PMCPMC6917503.

992 72. Viana M, Shirima GM, John KS, Fitzpatrick J, Kazwala RR, Buza JJ, et al. Integrating
993 serological and genetic data to quantify cross-species transmission: brucellosis as a case study.
994 Parasitology. 2016;143(7):821-34. Epub 2016/03/05. doi: 10.1017/S0031182016000044. PubMed
995 PMID: 26935267; PubMed Central PMCID: PMCPMC4873909.
996 73. Bataille A, Salami H, Seck I, Lo MM, Ba A, Diop M, et al. Combining viral genetic and animal
997 mobility network data to unravel peste des petits ruminants transmission dynamics in West Africa.
998 PLoS Pathog. 2021;17(3):e1009397. Epub 2021/03/19. doi: 10.1371/journal.ppat.1009397. PubMed
999 PMID: 33735294; PubMed Central PMCID: PMCPMC8009415 Mohamed Keyra, Abdellahi Salem
1000 Lella were unable to confirm their authorship contributions. On their behalf, the corresponding
1001 author has reported their contributions to the best of their knowledge.
1002 74. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform.
1003 Bioinformatics. 2010;26(5):589-95. Epub 2010/01/19. doi: 10.1093/bioinformatics/btp698. PubMed
1004 PMID: 20080505; PubMed Central PMCID: PMCPMC2828108.
1005 75. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
1006 Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. Epub 2009/06/10. doi:
1007 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PubMed Central PMCID:
1008 PMCPMC2723002.
1009 76. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new
1010 genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol.
1011 2012;19(5):455-77. Epub 2012/04/18. doi: 10.1089/cmb.2012.0021. PubMed PMID: 22506599;
1012 PubMed Central PMCID: PMCPMC3342519.
1013 77. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence
1014 alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059-66. Epub
1015 2002/07/24. doi: 10.1093/nar/gkf436. PubMed PMID: 12136088; PubMed Central PMCID:
1016 PMCPMC135756.

78. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016;2(1):vew007. Epub 2016/10/25. doi: 10.1093/ve/vew007. PubMed PMID: 27774300; PubMed Central PMCID: PMCPMC4989882.

79. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 2015;1(1):vev003. Epub 2015/05/26. doi: 10.1093/ve/vev003. PubMed PMID: 27774277; PubMed Central PMCID: PMCPMC5014473.

80. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870-4. Epub 2016/03/24. doi: 10.1093/molbev/msw054. PubMed PMID: 27004904.

81. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018;4(1):vey016. Epub 2018/06/27. doi: 10.1093/ve/vey016. PubMed PMID: 29942656; PubMed Central PMCID: PMCPMC6007674.

82. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol. 2018;67(5):901-4. Epub 2018/05/03. doi: 10.1093/sysbio/syy032. PubMed PMID: 29718447; PubMed Central PMCID: PMCPMC6101584.

83. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006;4(5):e88. Epub 2006/05/11. doi: 10.1371/journal.pbio.0040088. PubMed PMID: 16683862; PubMed Central PMCID: PMCPMC1395354.

84. Argimon S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom. 2016;2(11):e000093. Epub 2017/03/30. doi: 10.1099/mgen.0.000093. PubMed PMID: 28348833; PubMed Central PMCID: PMCPMC5320705.

85. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. Mol Biol Evol.

2016;33(8):2167-9. Epub 2016/05/18. doi: 10.1093/molbev/msw082. PubMed PMID: 27189542; PubMed Central PMCID: PMCPMC6398721.

86. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol. 2013;30(5):1196-205. Epub 2013/02/20. doi: 10.1093/molbev/mst030. PubMed PMID: 23420840; PubMed Central PMCID: PMCPMC3670733.

87. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 2005;22(5):1208-22. Epub 2005/02/11. doi: 10.1093/molbev/msi105. PubMed PMID: 15703242.

88. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. Mol Biol Evol. 2018;35(3):773-7. Epub 2018/01/05. doi: 10.1093/molbev/msx335. PubMed PMID: 29301006; PubMed Central PMCID: PMCPMC5850112.

89. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 2012;8(7):e1002764. Epub 2012/07/19. doi: 10.1371/journal.pgen.1002764. PubMed PMID: 22807683; PubMed Central PMCID: PMCPMC3395634.

90. Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. EasyCodeML: A visual tool for analysis of selection using CodeML. Ecol Evol. 2019;9(7):3891-8. Epub 2019/04/25. doi: 10.1002/ece3.5015. PubMed PMID: 31015974; PubMed Central PMCID: PMCPMC6467853.

91. Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics. 1998;148(3):929-36. Epub 1998/04/16. PubMed PMID: 9539414; PubMed Central PMCID: PMCPMC1460041.

92. Yang Z, Nielsen R, Goldman N, Pedersen AM. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics. 2000;155(1):431-49. Epub 2000/05/03. PubMed PMID: 10790415; PubMed Central PMCID: PMCPMC1461088.

93. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005;22(4):1107-18. Epub 2005/02/04. doi: 10.1093/molbev/msi097. PubMed PMID: 15689528.

94. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol. 2015;32(5):1342-53. Epub 2015/02/24. doi: 10.1093/molbev/msv022. PubMed PMID: 25697341; PubMed Central PMCID: PMCPMC4408413.

95. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, et al. Gene-wide identification of episodic selection. Mol Biol Evol. 2015;32(5):1365-71. Epub 2015/02/24. doi: 10.1093/molbev/msv035. PubMed PMID: 25701167; PubMed Central PMCID: PMCPMC4408417.

96. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res. 2003;31(13):3381-5. Epub 2003/06/26. doi: 10.1093/nar/gkg520. PubMed PMID: 12824332; PubMed Central PMCID: PMCPMC168927.

Supporting information

S1 Fig. Molecular detection of PPRV N gene in different tissues from wild Mongolian ungulates. RT-PCR for PPRV N gene was performed on 2ul RNA extracted from different tissue samples from the indicated host (RNA concentration ranged from 26-1125 ng/ul) prior to gel electrophoresis and UV transillumination. Lane numbers refer to sample IDs for different tissues given in Table 1. DNA ladder markers of different base pair (bp) lengths are shown. 'x' denotes an empty lane; '-' denotes a no template control and '+' denotes a positive control RNA template. Samples used subsequently for whole genome NGS are marked with *.

S2 Fig. Untraited Bayesian time-scaled Maximum Clade Credibility Tree. MCC tree from the combined output of three MCMC chains run in BEAST v1.10.4, inferred without partitioning the data by traits, and visualized in FigTree. The novel genomes from this study are shown in blue. x-

axis shows date. Lineages, referred to as LI, LII, LIII or LIV, are marked. * indicates posterior probability > 0.9 at the node opposite.

S3 Fig. TempEst analysis of temporal signal in the PPRV genome dataset. Plots from TempEst showing root-to-tip genetic distance against sampling date (A) and residuals (B) for a ML phylogeny of 85 PPRV genomes, using the best-fitting root. The correlation coefficient for the regression was 0.9362 and R^2 was 0.8764. An outlier sequence, shown in blue, was removed from the alignment before phylogenetic analysis.

S1 Table. Illumina NGS read summary for wildlife samples. Total read number, PPRV-specific read number and % total reads which were PPRV are shown for the five novel PPRV genomes from wildlife hosts. Average genome coverage was calculated as (read count * read length) / total genome size.

S2 Table. Recombination analysis using RPD4. The PPRV genome alignment (n=84) was analysed using Recombination Detection Program v4.101 (RPD4) using default settings. Genomes unambiguously identified as recombinant sequences are shown, for which at least 5 of 7 of the detection methods found significant evidence of recombination. Recombinant: genome sequence identified as potential recombinant; Lin-r: PPRV genetic lineage of Recombinant; Minor parental sequence: genome sequence identified as most likely minor parent of the recombinant, i.e. most closely related to the genome portion inserted; * indicates that other potential minor parents were also identified by RPD4 for that recombination event; Lin-mp: PPRV genetic lineage of minor parent; Begin: average genome position (in alignment) of the beginning breakpoint point of recombination; End: average genome position (in alignment) of the end breakpoint point of recombination; NS: not significant. The p-values for analysis with each of the 7 different

recombination detection algorithms are shown, after Bonferroni correction for multiple comparisons.

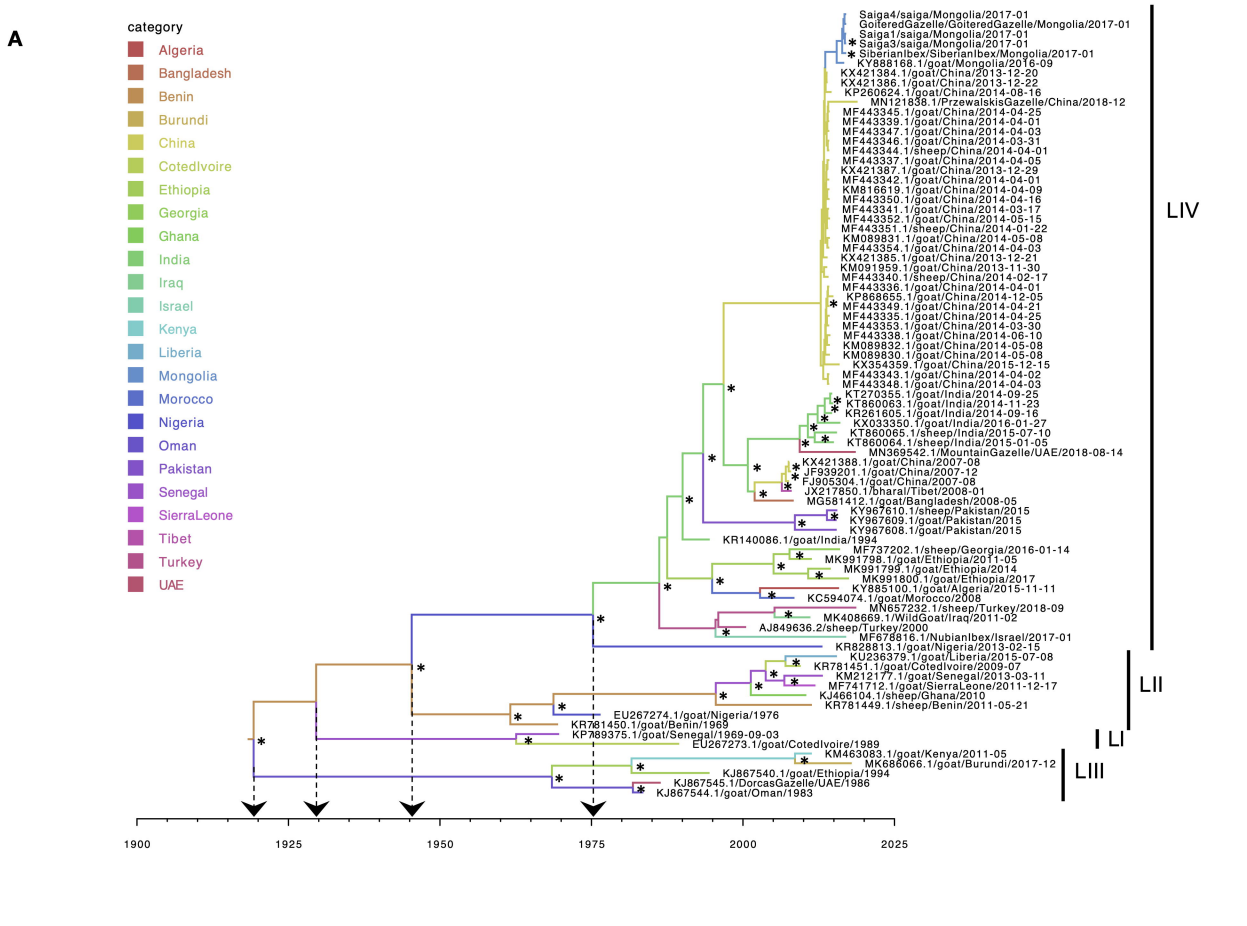
S3 Table. Bayes factors for spread of PPRV between countries. Using SpreaD3, Bayes Factors (BFs) were calculated using the log file from the BEAST BSSVS analysis and a discrete set of longitude and latitude coordinates for each country, coupled to a geoJSON formatted world map. The output gave BFs for all possible transitions between locations. The table shows the transitions with posterior probabilities >0.7. The transition from China to Mongolia, which had the highest BF of any transition, is highlighted in grey.

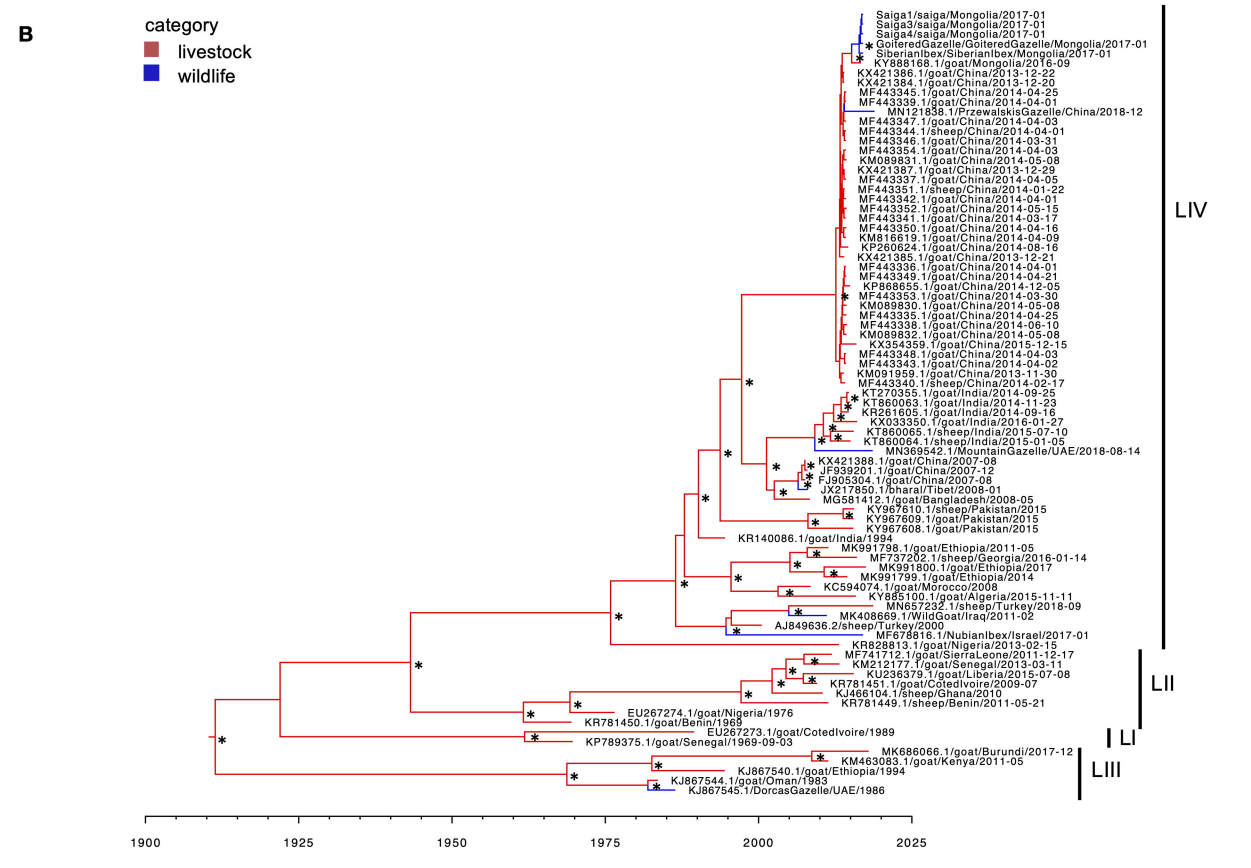
S4 Table. Values of Log-likelihood (lnL) for PPRV genes using different selection models in the CodeML analysis, and LRT comparing the two models. Two different site selection models (in which the ω ratio varies among codons) were implemented in CodeML: M7 (beta; no positive selection) and M8 (beta&ω; positive selection). For each gene, Log-likelihood (lnL) values are shown and the likelihood ratio test (LRT) to show the significance of model comparison. Bayes empirical Bayes (BEB) were used to calculate the posterior probabilities for site classes and identify sites under positive selection.

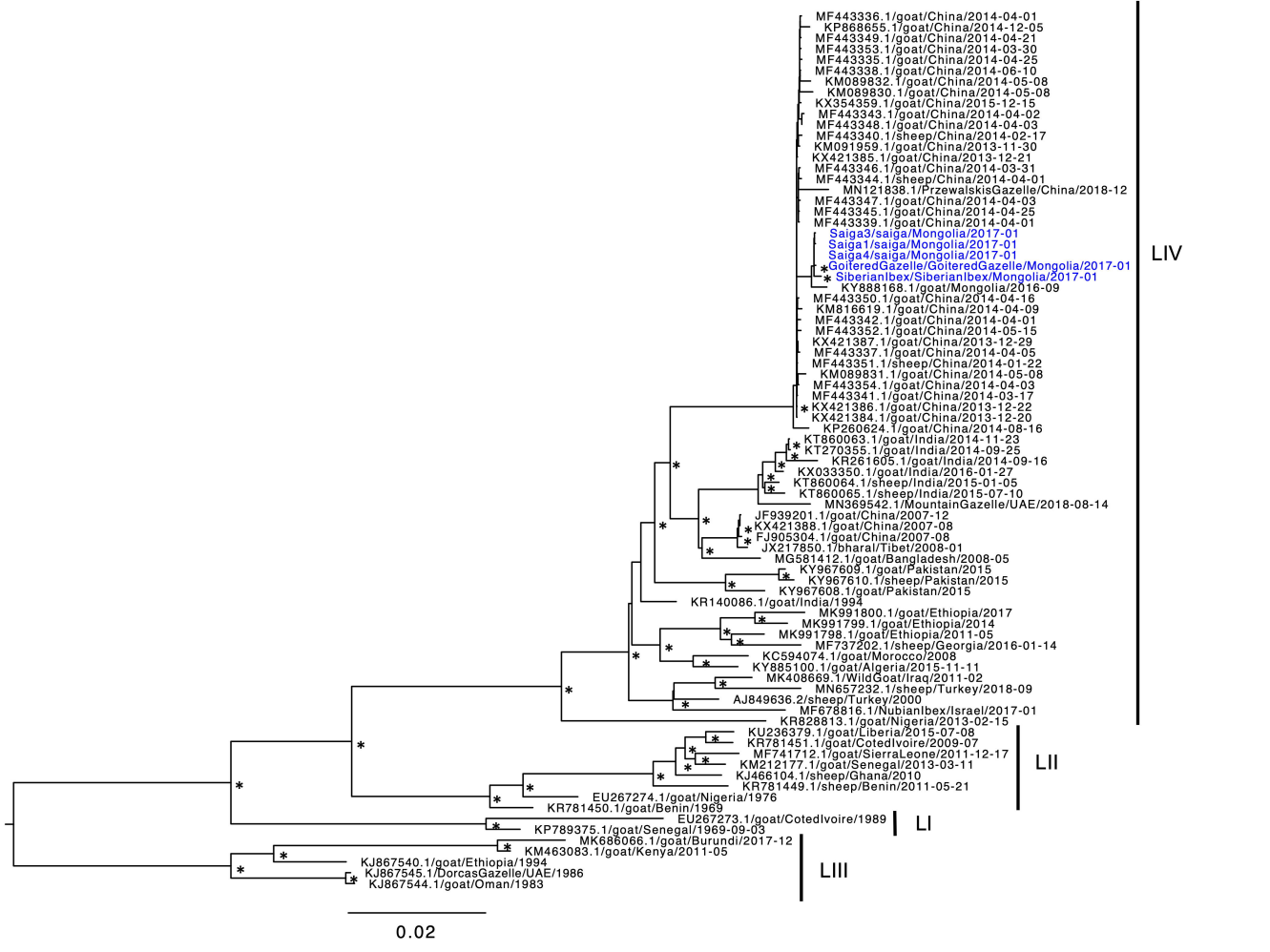
S5 Table: Evidence from aBSREL for episodic diversifying selection acting on the PPRV L gene. The PPRV L gene was analysed using aBSREL with the China/Mongolia clade selected the set of foreground branches on which to test for episodic diversifying selection. Significance was assessed using the likelihood ratio test statistic for selection (LRT) at a threshold of $p \leq 0.05$, after correcting for multiple testing. The two branches shown were identified as under positive selection, while all other branches were best described by a single ω rate category (ω1). The ω distribution shows inferred estimates for ω1 and ω2 and proportion of sites in each category.

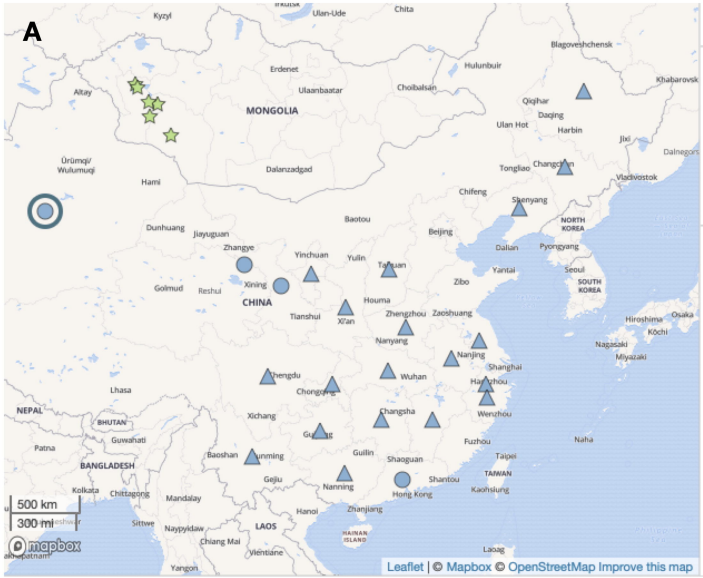
1146 **S1 File: Alignment of PPRV genomes (n=81)**

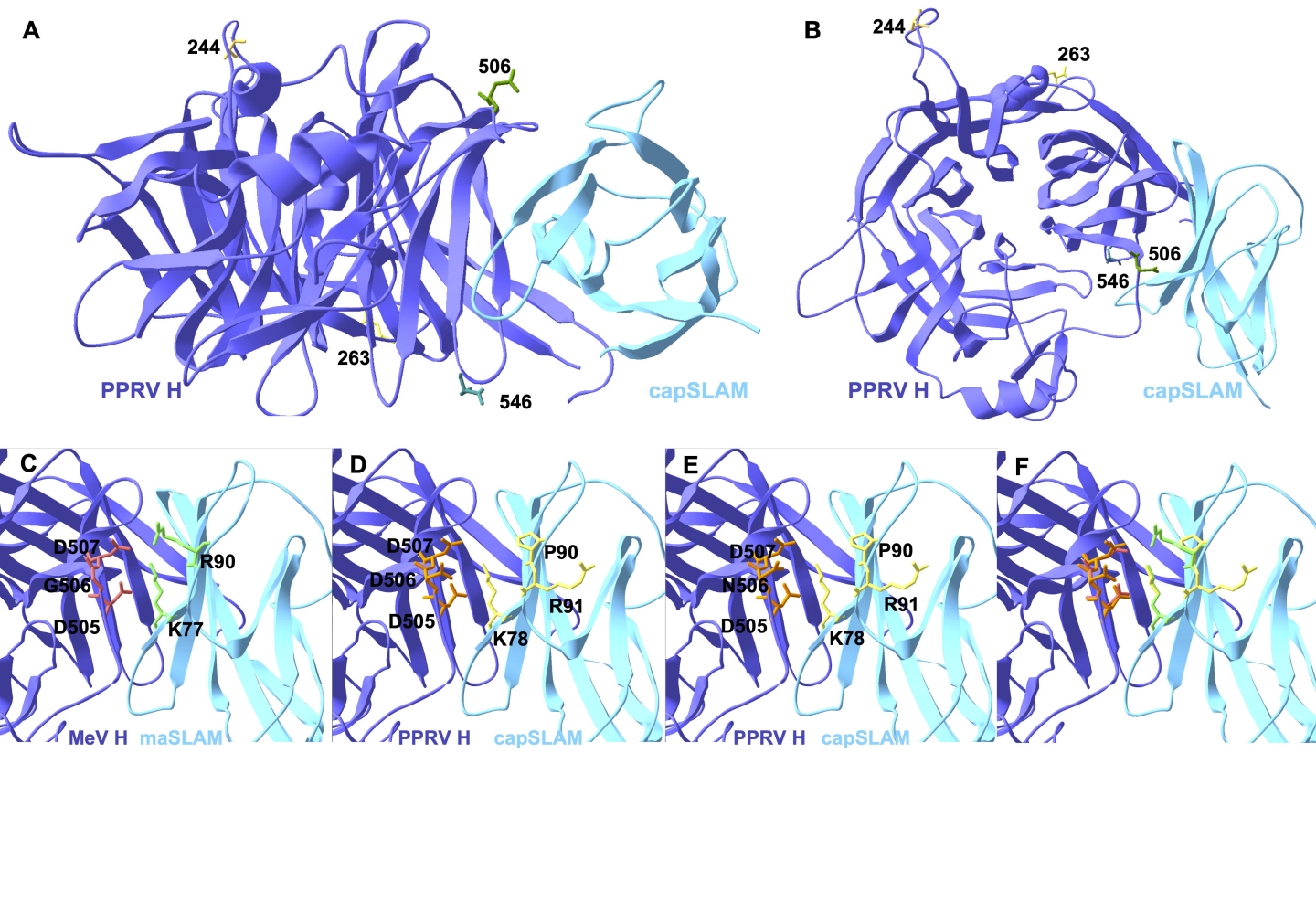
1147


A
category
Algeria
Bangladesh
Benin
Burundi
China
CotedIvoire
Ethiopia
Georgia
Ghana
India
Iraq
Israel
Kenya
Liberia
Mongolia
Morocco
Nigeria
Oman
Pakistan
Senegal
SierraLeone
Tibet
Turkey
UAE
Saiga4/saiga/Mongolia/2017-01
GolferedGazelle/GolferedGazelle/Mongolia/2017-01
Saiga1/saiga/Mongolia/2017-01
Saiga3/saiga/Mongolia/2017-01
SiberianIbex/SiberianIbex/Mongolia/2017-01
KY888168.1/goat/Mongolia/2016-09
KX421384.1/goat/China/2013-12-20
KX421386.1/goat/China/2013-12-22
KP260624.1/goat/China/2014-08-16
MN121838.1/PrzewalskisGazelle/China/2018-12
MF443345.1/goat/China/2014-04-25
MF443339.1/goat/China/2014-04-01
MF443347.1/goat/China/2014-04-03
MF443346.1/goat/China/2014-03-31
MF443344.1/sheep/China/2014-04-01
MF443337.1/goat/China/2014-04-05
KX421387.1/goat/China/2013-12-29
MF443342.1/goat/China/2014-04-01
KM816619.1/goat/China/2014-04-09
MF443350.1/goat/China/2014-04-16
MF443341.1/goat/China/2014-03-17
MF443352.1/goat/China/2014-05-15
MF443351.1/sheep/China/2014-01-22
KM089831.1/goat/China/2014-05-08
MF443354.1/goat/China/2014-04-03
KX421385.1/goat/China/2013-12-21
KM091959.1/goat/China/2013-11-30
MF443340.1/sheep/China/2014-02-17
MF443336.1/goat/China/2014-04-01
KP868655.1/goat/China/2014-12-05
MF443349.1/goat/China/2014-04-21
MF443335.1/goat/China/2014-04-25
MF443353.1/goat/China/2014-03-30
MF443338.1/goat/China/2014-06-10
KM089832.1/goat/China/2014-05-08
KM089830.1/goat/China/2014-05-08
KX354359.1/goat/China/2015-12-15
MF443343.1/goat/China/2014-04-02
MF443348.1/goat/China/2014-04-03
KT270355.1/goat/India/2014-09-25
KT860063.1/goat/India/2014-11-23
KR261605.1/goat/India/2014-09-16
KX033350.1/goat/India/2016-01-27
KT860065.1/sheep/India/2015-07-10
KT860064.1/sheep/India/2015-01-05
MN369542.1/MountainGazelle/UAE/2018-08-14
KX421388.1/goat/China/2007-08
JF939201.1/goat/China/2007-12
FJ905304.1/goat/China/2007-08
JX217850.1/bharal/Tibet/2008-01
MG581412.1/goat/Bangladesh/2008-05
KY967610.1/sheep/Pakistan/2015
KY967609.1/goat/Pakistan/2015
KY967608.1/goat/Pakistan/2015
KR140086.1/goat/India/1994
MF737202.1/sheep/Georgia/2016-01-14
MK991798.1/goat/Ethiopia/2011-05
MK991799.1/goat/Ethiopia/2014
MK991800.1/goat/Ethiopia/2017
KY885100.1/goat/Algeria/2015-11-11
KC594074.1/goat/Morocco/2008
MN657232.1/sheep/Turkey/2018-09
MK408669.1/WildGoat/Iraq/2011-02
AJ849636.2/sheep/Turkey/2000
MF678816.1/NubianIbex/Israel/2017-01
KR828813.1/goat/Nigeria/2013-02-15
EU023371.1/goat/Liberia/2015-07-08
KR781451.1/goat/CotedIvoire/2009-07
KM212177.1/goat/Senegal/2013-03-11
MF741712.1/goat/SierraLeone/2011-12-17
KJ486104.1/sheep/Ghana/2010
KR781449.1/sheep/Benin/2011-05-21
EU267274.1/goat/Nigeria/1976
KR781450.1/goat/Benin/1969
KP789375.1/goat/Senegal/1969-09-03
EU267273.1/goat/CotedIvoire/1989
KM463083.1/goat/Kenya/2011-05
MK686066.1/goat/Burundi/2017-12
KJ867540.1/goat/Ethiopia/1994
KJ867545.1/DorcasGazelle/UAE/1986
KJ867544.1/goat/Oman/1983
LIV
LII
LI
LIII
1900
1925
1950
1975
2000
2025


B
category
livestock
wildlife
Saiga1/saiga/Mongolia/2017-01
Saiga3/saiga/Mongolia/2017-01
Saiga4/saiga/Mongolia/2017-01
GoiteredGazelle/GoiteredGazelle/Mongolia/2017-01
SiberianIbex/SiberianIbex/Mongolia/2017-01
KY888168.1/goat/Mongolia/2016-09
KX421386.1/goat/China/2013-12-22
KX421384.1/goat/China/2013-12-20
MF443345.1/goat/China/2014-04-25
MF443339.1/goat/China/2014-04-01
MN121838.1/PrzewalskisGazelle/China/2018-12
MF443347.1/goat/China/2014-04-03
MF443344.1/sheep/China/2014-04-01
MF443346.1/goat/China/2014-03-31
MF443354.1/goat/China/2014-04-03
KM089831.1/goat/China/2014-05-08
KX421387.1/goat/China/2013-12-29
MF443337.1/goat/China/2014-04-05
MF443351.1/sheep/China/2014-01-22
MF443342.1/goat/China/2014-04-01
MF443352.1/goat/China/2014-05-15
MF443341.1/goat/China/2014-03-17
MF443350.1/goat/China/2014-04-16
KM816619.1/goat/China/2014-04-09
KP260624.1/goat/China/2014-08-16
KX421385.1/goat/China/2013-12-21
MF443336.1/goat/China/2014-04-01
MF443349.1/goat/China/2014-04-21
KP868655.1/goat/China/2014-12-05
MF443353.1/goat/China/2014-03-30
KM089830.1/goat/China/2014-05-08
MF443335.1/goat/China/2014-04-25
MF443338.1/goat/China/2014-06-10
KM089832.1/goat/China/2014-05-08
KX354359.1/goat/China/2015-12-15
MF443348.1/goat/China/2014-04-03
MF443343.1/goat/China/2014-04-02
KM091959.1/goat/China/2013-11-30
MF443340.1/sheep/China/2014-02-17
KT270355.1/goat/India/2014-09-25
KT860063.1/goat/India/2014-11-23
KR261605.1/goat/India/2014-09-16
KX033350.1/goat/India/2016-01-27
KT860065.1/sheep/India/2015-07-10
KT860064.1/sheep/India/2015-01-05
MN369542.1/MountainGazelle/UAE/2018-08-14
KX421388.1/goat/China/2007-08
JF939201.1/goat/China/2007-12
FJ905304.1/goat/China/2007-08
JX217850.1/bharal/Tibet/2008-01
MG581412.1/goat/Bangladesh/2008-05
KY967610.1/sheep/Pakistan/2015
KY967609.1/goat/Pakistan/2015
KY967608.1/goat/Pakistan/2015
KR140086.1/goat/India/1994
MK991798.1/goat/Ethiopia/2011-05
MF737202.1/sheep/Georgia/2016-01-14
MK991800.1/goat/Ethiopia/2017
MK991799.1/goat/Ethiopia/2014
KC594074.1/goat/Morocco/2008
KY885100.1/goat/Algeria/2015-11-11
MN657232.1/sheep/Turkey/2018-09
MK406669.1/WildGoat/Iraq/2011-02
AJ849636.2/sheep/Turkey/2000
MF678816.1/NubianIbex/Israel/2017-01
KR828813.1/goat/Nigeria/2013-02-15
MF74112.1/goat/SierraLeone/2011-12-17
KM212177.1/goat/Senegal/2013-03-11
KU236379.1/goat/Liberia/2015-07-08
KR781451.1/goat/CotedIvoire/2009-07
KJ466104.1/sheep/Ghana/2010
KR781449.1/sheep/Benin/2011-05-21
EU267274.1/goat/Nigeria/1976
KR781450.1/goat/Benin/1969
EU267273.1/goat/CotedIvoire/1989
KP789375.1/goat/Senegal/1969-09-03
MK686066.1/goat/Burundi/2017-12
KM463083.1/goat/Kenya/2011-05
KJ867540.1/goat/Ethiopia/1994
KJ867544.1/goat/Oman/1983
KJ867545.1/DorcasGazelle/UAE/1986
LIV
LII
LI
LIII
1900
1925
1950
1975
2000
2025


MF443336.1/goat/China/2014-04-01
KP868655.1/goat/China/2014-12-05
MF443349.1/goat/China/2014-04-21
MF443353.1/goat/China/2014-03-30
MF443335.1/goat/China/2014-04-25
MF443338.1/goat/China/2014-06-10
KM089832.1/goat/China/2014-05-08
KM089830.1/goat/China/2014-05-08
KX354359.1/goat/China/2015-12-15
MF443343.1/goat/China/2014-04-02
MF443348.1/goat/China/2014-04-03
MF443340.1/sheep/China/2014-02-17
KM091959.1/goat/China/2013-11-30
KX421385.1/goat/China/2013-12-21
MF443346.1/goat/China/2014-03-31
MF443344.1/sheep/China/2014-04-01
MN121838.1/PrzewalskisGazelle/China/2018-12
MF443347.1/goat/China/2014-04-03
MF443345.1/goat/China/2014-04-25
MF443339.1/goat/China/2014-04-01
Saiga3/saiga/Mongolia/2017-01
Saiga1/saiga/Mongolia/2017-01
Saiga4/saiga/Mongolia/2017-01
GoiteredGazelle/GoiteredGazelle/Mongolia/2017-01
SiberianIbex/SiberianIbex/Mongolia/2017-01
KY888168.1/goat/Mongolia/2016-09
MF443350.1/goat/China/2014-04-16
KM816619.1/goat/China/2014-04-09
MF443342.1/goat/China/2014-04-01
MF443352.1/goat/China/2014-05-15
KX421387.1/goat/China/2013-12-29
MF443337.1/goat/China/2014-04-05
MF443351.1/sheep/China/2014-01-22
KM089831.1/goat/China/2014-05-08
MF443354.1/goat/China/2014-04-03
MF443341.1/goat/China/2014-03-17
KX421386.1/goat/China/2013-12-22
KX421384.1/goat/China/2013-12-20
KP260624.1/goat/China/2014-08-16
LIV
KT860063.1/goat/India/2014-11-23
KT270355.1/goat/India/2014-09-25
KR261605.1/goat/India/2014-09-16
KX033350.1/goat/India/2016-01-27
KT860064.1/sheep/India/2015-01-05
KT860065.1/sheep/India/2015-07-10
MN369542.1/MountainGazelle/UAE/2018-08-14
JF939201.1/goat/China/2007-12
KX421388.1/goat/China/2007-08
FJ905304.1/goat/China/2007-08
JX217850.1/bharal/Tibet/2008-01
MG581412.1/goat/Bangladesh/2008-05
KY967609.1/goat/Pakistan/2015
KY967610.1/sheep/Pakistan/2015
KY967608.1/goat/Pakistan/2015
KR140086.1/goat/India/1998
MK991800.1/goat/Ethiopia/2017
MK991799.1/goat/Ethiopia/2014
MK991798.1/goat/Ethiopia/2011-05
MF737202.1/sheep/Georgia/2016-01-14
KC594074.1/goat/Morocco/2008
KY885100.1/goat/Algeria/2015-11-11
MK408669.1/WildGoat/Iraq/2011-02
MN657232.1/sheep/Turkey/2018-09
AJ849636.2/sheep/Turkey/2000
MF678816.1/NubianIbex/Israel/2017-01
KR828813.1/goat/Nigeria/2013-02-15
KU236379.1/goat/Liberia/2015-07-08
KR781451.1/goat/CotedIvoire/2009-07
MF741712.1/goat/SierraLeone/2011-12-17
KM212177.1/goat/Senegal/2013-03-11
KJ466104.1/sheep/Ghana/2010
KR781449.1/sheep/Benin/2011-05-21
LII
EU267274.1/goat/Nigeria/1976
KR781450.1/goat/Benin/1969
EU267273.1/goat/CotedIvoire/1989
KP789375.1/goat/Senegal/1969-09-03
LI
MK686066.1/goat/Burundi/2017-12
KM463083.1/goat/Kenya/2011-05
KJ867540.1/goat/Ethiopia/1994
KJ867545.1/DorcasGazelle/UAE/1986
KJ867544.1/goat/Oman/1983
LIII
0.02


A
MONGOLIA
CHINA
NEPAL
BHUTAN
BANGLADESH
LAOS
TAIWAN
NORTH KOREA
SOUTH KOREA
Kyzyl
Ulan-Ude
Chita
Erdenet
Hulunbuir
Blagoveshchensk
Khabarovsk
Altay
Ulaanbaatar
Choibalsan
Qiqihar
Daqing
Ulan Hot
Harbin
Jixi
Ürümqi/ Wulumuqi
Hami
Dalanzadgad
Tongliao
Changchun
Vladivostok
Chifeng
Shenyang
Dunhuang
Jiayuguan
Baotou
Beijing
Zhangye
Yinchuan
Yulin
Taiyuan
Dalian
Pyongyang
Seoul
Golmud
Reshui
Xining
Houma
Zibo
Yantai
Tianshui
Xi'an
Zhengzhou
Zaoshuang
Hiroshima
Osaka
Kochi
Nagasaki
Miyazaki
Nanyang
Nanjing
Shanghai
Chengdu
Wuhan
Hangzhou
Lhasa
Chongqing
Wenzhou
Xichang
Changsha
Fuzhou
Naha
Patna
Guwahati
Baoshan
Kunming
Guilin
Shaoguan
Taipei
Kolkata
Chittagong
Gejiu
Nanning
Shantou
Hong Kong
Kaohsiung
Mandalay
Hanoi
Zhanjiang
Sittwe
Naypyidaw
Chiang Mai
Vientiane
Yangon
Laoag
500 km
300 mi
mapbox
Leaflet | © Mapbox © OpenStreetMap Improve this map
B
KX421388.1/goat/China/2007-08
JF939201.1/goat/China/2007-12
MN369542.1/MountainGazelle/UAE/2018-08-14
KT860065.1/sheep/India/2015-07-10
KT860064.1/sheep/India/2015-01-05
KX033350.1/goat/India/2016-01-27
KR261605.1/goat/India/2014-09-16
KT270355.1/goat/India/2014-09-25
KT860063.1/goat/India/2014-11-23
MF443340.1/sheep/China/2014-02-17
KM091959.1/goat/China/2013-11-30
MF443343.1/goat/China/2014-04-02
MF443348.1/goat/China/2014-04-03
KX354359.1/goat/China/2015-12-15
KM089830.1/goat/China/2014-05-08
KM089832.1/goat/China/2014-05-08
MF443338.1/goat/China/2014-06-10
MF443335.1/goat/China/2014-04-25
MF443353.1/goat/China/2014-03-30
MF443349.1/goat/China/2014-04-21
MF443336.1/goat/China/2014-04-01
KP868655.1/goat/China/2014-12-05
KX421385.1/goat/China/2013-12-21
MF443342.1/goat/China/2014-04-01
KM089831.1/goat/China/2014-05-08
MF443354.1/goat/China/2014-04-03
MF443350.1/goat/China/2014-04-16
KM816619.1/goat/China/2014-04-09
KX421387.1/goat/China/2013-12-29
MF443337.1/goat/China/2014-04-05
MF443352.1/goat/China/2014-05-15
MF443351.1/sheep/China/2014-01-22
MF443341.1/goat/China/2014-03-17
MF443346.1/goat/China/2014-03-31
MF443344.1/sheep/China/2014-04-01
MF443345.1/goat/China/2014-04-25
MN121838.1/PrzewalskisGazelle/China/2018-12
MF443339.1/goat/China/2014-04-01
MF443347.1/goat/China/2014-04-03
KP260624.1/goat/China/2014-08-16
KX421386.1/goat/China/2013-12-22
KX421384.1/goat/China/2013-12-20
KY888168.1/goat/Mongolia/2016-09
SiberianIbex/SiberianIbex/Mongolia/2017-01
Saiga4/saiga/Mongolia/2017-01
GoiteredGazelle/GoiteredGazelle/Mongolia/2017-01
Saiga3/saiga/Mongolia/2017-01
Saiga1/saiga/Mongolia/2017-01
5.62
C
Feb 17, 2013
Sep 1, 2013
Mar 16, 2014
Sep 28, 2014
Apr 12, 2015
Oct 25, 2015
May 8, 2016
Nov 20, 2016
Jun 4, 2017
Dec 17, 2017
Dec 16, 2018


A
244
506
263
546
PPRV H
capSLAM
B
244
263
506
546
PPRV H
capSLAM
C
D507
G506
D505
R90
K77
MeV H
maSLAM
D
D507
D506
D505
P90
R91
K78
PPRV H
capSLAM
E
D507
N506
D505
P90
R91
K78
PPRV H
capSLAM
F