VIRUS EVOLUTION

Virus Evolution, 2024, **10(1)**, vead087
DOI: https://doi.org/10.1093/ve/vead087
Advance Access Publication 12 January 2024
Research Article

Equine herpesvirus 4 infected domestic horses associated with Sintashta spoke-wheeled chariots around 4,000 years ago

Ophélie Lebrasseur,[1,2,*,§] Kuldeep Dilip More,[1,*] and Ludovic Orlando[1,‡]

[1]Centre for Anthropobiology and Genomics of Toulouse (CAGT), CNRS/Université Paul Sabatier, 37 Allées Jules Guesde, 31000, Toulouse, France and [2]Instituto Nacional de Antropología y Pensamiento Latinoamericano, 3 de Febrero 1370 (1426), Ciudad Autónoma de Buenos Aires, Argentina
[§]https://orcid.org/0000-0003-0687-8538
[‡]https://orcid.org/0000-0003-3936-1850
*Corresponding author: E-mails: ophelie.lebrasseur@arch.ox.ac.uk; kdmore@buffalo.edu

Abstract

Equine viral outbreaks have disrupted the socio-economic life of past human societies up until the late 19th century and continue to be of major concern to the horse industry today. With a seroprevalence of 60–80 per cent, equine herpesvirus 4 (EHV-4) is the most common horse pathogen on the planet. Yet, its evolutionary history remains understudied. Here, we screen the sequenced data of 264 archaeological horse remains to detect the presence of EHV-4. We recover the first ancient EHV-4 genome with 4.2× average depth-of-coverage from a specimen excavated in the Southeastern Urals and dated to the Early Bronze Age period, approximately 3,900 years ago. The recovery of an EHV-4 virus outside the upper respiratory tract not only points to an animal particularly infected but also highlights the importance of post-cranial bones in pathogen characterisation. Bayesian phylogenetic reconstruction provides a minimal time estimate for EHV-4 diversification to around 4,000 years ago, a time when modern domestic horses spread across the Central Asian steppes together with spoke-wheeled Sintashta chariots, or earlier. The analyses also considerably revise the diversification time of the two EHV-4 subclades from the 16th century based solely on modern data to nearly a thousand years ago. Our study paves the way for a robust reconstruction of the history of non-human pathogens and their impact on animal health.

Keywords: equine herpesvirus 4; sintashta; virus evolution; palaeogenomics; horse.

Introduction

Equine herpesviruses are pathogens of major economic, clinical, and epidemiological importance to the horse industry worldwide (Patel and Heldens 2005; Lunn et al. 2009; Ma, Azab, and Osterrieder 2013; EFSA Panel on Animal Health and Welfare (AHAW) et al. 2022). The equine alphaherpesvirus 1 (EHV-1) outbreak of early 2021 in Europe was the most severe in decades, which led the European Food Safety Authority to implement disease prevention and control measures in accordance with the Animal Health Law (Patel and Heldens 2005; EFSA Panel on Animal Health and Welfare, 2022). Following infection by EHV-1, horses develop respiratory syndromes that can not only lead to abortions in pregnant mares and increase mortality rates in neonatal foals but can also evolve into myeloencephalitis (Telford et al. 1992; Studdert et al. 2003; Pavulraj et al. 2021). Other equine herpesviruses, such as equine gammaherpesviruses 2 and 5, are ubiquitous in horse populations (70–90 percent seroprevalence) but rarely lead to symptomatic infections, while the EHV-3 alphaherpesvirus causes equine coital exanthema, a venereal mucocutaneous disease (Stasiak, Dunowska, and Rola 2018; Vissani, Damiani, and Barrandeguy 2021). The EHV-4 alphaherpesvirus, on the contrary, primarily causes upper respiratory tract disease but only occasionally advances into neurological disorders and abortions (Telford et al. 1998; Khattab et al. 2022). Yet, EHV-4 remains a major concern due to its higher seroprevalence across the world (60–80 per cent among horses, donkeys, and their mule hybrids (Mekonnen, Eshetu, and Gizaw 2017; Pavulraj et al. 2021)), which may derive from its capacity to develop year-round infections as opposed to winter for EHV-1 (Matsumura et al. 1992; Ma, Azab, and Osterrieder 2013; Vaz et al. 2016; Pavulraj et al. 2021). It is also known to establish life-long latency in the trigeminal ganglion and the submandibular lymph nodes following primary infection, leading to likely periodical reactivation (Matsumura et al. 1992; Borchers, Wolfinger, and Ludwig 1999; Patel and Heldens 2005; Osterrieder and Van de Walle 2010; Ma, Azab, and Osterrieder 2013; Vargas-Bermudez et al. 2018; Pavulraj et al. 2021). Owing to the considerable economic threat EHV-4 poses, much attention has been given to its pathology. Yet, its evolutionary history in the context of horse domestication remains uninvestigated.

Textual sources provide evidence for respiratory tract disease in horses as early as 412 BCE in Sicily (Williams 1924). Whether EHV-4 or any other virus causing influenza-type diseases was

responsible for these outbreaks cannot, however, be determined from historical accounts nor from archaeological bone assemblages in the absence of obvious pathological lesions. As a consequence, the earliest confirmed record of an equine herpesvirus-like disease dates back to 1936, when the first EHV infection was described in the USA by Dimock and Edwards (Dimock and Edwards 1933, 1936; Bryans and Allen 1989). Further reports followed shortly after, for example, in Hungary in 1941 (Manninger and Csontos 1941; Bryans and Allen 1989). EHV-1 and EHV-4 were, however, only recognised as genetically closely related yet distinct viruses in 1981 (Telford et al. 1998). Therefore, determining which of these two virus types was responsible for the pre-1981 outbreaks remained difficult until serological tests became available from 1985 (Duxbury and Oxer 1968; Burrows and Goodridge 1974; Yeargan, Allen, and Bryans 1985; Allen and Bryans 1986; Matsumura et al. 1992; Crabb and Studdert 1995; Telford et al. 1998; Gilkerson et al. 1999; Patel and Heldens 2005; Vaz et al. 2016). Since such retrospective genetic analyses have been limited to samples collected in the second half of the 20th century, the potential importance of EHV-4 outbreaks in deeper historical time remains overlooked.

Over the last decade, methodological advances in ancient DNA research have gained access to genetic information of past populations (Orlando et al. 2021). This has revolutionised our understanding of human history, not only revealing changing patterns of mobility, admixture, and selection through space and time (Nielsen et al. 2017), but also the causative agents of past epizootics (Spyrou et al. 2019a). Ancient DNA analyses of human bone assemblages have, thus far, led to the successful genome characterisation of a diversity of pathogens, including both bacterial (e.g. *Yersinia pestis* (Spyrou et al. 2022), *Mycobacterium leprae* (Schuenemann et al. 2013), and *Salmonella enterica* (Key et al. 2020)) and viral (e.g. Hepatitis B virus (HBV) (Kocher et al. 2021), smallpox (Duggan et al. 2016), and Herpes Simplex virus (HSV) (Guellil et al. 2022)). Faunal archaeological remains have, however, received considerably less attention with respect to pathogen DNA characterisation, despite the reported impact of animal disease outbreaks throughout history (Roman-Binois 2017; Frantz et al. 2020). For example, glanders, another established respiratory tract disease, and its impact on cavalries may have been crucial in deciding the outcome of battles, which is a subject of repeated debate among historians (Chandler 1963; Sharrer 1995; Dvorak and Spickler 2008). The extensive sequencing of horse archaeological remains dating to the last 50,000 years provide a unique opportunity to start mapping infectious diseases prior to, during, and following domestication around 4,200 years ago in the Western Eurasian steppes (Librado et al. 2021; Orlando et al. 2021), with the potential to solve long-standing historical debates.

In this study, we screened the sequencing data of 264 ancient horses first presented by Librado and colleagues (Librado et al. 2021) for the presence of EHV-1 and EHV-4. We identify one Bronze Age horse as positive for EHV-4. The underlying archaeological bone was radiocarbon dated to 1,853 cal BCE and found in association with Sintashta material culture, which developed the spoked wheel technology and spread horse-drawn chariots across the Asian steppes between ca. 3,800 and 4,100 years ago (Anthony 2007). The available sequence data led to the characterisation of a partial EHV-4 genome sequence, with a 4.2× average depth-of-coverage. Phylogenetic modelling reveals that EHV-4 was pervasive in domestic horse populations for at least 4,000 years. We identify two amino acid changes that have been positively selected following the emergence of modern subclades around 1,000 years ago in ORF48, which codes for protein UL14 supposedly important in virion morphogenesis. This work, and the exponentially increasing sequencing data made available from faunal archaeological remains (Frantz et al. 2020), opens for a deeper understanding of the past history of infectious diseases in domestic animals and their impact on health, economy, and societies.

Results

EHV-4 identification and authentication

Raw FastQ sequence files underlying the comprehensive dataset of ancient horse genomes published in Librado et al. (2021) were mapped against de novo assemblies of EHV-1 and EHV-4 genomes that were used as reference genomes (Telford et al. 1992, 1998). The sequence data from the five libraries constructed from the DNA extracts of the tibia of a mare labelled UR17x29 (BioSample: SAMEA9533417) and recovered on the burial site of Kamennyi Ambar 5 (kurgan 8, grave 2), Kartaly District, south of the Chelyabinsk Oblast, Russia (Fig. 1A; see Librado et al. 2021 for complete archaeological details), showed a minute fraction of high-quality alignments against the EHV-4 genome (isolate 306–74, Genbank Acc. KT324743, (Vaz et al. 2016)). This specimen was previously radiocarbon dated to 1,936–1,770 cal BCE (95.4 per cent probability, InCal2020 (Reimer et al. 2020); 3,530 ± 30 uncal. BP, UCIAMS-199,248), providing a median age estimate of 1,853 cal BCE. In order to confirm the abovementioned results as true-positives for EHV-4, we carried out competitive mapping against candidate reference genomes in HAYSTAC (Dimopoulos et al. 2022) and alignment-free taxonomic assignment in Kraken2 (Wood, Lu, and Langmead 2019). The potential candidates were selected to represent the default list of pathogens from Heuristic Operations for Pathogen Screening (HOPS) (n= 220) which we complemented with 27 horse-specific pathogens (see Supplementary Table S1 for details). Both approaches returned consistent results across all libraries (ERR6466108-ERR6466112), supporting the presence of opportunistic pathogens abundant in the environment (e.g. *Mycobacterium* complex), as well as EHV-4 and another equine-specific pathogen, namely, *Prescottella equi* (Supplementary Tables S2 and S3). While *P. equi* represents a well-known opportunistic pathogen in immuno-compromised horses, we decided to focus the downstream analyses on EHV-4, as the only one candidate supported with a sufficient number of reads to assess post-mortem DNA damage profile and associated with significant genome coverage (>0.05×). We therefore considered EHV-4 BAM alignment that we filtered for PCR duplicates and alignments showing mapping quality scores inferior to 37. This left a total of 13,019 sequence alignments, corresponding to an average depth-of-coverage of 4.24-fold, with 68.57 per cent of the sites covered ≥3× (Fig. 2A). Local coverage dropped within GC-rich and/or duplicated repeat regions including ORF64–ORF66, as expected (Fig. 2A).

All the aligned sequences consisted of read pairs showing sufficient overlap to be collapsed as part of a single template, due to the extreme short size of the DNA molecules recovered (average = 46.9 bp; median= 42 bp; 95 per cent confidence interval = 27–84 bp). Additionally, the genomic positions immediately preceding alignment starts were enriched in cytosine residues, while those immediately following were enriched in guanine residues. This was true for both horse and EHV-4 read alignments (Supplementary Fig. S1A and S1B, respectively), and reflects the enzymatic treatment of DNA extracts with the USER mix (Rohland et al. 2015), which is

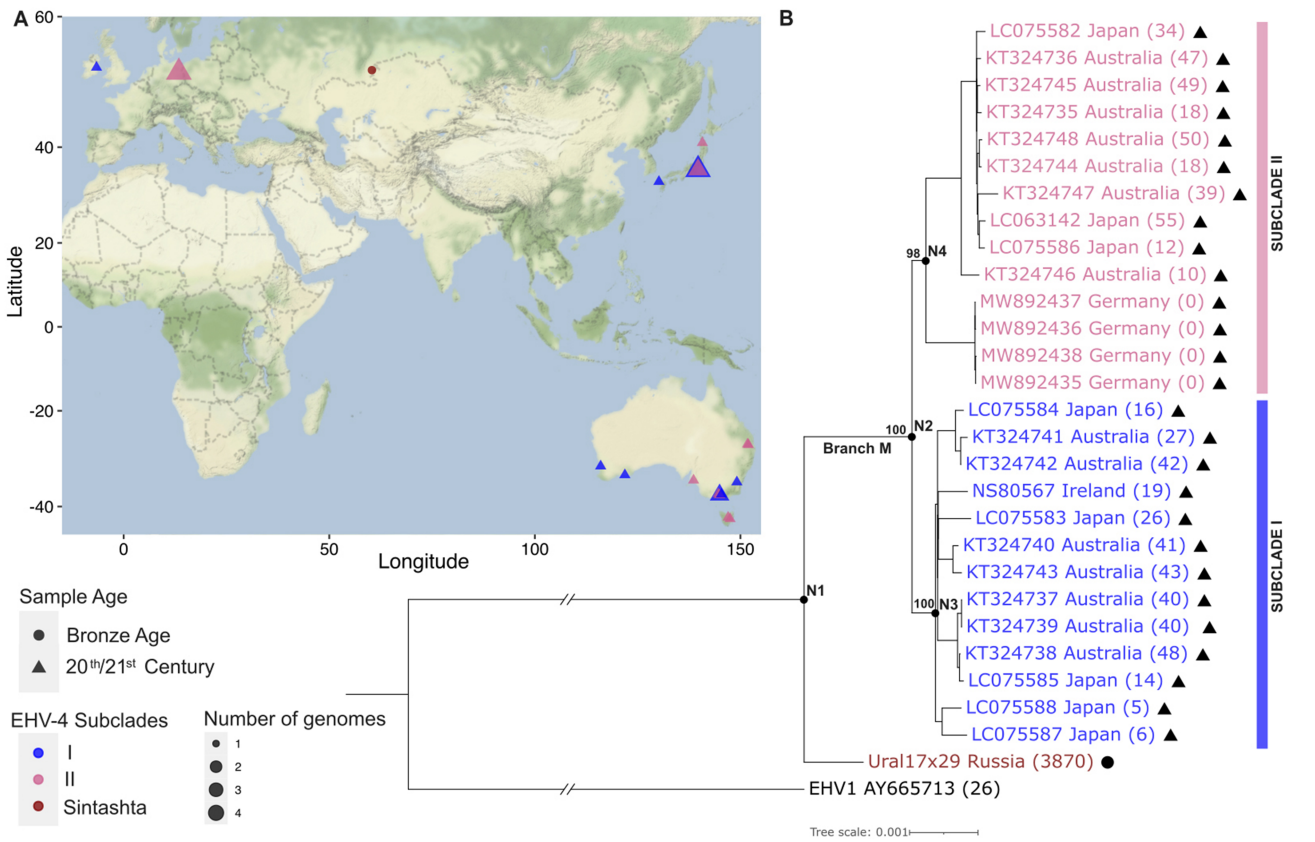

Figure 1. Sample map and phylogenetic relationships. (A) Location of the samples used in this study, including subclades, sample dates, and numbers of genomes. For samples with geographic information at country-level, the coordinates of the regional or country capital were used. (B) Maximum-likelihood tree and node bootstrap support. Sample names comprise Genbank accession numbers, sampling locations and, in parentheses, sampling dates in years BP with '2017' set as 'Present' based on the most recently sampled isolates. The symbols refer to Fig. 1A. EHV-1 was used as an outgroup.

known to cleave DNA strands at those cytosines deaminated into uracils post-mortem, leaving only moderate C→T nucleotide mis-incorporations characteristic of ancient DNA damage (Orlando et al. 2021) (0.706 per cent at alignment starts versus 0.211 per cent on average across the following 24 positions). Complementary G→A mis-incorporation rates were inflated at alignment stops (0.468 percent versus 0.211 percent on average across the preceding 24 positions). The observed sequence complementarity of mis-incorporation signatures at read termini is typical of ancient DNA data, once incorporated into double stranded DNA libraries following USER enzymatic treatment (Orlando et al. 2021). Together with the short size of the DNA templates characterised and the stringent quality filters used in sequence alignments, these findings suggest that, despite representing only a limited fraction of the metagenomic content of the archaeological bone analysed, authentic ancient EHV-4 DNA molecules were sequenced together with the DNA molecules of their host, which covered the horse reference genome at an average of 2.08-fold depth-of-coverage (Librado et al. 2021).

Phylogenetic analyses

At the time of writing, a total of 27 published EHV-4 genomes were available from previous studies. These were generated from fresh biological tissues that were collected in Australia, Germany, Ireland, and Japan between 1962 and 2017 (Telford et al. 1998; Vaz et al. 2016; Izume et al. 2017; Pavulraj et al. 2021) (Supplementary Table S4). However, using MAFFT v7.453 (Katoh and Standley 2013), these sequences and the ancient UR17x29 genome characterised here for the first time proved difficult to align against the EHV-1 reference sequence (strain Ab4p; AY665713). The inclusion of EHV-1 as an outgroup is, however, required to polarise phylogenetic reconstructions. We, thus, aligned each ORF independently and concatenated all the resulting alignments together before undertaking phylogenetic reconstruction, with the exception of ORFs 24, 47/44 and 71, which include numerous tandem repeats, duplicated ORFs (corresponding to ORF64–ORF67 for EHV-1 but to ORF64–ORF66 for EHV-4; Fig. 2A), and previously identified tandem repeat regions (Izume et al. 2017). The concatenation of ORF sequences provided a final alignment comprising 104,418 orthologous sites. Maximum-likelihood phylogenetic analysis reiterated the previously reported subdivision of EHV-4 viruses within two main subclades, referred to as subclades I and II (Fig. 1B). While these subclades are not defined by geography and/or pathogenicity, subclade I is known to almost exclusively consist of abortion isolates (Izume et al. 2017; Kolb et al. 2017). The phylogenetic reconstruction also returned strong support for the existence of a subgroup within subclade II defined by a single T-to-C point mutation affecting position 624 of ORF33 (61,770 on EHV-4 genome reference AF030027) (Telford et al. 1998; Pavulraj et al. 2021). Our maximum-likelihood phylogenetic analysis and pairwise genetic distance matrix revealed that the UR17x29 ancient sequence was considerably distant to the EHV-1 outgroup (0.32739) but represented a close relative to all modern EHV-4 sequences hitherto sequenced, showing pairwise genetic distances no

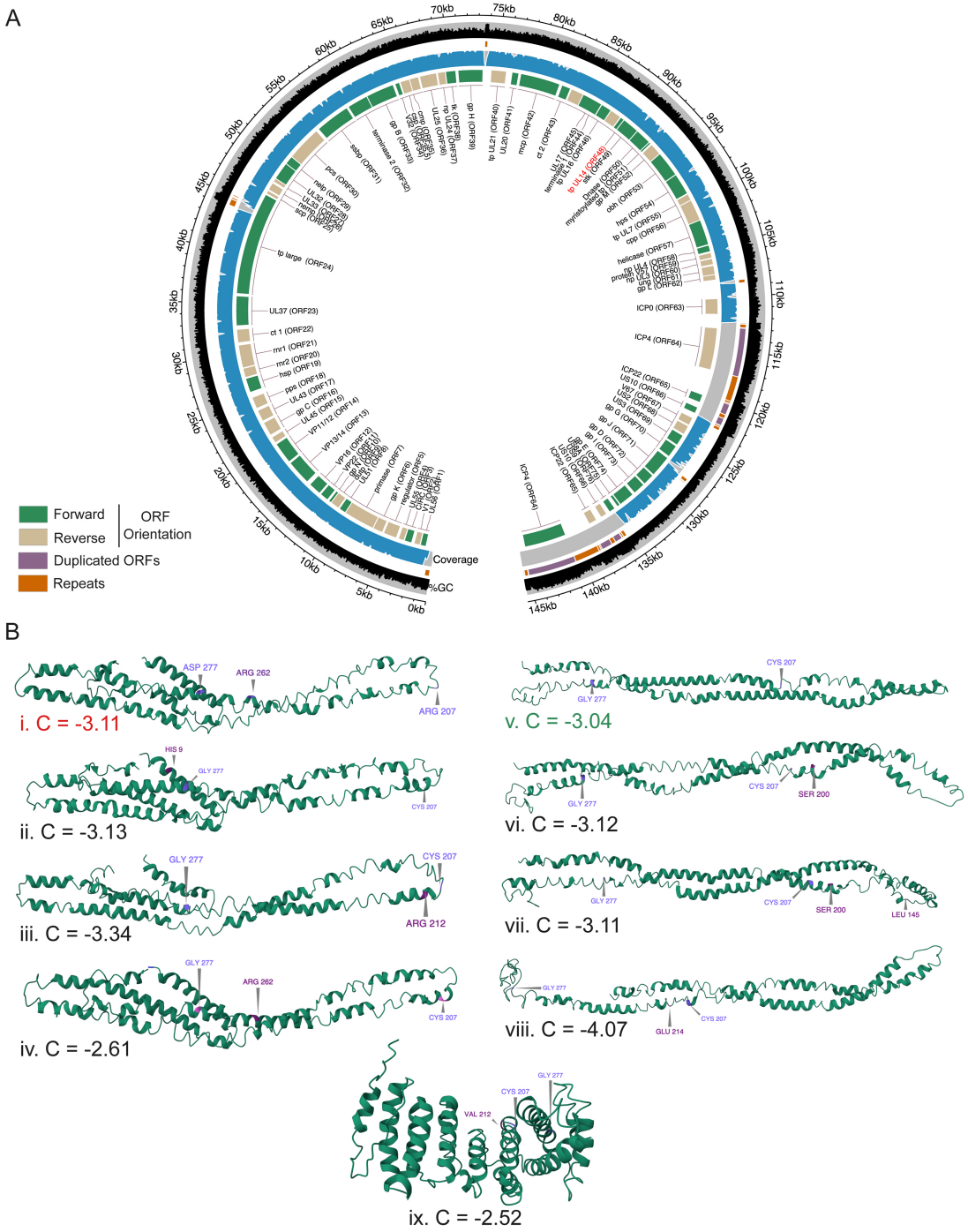

Figure 2. EHV-4 genome organisation and protein structure predictions. (A) %GC and coverage was calculated within non-overlapping 50 bp and 100 bp sliding windows, respectively. Coverage is given as a percentage based on a maximum of 7.3×. The open reading frame encoding for the tegument protein (tp) UL14 (highlighted in red) was found to have been positively selected along the branch ancestral to subclades I and II (Branch M, Fig. 1B). The grey area on the coverage track shows the drop in coverage represented by the duplicated repeat regions including ORF64–ORF66. cmp: capsid maturation protein; cpp: capsid portal protein; csp: capsid scaffold protein; ct: capsid triplex; dutp: deoxiuridine triphosphatase; gp: glycoprotein; hps: helicase primase subunit; hsp: host shutoff protein; mcp: major capsid protein; nelp: nuclear egress lamina protein; nemp: nuclear egress membrane protein; np: nuclear protein; pcs: DNA polymerase catalytic subunit; obh: origin binding helicase; pps: DNA polymerase processivity subunit; rnr: ribonucleotide reductase; scp: small capsid protein; ssbp: single-stranded DNA binding protein; stk: serine/threonine protein kinase; tk: thymidine kinase; tp: tegument protein; ung: uracil DNA glycosylase. (B) Secondary structures predicted by I-TASSER for UL14 proteins. Amino acid substitutions are shown in light purple (substitutions differing between the ancient and modern versions) and in dark purple (substitutions differing from the most commonly observed version). The confidence score C denotes the quality of predicted models ranging from –5 to 2. The predictions for the most common version and UR17x29 are shown in green and red c-scores, respectively.

larger than 0.00369 (Fig. 1B, Supplementary Table S5). The UR17x29 genome was, however, found to fall outside the monophyletic group of modern EHV-4 strains including both subclades. This supports the UR17x29 genome as deriving from an ancient, divergent EHV-4 strain not segregating today, ruling out modern contamination.

Recombination

The extensive recombination previously reported in EHV-4 (Vaz et al. 2016; Kolb et al. 2017) may have introduced bias in the phylogenetic reconstructions mentioned earlier, as these were based on genome-wide ORF alignments. The multiple reticulate and pairwise homoplasy (Phi) test analyses implemented in SplitsTree4 (Huson and Bryant 2006) indeed provided significant statistical support for the presence of recombination in our dataset (P-value = 2.628E-8). To further address the extent to which such recombination may have affected our phylogenetic findings, we mapped the recombination breakpoints using the most commonly used detection methods (RDP (Martin and Rybicki 2000), GENECONV (Padidam, Sawyer, and Fauquet 1999), Chimaera (Posada and Crandall 2001), MaxChi (Maynard Smith 1992), BootScan (Martin et al. 2005), SiScan (Gibbs, Armstrong, and Gibbs 2000), and 3Seq (Lam, Ratmann, and Boni 2018)) implemented in RDP5 (Martin et al. 2021) and relying on two successive analytical steps. Considering all methods together, as well as conservative P-values and recombination scores, these analyses predicted the presence of 12 recombination events throughout the EHV-4 genome (Supplementary Table S6). The largest genomic block predicted to not be affected by recombination spanned approximately 85 kb (nucleotide coordinates 4,254–90,128 of the EHV-4 reference genome AF030027 (Telford et al. 1998) with tandem repeats excluded), which was further confirmed to be devoid of recombination through the Phi test implemented in SplitsTree4 (P = 0.1121). Phylogenetic reconstruction through IQ-TREE (Minh et al. 2020) (Supplementary Fig. S2) returned the same topology as that obtained on the genome-wide ORF alignments, ruling out any significant impact of recombination in the evolutionary history reconstructed above.

Divergence times

The characterisation of a deeply divergent ancient genome sequence provided the first opportunity to calibrate the timing of major phylogenetic divergence events underlying the EHV-4 evolutionary tree. To achieve this, we applied the Bayesian framework implemented in BEAST v2.6.7 (Bouckaert et al. 2019) to the 85 kb recombination-free alignment. Bayesian phylogeny was reconstructed under the birth–death skyline serial model with tip-dating and assuming either a strict or a relaxed lognormal clock. Model selection through Nested Sampling analyses (Skilling 2006; Maturana-Russel et al. 2019) as implemented in the Beast2 NS package v.1.1.0. revealed that, while no one model could be confidently selected, both models were comparable in their results (Fig. 3A and Supplementary Fig. S3). To follow a conservative approach, we applied the strict clock model for assessing divergence times. We assessed the temporal signal in our dataset using an approach published by Duchêne and Duchêne (2020) which involves comparing the prior and posterior distributions of the tree height to assess the informativeness of our sequenced data. With a CV_{ratio} of 4.04, our results implied a more informative posterior; though as the strength of this signal is unknown, we caution that our data may not actually display temporal signal (see Duchêne and Duchêne 2020 for methodological details). Hence, to address this possibility, we included a molecular clock rate prior as an additional calibration to our tip-dating.

The estimated age of the EHV-4 tMRCA (time to the Most Recent Common Ancestor) was dated to 1,905 BCE (median; 95 per cent confidence interval (CI) = 2,090–1,853 BCE; node N1, Fig. 3A and B). This date estimate is contemporary to the spread of the Sintashta culture, which is also associated with the archaeological context into which sample UR17x29 was found. Therefore, our work provides definitive proof that EHV-4 infected horses as far back as the Bronze Age. We caution that the 1,905 BCE age estimate should not be confused with the earliest horse infection by EHV-4. In fact, it is reasonable that EHV-4-like alphaherpesviruses already infected equine ancestors as they diversified in modern species 4 to 4.5 million years ago (Orlando et al. 2013), given that the alphaherpesvirus evolutionary tree mirrors that of their mammalian hosts, in line with a co-divergence history of viruses with their hosts. The Bronze Age virus characterised in this study also significantly revised the divergence time between subclades I and II (node N2) to around 1,071 CE (Fig. 3A, median; 95 per cent CI = 899–1,214 CE), versus 1,531 CE (Supplementary Fig. S4, 95 per cent CI = 1,361–1,669 CE). Therefore, the two main phylogenetic subclades structuring modern EHV-4 diversity already emerged a thousand years ago.

Positive selection

We next leveraged the robust phylogeny reconstructed above to test for the presence of sites that were positively selected along the branch leading to modern subclades I and II (branch M, Figs 1B and 3A). The branch-site model implemented in PAML v.4.9 (Yang 2007) revealed ORF48 as the only positively selected ORF ($P < 0.05$), with two of the five amino acid substitutions suggested as potentially adaptive appearing fixed in all modern strains (C619T and A830G changing Cysteine into Arginine and Aspartate into Glutamate, respectively). The prediction of three other polymorphic sites as potentially adaptive along branch M may result from a combination of a backward mutation in the German samples clustering within subclade II (G785A changing Arginine into Glutamine), and confounding consequences of recombination as ORF48 includes one predicted recombination breakpoint. Similar selection scans failed at detecting selection candidates between subclades I and II. Testing for episodic diversifying selection at individual sites through MEME (Mixed Effects Model of Evolution) (Murrell et al. 2012) within the Hyphy (Kosakovsky Pond et al. 2020) suite of analyses did not detect any evidence of such selection for any of our ORFs ($p < 0.1$). This implies that the positive selection observed on ORF48 through PAML is not episodic or found in a specific subset of branches. Instead, it appears to have been consistently favoured sometime after 1,853 cal BCE (corresponding to the age of our sample) but prior to the diversification of the two EHV-4 subclades.

ORF48 encodes for the UL14 tegument protein, which is hypothesised to be involved in virion morphogenesis, by transporting other tegument proteins from the nucleus into the cytoplasm (Telford et al. 1998). The best secondary structures predicted by I-TASSER (Yang et al. 2015) showed two different prominent models comprising eight versions of UL14 proteins found amongst modern strains and ancient UR17x29 (Fig. 2B) with an exception of the Irish strain NS80567 (Fig. 2B, ix). No structural difference was observed between UL14 protein structures of modern strains and of UR17x29 as shown in Fig. 2B, suggesting that the two point mutations observed above may not play a decisive role in the secondary structure of UL14 protein. However, due to the lack of confirmed structure/s of homologous protein/s, our analyses cannot be considered as conclusive but rather remain predictive in nature. UL14 has been shown to possess anti-apoptotic and heat-shock protein-like activities while playing a pivotal role in protein localisation in various herpesviruses increasing longevity of an infected cell (Cunningham et al. 2000; Yamauchi et al. 2002, 2008; De Martino et al. 2007; Oda et al. 2016). Therefore, the strong signature of positive selection detected for UL14 suggests that key evolutionary changes may have contributed to

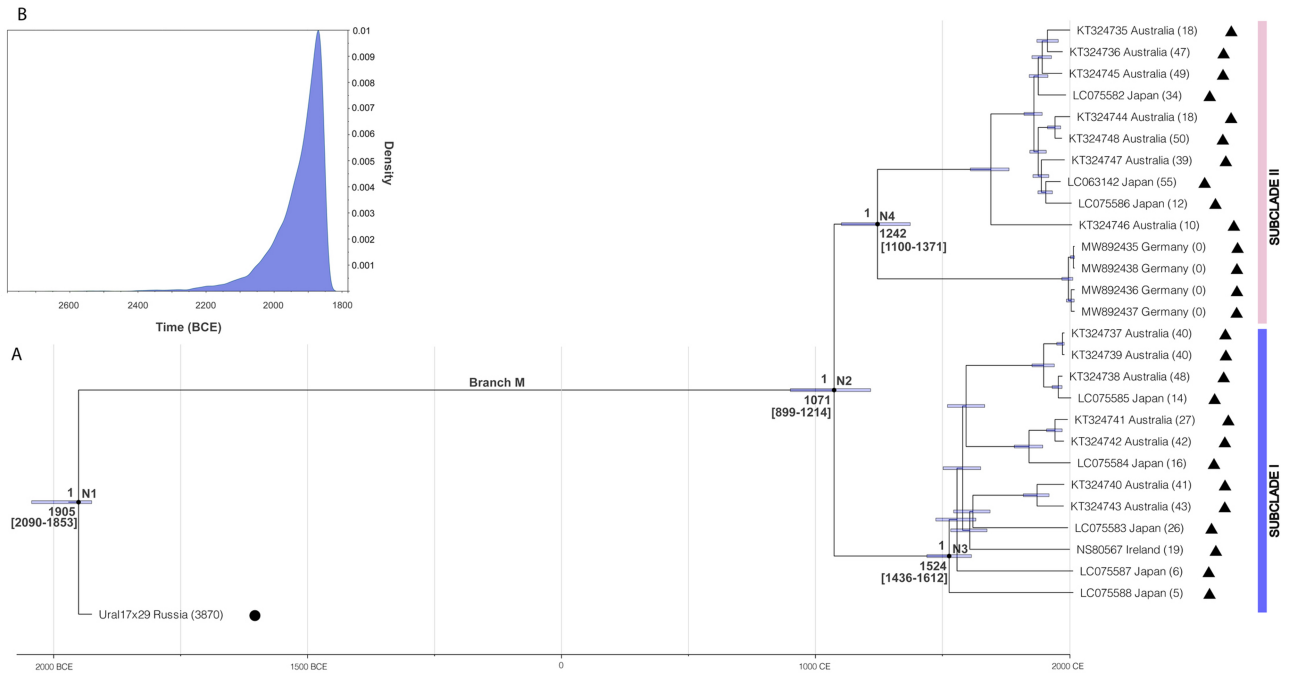

Figure 3. Time-scaled EHV-4 evolutionary history. (A) Bayesian phylogeny under the strict clock and birth–death skyline serial models based on the ancient and modern dataset. Node posterior supports are provided above nodes to the left, while median estimates of divergence times are provided below nodes together with the 95% CI in square brackets [Height_95 per cent_HPD]. Sequence labels refer to Genbank accession numbers, sampling locations and, in parentheses, sampling dates formatted as years BP with 'Present' set to '2017'. These dates were used for tip-dating calibration. The symbols refer to Fig. 1A. (B) Posterior distribution of the tMRCA of all EHV-4 strains considered in this study (N1). Time is indicated in calendar years (BCE: Before Common Era, CE: Common Era).

post-replication viral sustenance in infected cells. *In-vitro* study of UL14 deletion mutants and X-ray crystallography-based structure reconstructions are, however, needed to fully understand the role and the exact functional consequences of the two point mutations identified here.

Discussion

This study demonstrates that EHV-4 was present in the south-eastern Urals in 1,853 BCE. This pushes back the age of the oldest EHV-4 virus characterised so far by nearly 3,800 years. The viral DNA material was preserved in a tibia that was found in a cemetery associated with the Sintashta culture. This culture is well-known for triggering important socio-economic changes in the Bronze Age Eurasian steppes. These range from proto-urban developments in settlement organisation, metallurgy (Hanks et al. 2018; Librado et al. 2021), and the invention of spoke-wheeled chariotry (Hanks et al. 2018; Klecel and Martyniuk 2021; Librado et al. 2021), which not only revolutionised transport but also accompanied the rise of innovative warfare technologies (Drews 1993; Anthony 2009; Klecel and Martyniuk 2021). The spread of Sintashta–associated populations together with their horse-driven spoke-wheeled chariots fuelled the expansion of Indo-Iranian languages across Central Asia and beyond (Allentoft et al. 2015; Narasimhan et al. 2019; Lazaridis et al. 2022). Interestingly, patterns of mitochondrial, Y-chromosomal, and nuclear variation in ancient horses (Gaunitz et al. 2018; Fages et al. 2019; Librado et al. 2021) recently showed that the genetic lineage that gave rise to modern domestic horses (DOM2) expanded both demographically and geographically in the late 3rd millennium BCE. The period spanning 2,090–1,853 BCE, which we estimated for the EHV-4 tMRCA, not only coincides with the rapid increase of the DOM2 effective size by at least an order of magnitude but also their expansion outside the domestication homeland, from the lower Volga-Don basin into Moldova, Anatolia, and Central Asia (Librado et al. 2021). In fact, this period marks the time when horses became paramount to human societies (Kelekna 2009). The increasingly global demand for horses in relation to chariotry, warfare, and mobility, as well as the underlying changes in management practices would, thus, have expedited the spread of EHV-4. This historical turning point may also have facilitated the dispersal of other equine diseases across long distances, and human diseases, as recently suggested for the plague (Rascovan et al. 2019; Valtueña et al. 2022). The tMRCA of all EHV-4 strains including our ancient genome slightly postdates the rise of DOM2 horse domestic bloodlines. This may indicate that domestication resulted in considerable subsampling of EHV-4 diversity, as a result of a demographic bottleneck within host populations, and/or that the rapid demographic growth and expansion of DOM2 horses has favoured the surfing of only a subset of the EHV-4 diversity then present. The latter has been previously reported for parasites, in which diversity is lost as the host population expands away from its original environment (White and Perkins 2012). Disentangling both scenarios will require mapping out the EHV-4 diversity in pre-DOM2 lineages. Given the relatively limited EHV-4 prevalence rate measured through shotgun sequencing in our ancient horse panel, this will likely require dedicated target-enrichment approaches. Furthermore, our knowledge of modern EHV-4 diversity is currently limited to a few countries only and should be characterised globally before the true impact of EHV-4 infection during domestication can be measured.

In this study, only a single EHV-4 genome could be characterised amongst 135 ancient DOM2 individuals (0.7 per cent). This is in striking contrast with the high seroprevalence observed in

unvaccinated equids today (60–80 perc ent (Mekonnen, Eshetu, and Gizaw 2017; Pavulraj et al. 2021)). At first glance, this would suggest a low seroprevalence rate in ancient populations. However, EHV-4 infections generally start through the inhalation of aerosols containing viral particles (Ma, Azab, and Osterrieder 2013; Khusro et al. 2020), and viral replication remains primarily restricted to the epithelial cells of the upper respiratory tract (Osterrieder and Van de Walle 2010; Vandekerckhove et al. 2011; Ma, Azab, and Osterrieder 2013). Such soft tissues are generally not available in the archaeological record, which considerably limits our detection capacity. Regardless, the detection rate of viral material in extant animals is known to not necessarily reflect seroprevalence, as reported in other herpesviruses, such as the human Varicella zoster virus (Moustafa et al. 2017). Therefore, the single positive case detected in our study does not inform on past seroprevalence rates but only reflects a horse experiencing high viral replication with substantial viral particles in the systemic blood circulation (Pusterla et al. 2005; Vandekerckhove et al. 2011).

Our work, although limited to only a partial sequence of the EHV-4 genome (68.57 per cent), provides the first temporal calibration for the evolutionary tree of EHV-4 viruses. In contrast to horses for which an abundant palaeontological record provides fine-grained resolution into macroevolutionary changes (Macfadden 2005), equine viruses and pathogens hardly leave any obvious fossil remains. Therefore, the molecular traces left by the viral pathogens once infecting ancient equine populations offer the unique opportunity to access and chart past epidemiological outbreaks through space and time. The cultural and osteological material composing the archaeological context also adds crucial information to refine the social, societal, and environmental conditions driving the spread of past infectious diseases. The immense potential of ancient DNA for advancing understanding of the evolutionary history of pathogens is now fully established for humans (Spyrou et al. 2019a), following pioneering work characterising ancient plague (Spyrou et al. 2019b; Valtueña et al. 2022), tuberculosis (Bos et al. 2014; Kay et al. 2015; Kerner et al. 2021), leprosy (Schuenemann et al. 2013, 2018; Pfrengle et al. 2021), and more. Beyond bacterial pathogen vectors, this approach has also started to illuminate the past history of both DNA (e.g. HBV (Kocher et al. 2021), HSV (Guellil et al. 2022)) and RNA (e.g. Influenza (Taubenberger et al. 2005), Measles (Düx et al. 2020)) viruses. Despite such potential and the growing interest for animal domestication in ancient DNA research (Frantz et al. 2020), this approach remains overlooked for non-human animals. Yet, the domestication process itself, as well as the mass production sustaining an ever-increasing human population, are known to have increasingly exposed animals to infectious diseases, with particularly detrimental consequences on their health (Morand, McIntyre, and Baylis 2014). Therefore, the methodology developed in our study, and more generally the tools supporting ancient DNA research, provide an effective framework to both map out animal pathogens in the past and to assess the impact of domestication and changing management practices on animal health.

Materials and methods

Sample and sequence alignment

We mapped the shotgun sequence data underlying the comprehensive dataset of ancient horse genomes published in Librado et al. (2021) against the complete reference genomes of EHV-1 (AY665713 (Telford et al. 1992)) and EHV-4 (KT324743 (Vaz et al. 2016)). Complete archaeological details for all samples pertaining to this comprehensive dataset can be found in the Supplementary Methods of Librado et al. (2021). The sequence data were processed using PALEOMIX (Schubert et al. 2014) and the BWA aligner (Li and Durbin 2009), disregarding seed (-l 1024) and using the stringent alignment filters of (Spyrou et al. 2019b) (i.e. -n 0.1, quality scores greater or equal to 37). The UR17x29 sample (BioSample: SAMEA9533417) returned substantial high-quality alignments for EHV-4. To confirm if the reads were true positives, we ran HAYSTAC (Dimopoulos et al. 2022) and Kraken2 (Wood, Lu, and Langmead 2019) on all five libraries (ERR6466108-ERR6466112). For both HAYSTAC and Kraken2, we referred to the HOPS default pathogen list (https://github.com/rhuebler/HOPS/blob/external/Resources/default_list.txt), which comprised 356 pathogens, though two entries were duplicates, hence bringing the total of pathogens to 354. We removed pathogens listed as Family or Genus level as no specific reference genome could be given. We also excluded substrains and incomplete genomes. We complemented this dataset by adding 27 horse-specific pathogens, bringing our final reference database to 247 pathogens (see Supplementary Table S1 for a detailed list of pathogens considered including NCBI accession numbers and reasons for exclusion). We custom-built the database using HAYSTAC (Dimopoulos et al. 2022) by providing the NCBI Accession numbers and ran the analysis using the abundance mode (—mode abundances) and a posterior probability for taxon assignment $\geq$0.75 (default parameter). Kraken2 was run with default parameters (Wood, Lu, and Langmead 2019). We selected EHV-4 for downstream analyses based on the following HAYSTAC thresholds: Dirichlet read number >1000 is required for reliable mapDamage analysis, an evenness of coverage <10, and a coverage of >0.05×. Fragmentation and nucleotide misincorporation plots were generated using mapDamage2 v2.0.8 (Jónsson et al. 2013), using default parameters (Supplementary Fig. S1A and S1B). The genome sequence was reconstructed using Bcftools (Danecek et al. 2021) mpileup, considering base phred quality scores of at least 30. The pileup sequence format (mpileup) provided the input for calling genotypes, assuming haploidy. Following a normalization step (norm -c x -d all), sites showing coverage below 3× or genotype phred qualities below 30 were disregarded. The resulting VCF file was indexed using tabix v1.7–41-g816a220 (Li et al. 2009), and the genome sequence was converted into fasta format using the vcf_to_fasta command from the PALEOMIX pipeline (Schubert et al. 2014).

Phylogenetic analyses

To investigate the phylogenetic placement of our ancient sequence, we first used NCBImeta v0.8.3 (Eaton 2020) to identify all complete EHV-4 genomes previously published in NCBI. It returned 141 entries of which 27 were complete unique genomes from four countries only, namely, Australia, Germany, Ireland, and Japan. We hence compiled a dataset of these 27 published EHV-4 genomes comprising the Irish strain NS80567 (Telford et al. 1998), 14 isolates from Australia (Vaz et al. 2016), eight isolates from Japan (Izume et al. 2017), and four isolates from Germany (Pavulraj et al. 2021) (Supplementary Table S4). The sequence of the EHV-1 reference strain Ab4p (AY665713) was added as an outgroup (Telford et al. 1992). We conducted a multiple sequence alignment using MAFFT v7.453 (Katoh and Standley 2013), but the EHV-4 sequences were difficult to align against the EHV-1 reference sequence. Therefore, with EHV-1 and EHV-4 genomes both comprising 76 orthologous genes, we proceeded to the independent alignment of each ORF using MAFFT v7.453, followed by a manual verification on AliView v1.28 (Larsson 2014) and Seaview

v4.7 (Gouy, Guindon, and Gascuel 2010). We observed considerable disparity significantly caused by direct tandem repeats in the alignments of ORF 24, 47/44 and 71 (previously observed in Izume et al. (2017)); these ORFs were thus removed from the alignment. The duplicated genes ORF64–ORF67 for EHV-1 and ORF64–ORF66 for EHV-4 were also discarded, as were the direct tandem repeat regions previously reported in the literature (Izume et al. 2017). Our procedure resulted in a multiple alignment of 104,418 orthologous sites comprising the sequences of 28 individual EHV-4 strains including our ancient sample and one EHV-1 outgroup. We then generated a maximum likelihood tree using IQ-Tree v2.0.3 (Minh et al. 2020), with 1,000 bootstrap pseudoreplicates (option -b 1000) and the EHV-1 reference strain Ab4p as the outgroup (option -o). The integrated program ModelFinder (Kalyaanamoorthy et al. 2017) identified the best substitution model according to BIC as K3Pu + F + G4 (option -m TEST), and the EHV-1 reference strain Ab4p was rooted as an outgroup when visualising the resulting tree topology in iTOL v6.7 (Letunic and Bork 2021). Pairwise genetic distances were computed between pairs at tips from the maximum likelihood phylogenetic tree using its branch lengths through the cophenetic.phylo function in the R package ape (Paradis and Schliep 2019) (Supplementary Table S5).

Recombination analyses

Prior to conducting Bayesian phylogenetic analyses, we first tested for recombination in our dataset. We generated a multiple sequence alignment comprising our ancient and all 27 modern EHV-4 full genomes, using MAFFT v7.453 (Katoh and Standley 2013). We manually verified the alignment and removed the 17 direct tandem repeat regions identified in Izume et al. (2017) using Seaview v4.7 (Gouy, Guindon, and Gascuel 2010), which provided a final alignment of 28 genomes along 142,260 sites. Recombination was investigated using SplitsTree4 v4.18.3 (Huson and Bryant 2006) and RDP5 (Martin et al. 2021). Phylogenetic recombination networks were generated in SplitsTree4 using default parameters, and the statistical analyses of the recombination networks were performed using the pairwise homoplasy (Phi) test (Bruen, Philippe, and Bryant 2006) as implemented in the software. Breakpoint identification was conducted using the seven most commonly used recombination detection methods implemented in RDP5 (v5.29), comprising RDP (Martin and Rybicki 2000), GENECONV (Padidam, Sawyer, and Fauquet 1999), Chimaera (Posada and Crandall 2001), MaxChi (Maynard Smith 1992), BootScan (Martin et al. 2005), SiScan (Gibbs, Armstrong, and Gibbs 2000), and 3Seq (Lam, Ratmann, and Boni 2018) and with the following parameters: linear sequences, Bonferroni correction, and highest acceptable P-value of 0.05. Recombination events identified by at least two of the seven methods, with $P < 0.05$ and good recombination scores (>0.4, (Martin et al. 2021)) were kept for downstream analyses to ensure as conservative an approach as possible. This resulted in 12 recombination events (Supplementary Table S6) leading to a conserved region of 84,900 sites common to all sequences (positions 4,254–90,128 of the EHV-4 reference genome AF030027 (Telford et al. 1998) with direct tandem repeat regions removed (Izume et al. 2017)).

Bayesian phylogenetics

To investigate the tMRCA of all EHV-4 sequences and the divergence time of phylogenetic subclades I and II, we constructed a time-calibrated Bayesian phylogeny based on the modern sequences and our ancient radiocarbon dated sample using BEAST v2.6.7 (Bouckaert et al. 2019), employing a birth–death skyline serial model and tip-dating. Tip dates were specified numerically as years Before Present (BP) with 'Present' set as '2017' based on the most recently sampled isolates (Pavulraj et al. 2021). To ascertain the best-fitting clock prior between a strict clock and a relaxed lognormal clock, we performed a model selection test using the Nested Sampling (NS) approach (Skilling 2006; Maturana-Russel et al. 2019) available through the NS package (v.1.1.0) and based upon established guidelines from the dedicated Taming the Beast (Barido-Sottani et al. 2018) tutorial (https://taming-the-beast.org/tutorials/NS-tutorial/NS-tutorial.pdf) and the GitHub Nested Sampling FAQ (https://github.com/BEAST2-Dev/nested-sampling/wiki/FAQ). The NS analysis was run with a Chain Length of 1E + 12, a subchain length of 860,000 and up to 600 particles, which ensured the points generated at each iteration were independent and the marginal likelihoods were comparable between models (Strict clock: SD = 0.51; Relaxed clock log normal: SD = 0.52). The resulting log BF ($\log ML_{Relaxed}$—$\log ML_{Strict}$) was found to be 1.229, but we could not ascertain that this was not due to randomisation (log BF < 2*sqrt(SD1*SD1 + SD2*SD2) = 1.462) according to the Github Nested Sampling FAQ. However, both the Strict and Relaxed lognormal clock models display very little variation in the median of their tMRCA estimates (Strict = 1905 BCE versus Relaxed lognormal = 1947 BCE). We selected the Strict clock as it represents the more conservative model, whilst the Relaxed lognormal clock is available in our Supplementary Data (Supplementary Fig. S3). Temporal signal could not be investigated through a root-to-tip regression as a single ancient point would necessarily lead to a linear regression, rendering such a test non-applicable. Instead, we explored temporal signal based on the approach detailed in Duchêne and Duchêne (2020), which involves comparing the prior and posterior distributions for the age of the root node to assess the informativeness of our sequenced data. Quantifying this information can be done by taking the 95 per cent quantile width and dividing it by the mean for both the prior and the posterior, which gives us CV_{prior} and $CV_{posterior}$ values. We then calculate the CV_{ratio} based on the formula $CV_{prior}/CV_{posterior}$ (see Duchêne and Duchêne 2020 for further details on the method). Hence, to conduct this analysis, we ran the same birth–death skyline serial model with the same priors and MCMC chain length but with the option 'Sample from Prior'. We next measured the informativeness of the tree height prior and posterior distributions and obtained a CV_{prior} of 0.242210 and a $CV_{posterior}$ of 0.059952. This resulted in a CV_{ratio} of 4.04, which implies the posterior is more informative than the prior. However, as we do not know the strength of this signal, we cannot positively ascertain our data displays temporal signal (Duchêne and Duchêne 2020). As such, as well as our tip-dating, we specified a molecular clock rate prior as an additional calibration.

For our non-recombined EHV-4 dataset including UR17x29, K3Pu + F + I was identified as the best substitution model by ModelFinder (Kalyaanamoorthy et al. 2017; Minh et al. 2020) according to BIC. As this model was not available in BEAST v2.6.7, the next best model identified and available was GTR + F + I. All rates were set to the default value of 1.0 and estimated relative to the rate parameter of CT. Frequencies were set to Empirical. The Gamma Category Count was set to 0 and the Proportion of Invariant was estimated with a starting value of 0.9 based upon the ModelFinder estimate of 0.913. Proper priors were used throughout. Considering we cannot ensure temporal signal, we used a molecular clock rate prior as an additional calibration to the sequence sampling time. To reflect the epidemiology of EHV-4, we set the substitution rate to the default value of 1.0, and the clock rate to be

estimated with a starting value of 1.0E-6 substitutions per site per year, which corresponds to the average overall substitution rate for double-stranded DNA viruses (Sanjuán et al. 2010). We specified the lower and upper substitution rate boundaries to be 1.0E-8 and 1.0E-5, respectively, to reflect current knowledge on herpesviruses and used a lognormal prior distribution. For the effective reproductive number, Re, we used five equidistant intervals between the root and the last sample with a Log Normal (0, 1.25) prior distribution resulting in a median of 1 and most weight being placed below 7.82 but still encompassing the upper boundary. Re was set to be estimated with a starting value of 3.0 and a lower boundary of 0 and an upper boundary of 15 based on known Re infection rate of EHV-1 considering they belong to the same subtype (Meade 2012). The infection period of EHV-4 usually lasts 14 days but can extend up to 60 days (Pavulraj et al. 2021), meaning the 'become uninfectious rate' falls between 6 and 26 per year. We set the value for M of the log normal distribution in real space to 26, with a standard deviation of 0.75. This places our prior on the period of infectiousness primarily between 5.4 and 81 days (see the Taming the Beast (Barido-Sottani et al. 2018) tutorial 'Prior selection and clock calibration using Influenza A data by Bošková, Mitov, and du Plessis available at 'https://taming-the-beast.org/tutorials/Prior-selection/). The sampling proportion was set to a beta distribution with A = 1.0 and B = 999.0, reflecting the very low sampling proportion. We specified the origin of our tree prior to follow a uniform distribution, and to be estimated with a starting value of 4,000 years with lower and upper boundaries of 3,871 and Infinity, respectively, based on the age of our oldest sample. Our ucldMean prior followed a uniform distribution and was based on the clock rate input. Finally, the MCMC chain length was run for a total of 200 million iterations for the relaxed lognormal clock model, and for a total of 100 million iterations for the strict clock model. Convergence was visually confirmed by ensuring the effective sample sizes (ESS) estimation for all parameters was >200, disregarding the first 10 per cent iterations as burn-in, using Tracer v.1.7.2 (Rambaut et al. 2018). Maximum clade credibility trees with median node heights and following a 10 per cent burn-in were generated using Tree Annotator v2.6.6 as implemented in BEAST v2.6.7 (Bouckaert et al. 2019) and visualised with FigTree v1.4.4. (http://tree.bio.ed.ac.uk/software/figtree/).

For our non-recombined modern dataset, TIM + F + I was identified as the best substitution model by ModelFinder (Kalyaanamoorthy et al. 2017; Minh et al. 2020) according to BIC. As this model was unavailable in BEAST v2.6.7; the next best model identified and available was GTR + F + I. All parameters were identical to our non-recombined ancient dataset, with the following two exceptions: (1) the Proportion of Invariant sites was estimated with a starting value of 0.89 based upon the ModelFinder estimate, and (2) our origin prior followed a uniform distribution and was estimated with a starting value of 167 (1,850 CE) with lower and upper boundaries of 145 (1,872) and Infinity respectively based on historical reviews. Indeed, the earliest record of a mild form of a non-zoonotic equine influenza lacking abortions, thus resembling EHV-4, is the American equine epidemic of 1872–1873 (Williams 1924). Hence, this outbreak served as a guided hypothesis for our origins prior. ESS was >200 for all parameters with the first 10 per cent iterations disregarded as burn-in using Tracer v.1.7.2 (Rambaut et al. 2018), thus confirming convergence. Maximum clade credibility trees were generated and visualised through BEAST v2.6.7 (Bouckaert et al. 2019) and FigTree v1.4.4, respectively, as described earlier.

Positive selection

Positive selection between our ancient and published modern sequences was tested using the CODEML branch-site model implemented in the PAML v.4.9 package (Yang 2007). The branch-site model compares the branch-site model A with the corresponding null model which fixes $\omega_2 = 1$, and has a likelihood ratio test of 1 degree of freedom (model = 2; NSsites = 2; fix_omega = 0/1; omega =1.5/1). We used our previously generated phylogenetic tree from IQ-Tree based upon our ORF-concatenated alignment comprising the 28 EHV-4 sequences and the EHV-1 reference strain Ab4p as an outgroup to constrain the phylogenetic placement of our ancient sequence. The EHV-1 sequence was then removed, and the tree was saved in a Newick format. We then manually assigned the UR17x29 ancient sequence as a background branch and the modern sequences as foreground branches (from Branch M onwards, Fig. 1B). The codon frequencies were calculated from the average nucleotide frequencies. Testing for positive selection in subclades I and II followed the same protocol as above: positive selection in subclade I was tested by assigning the UR17x29 sequence and subclade II sequences as background branches and subclade I sequences as foreground branches, and vice-versa for positive selection in subclade II. We used the χ_1^2 with critical value 3.84 (df = 1). To investigate whether episodic diversifying selection occurred, we ran a HyPhy MEME (Mixed Effects Model of Evolution) analysis (Murrell et al. 2012; Kosakovsky Pond et al. 2020) on our individual EHV-4 ORF alignments ($P < 0.1$) available through the datamonkey web server (www.datamonkey.org/meme; (Weaver et al. 2018)).

Protein structure prediction

Protein structure prediction for the eight versions of tegument protein UL14 observed in modern strains and UR17x29 was performed using I-TASSER (Yang et al. 2015). Amino acid sequences representing UL14 from ancient and modern populations were submitted to I-TASSER for automated protein structure prediction using default parameters. Unassigned nucleotides from the ancient genome were assumed to have the same nucleotides as the reference genome for I-TASSER submission. I-TASSER generates full-length atomic structural models from multiple threading alignments and predicts the functions from known protein function databases based on sequence and structure profile comparisons. I-TASSER generated five models, out of which the best model with highest C-score was chosen for structural analysis. Pdb files were visualised using RCSB PDB Mol* viewer (Sehnal et al. 2021). Lack of protein homolog structures of UL14 confirmed by X-ray crystallography or otherwise from other equine herpesviruses, makes it impossible to go beyond primary structure prediction.

Data availability

The data are available on NCBI under BioSample: SAMEA9533417 with further details available at https://doi.org/10.1038/s41586-021-04018-9.

Supplementary data

Supplementary data is available at *VEVOLU* Journal online.

Funding

This project has received funding from the Centre National de la Recherche Scientifique (CNRS) and University Paul Sabatier (AnimalFarm IRP), the European Research Council (ERC) under the

European Union's Horizon 2020 research and innovation program (grant agreement 681605-PEGASUS) and the European Research Council (ERC) through the Horsepower Synergy grant (Agreement no. 101071707). O.L. and K.M. were supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreements no. 895107 and no. 897931, respectively.

Conflict of interest: None declared.

References

Allen, G. P., and Bryans, J. T. (1986) 'Molecular Epizootiology, Pathogenesis, and Prophylaxis of Equine Herpesvirus-1 Infections', *Progress in Veterinary Microbiology and Immunology*, 2: 78–144.

Allentoft, M. E. et al. (2015) 'Population Genomics of Bronze Age Eurasia', *Nature*, 522: 167–72.

Anthony, D. W. (2007) *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton University Press: Princeton, NJ, USA.

——— (2009) 'The Sintashta Genesis: The Roles of Climate Change, Warfare, and Long-Distance Trade', in Hanks, B. K., and Linduff, K. M. (eds) *Social Complexity in Prehistoric Eurasia. Monuments, Metals, and Mobility*. New York, NY, USA: Cambridge University Press, pp. 47–73

Barido-Sottani, J. et al. (2018) 'Taming the BEAST – A Community Teaching Material Resource for BEAST 2', *Systematic Biology*, 67: 170–4.

Borchers, K., Wolfinger, U., and Ludwig, H. (1999) 'Latency-Associated Transcripts of Equine Herpesvirus Type 4 in Trigeminal Ganglia of Naturally Infected Horses', *The Journal of General Virology*, 80: 2165–71.

Bos, K. I. et al. (2014) 'Pre-Columbian Mycobacterial Genomes Reveal Seals as a Source of New World Human Tuberculosis', *Nature*, 514: 494–7.

Bouckaert, R. et al. (2019) 'BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis', *PLoS Computational Biology*, 15: e1006650.

Bruen, T. C., Philippe, H., and Bryant, D. (2006) 'A Simple and Robust Statistical Test for Detecting the Presence of Recombination', *Genetics*, 172: 2665–81.

Bryans, J. T., and Allen, G. P. (1989) 'Herpesviral Diseases of the Horse', in Wittmann, G. (ed.), *Herpesvirus Diseases of Cattle, Horses, and Pigs*. Springer US: Boston, MA, pp. 176–229.

Burrows, R., and Goodridge, D.. 1974. "In Vivo and in Vitro Studies of Equine Rhinopneumonitis Virus Strains". In Karger S. edited by, *Proceedings of the Third International Conference on Equine Infectious Diseases* Paris July 17th–21st, 1972, pp. 306–21.

Chandler, D. G. (1963) 'From the Other Side of the Hill Blenheim, 1704', *Journal of the Society for Army Historical Research. Society for Army Historical Research*, 41: 79–93.

Crabb, B. S., and Studdert, M. J. (1995) 'Equine Herpesviruses 4 (Equine Rhinopneumonitis Virus) and 1 (Equine Abortion Virus)', *Advances in Virus Research*, 45: 153–90.

Cunningham, C. et al. (2000) 'Herpes Simplex Virus Type 1 Gene UL14: Phenotype of a Null Mutant and Identification of the Encoded Protein', *Journal of Virology*., 74: 33–41.

Danecek, P. et al. (2021) 'Twelve Years of SAMtools and BCFtools', *GigaScience*, 10: giab008.

De Martino, L. et al. (2007) 'Antiapoptotic Activity of Bovine Herpesvirus Type-1 (BHV-1) UL14 Protein', *Veterinary Microbiology*, 123: 210–6.

Dimock, W. W., and Edwards, P. R. (1933) 'Is There a Filterable Virus of Abortion in Mares?', *Kentucky Agricultural Experiment Station Bulletin*, 333: 297–301.

——— (1936) 'The Differential Diagnosis of Equine Abortion with Special Reference to a Hitherto Undescribed Form of Epizootic Abortion of Mares', *The Cornell Veterinarian*, 26: 231–40.

Dimopoulos, E. A. et al. (2022) 'HAYSTAC: A Bayesian Framework for Robust and Rapid Species Identification in High-Throughput Sequencing Data', *PLOS Computational Biology*, 18: e1010493.

Drews, R. (1993) *The End of the Bronze Age*. Princeton, NJ, USA: Princeton University Press.

Duchêne, S., and Duchêne, D. A. (2020) 'Estimating Evolutionary Rates and Timescales from Time-Stamped Data', in Ho, S. Y. W. (ed.), *The Molecular Evolutionary Clock: Theory and Practice*. Switzerland: Springer Cham, pp. 157–74.

Duggan, A. T. et al. (2016) '17th Century Variola Virus Reveals the Recent History of Smallpox', *Current Biology: CB*, 26: 3407–12.

Düx, A. et al. (2020) 'Measles Virus and Rinderpest Virus Divergence Dated to the Sixth Century BCE', *Science*, 368: 1367–70.

Duxbury, A. E., and Oxer, D. T. (1968) 'Isolation of Equine Rhinopneumonitis Virus from an Epidemic of Acute Respiratory Disease in Horses', *Australian Veterinary Journal*, 44: 58–63.

Dvorak, G. D., and Spickler, A. R. (2008) 'Glanders', *Journal of the American Veterinary Medical Association*, 233: 570–7.

Eaton, K. (2020) 'NCBImeta: Efficient and Comprehensive Metadata Retrieval from NCBI Databases', *Journal of Open Source Software*, 5: 1990.

Fages, A. et al. (2019) 'Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series', *Cell*, 177: 1419–35.e31.

Frantz, L. A. F. et al. (2020) 'Animal Domestication in the Era of Ancient Genomics', *Nature Reviews Genetics*, 21: 449–60.

Gaunitz, C. et al. (2018) 'Ancient Genomes Revisit the Ancestry of Domestic and Przewalski's Horses', *Science*, 360: 111–4.

Gibbs, M. J., Armstrong, J. S., and Gibbs, A. J. (2000) 'Sister-Scanning: A Monte Carlo Procedure for Assessing Signals in Recombinant Sequences', *Bioinformatics*, 16: 573–82.

Gilkerson, J. R. et al. (1999) 'Epidemiology of EHV-1 and EHV-4 in the Mare and Foal Populations on a Hunter Valley Stud Farm: Are Mares the Source of EHV-1 for Unweaned Foals', *Veterinary Microbiology*, 68: 27–34.

Gouy, M., Guindon, S., and Gascuel, O. (2010) 'SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building', *Molecular Biology and Evolution*, 27: 221–4.

Guellil, M. et al. (2022) 'Ancient Herpes Simplex 1 Genomes Reveal Recent Viral Structure in Eurasia', *Science Advances*, 8: eabo4435.

Hanks, B. et al. (2018) 'Bronze Age Diet and Economy: New Stable Isotope Data from the Central Eurasian Steppes (2100-1700 BC)', *Journal of Archaeological Science*, 97: 14–25.

Huson, D. H., and Bryant, D. (2006) 'Application of Phylogenetic Networks in Evolutionary Studies', *Molecular Biology and Evolution*, 23: 254–67.

Izume, S. et al. (2017) 'The Full Genome Sequences of 8 Equine Herpesvirus Type 4 Isolates from Horses in Japan', *The Journal of Veterinary Medical Science/the Japanese Society of Veterinary Science*, 79: 206–12.

Jónsson, H. et al. (2013) 'mapDamage2.0: Fast Approximate Bayesian Estimates of Ancient DNA Damage Parameters', *Bioinformatics*, 29: 1682–4.

Kalyaanamoorthy, S. et al. (2017) 'ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates', *Nature Methods*, 14: 587–9.

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

Kay, G. L. et al. (2015) 'Eighteenth-Century Genomes Show that Mixed Infections Were Common at Time of Peak Tuberculosis in Europe', *Nature Communications*, 6: 6717.

Kelekna, P. (2009) *The Horse in Human History*. New York, USA: Cambridge University Press.

Kerner, G. et al. (2021) 'Human Ancient DNA Analyses Reveal the High Burden of Tuberculosis in Europeans over the Last 2,000 Years', *The American Journal of Human Genetics*, 108: 517–24.

Key, F. M. et al. (2020) 'Emergence of Human-Adapted Salmonella Enterica Is Linked to the Neolithization Process', *Nature Ecology and Evolution*, 4: 324–33.

Khattab, O. M. et al. (2022) 'Equine Herpes Virus 4 (EHV4) Investigation in Aborted Egyptian Mares; Molecular Detection, Isolation, and Phylogeny for Viral Glycoprotein B', *Advances in Animal and Veterinary Sciences*, 10: 1907–15.

Khusro, A. et al. (2020) 'Equine Herpesvirus-I Infection in Horses: Recent Updates on Its Pathogenicity, Vaccination, and Preventive Management Strategies', *Journal of Equine Veterinary Science*, 87: 102923.

Klecel, W., and Martyniuk, E. (2021) 'From the Eurasian Steppes to the Roman Circuses: A Review of Early Development of Horse Breeding and Management', *Animals*, 11: 1859.

Kocher, A. et al. (2021) 'Ten Millennia of Hepatitis B Virus Evolution', *Science*, 374: 182–8.

Kolb, A. W. et al. (2017) 'Phylogenetic and Recombination Analysis of the Herpesvirus Genus Varicellovirus', *BMC Genomics*, 18: 887.

Kosakovsky Pond, S. L. et al. (2020) 'HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies', *Molecular Biology and Evolution*, 37: 295–9.

Lam, H. M., Ratmann, O., and Boni, M. F. (2018) 'Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm', *Molecular Biology and Evolution*, 35: 247–51.

Larsson, A. (2014) 'AliView: A Fast and Lightweight Alignment Viewer and Editor for Large Datasets', *Bioinformatics*, 30: 3276–8.

Lazaridis, I. et al. (2022) 'The Genetic History of the Southern Arc: A Bridge between West Asia and Europe', *Science (New York, N.Y.)*, 377: eabm4247.

Letunic, I., and Bork, P. (2021) 'Interactive Tree of Life (Itol) V5: An Online Tool for Phylogenetic Tree Display and Annotation', *Nucleic Acids Research.*, 49: W293–96.

Li, H. et al. and 1000 Genome Project Data Processing Subgroup (2009) 'The Sequence Alignment/Map Format and SAMtools', *Bioinformatics*, 25: 2078–9.

Librado, P. et al. (2021) 'The Origins and Spread of Domestic Horses from the Western Eurasian Steppes', *Nature*, 598: 634–40.

Li, H., and Durbin, R. (2009) 'Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform', *Bioinformatics*, 25: 1754–60.

Lunn, D. P. et al. (2009) 'Equine Herpesvirus-1 Consensus Statement', *Journal of Veterinary Internal Medicine/American College of Veterinary Internal Medicine*, 23: 450–61.

Ma, G., Azab, W., and Osterrieder, N. (2013) 'Equine Herpesviruses Type 1 (EHV-1) and 4 (Ehv-4)–masters of Co-Evolution and a Constant Threat to Equids and Beyond', *Veterinary Microbiology*, 167: 123–34.

Macfadden, B. J. (2005) 'Evolution. Fossil Horses–Evidence for Evolution', *Science*, 307: 1728–30.

Manninger, R., and Csontos, J. (1941) 'Virusabortus Der Stuten', *Dtsch Tierartztl Wschr*, 49: 105–11.

Martin, D. P. et al. (2021) 'RDP5: A Computer Program for Analyzing Recombination in, and Removing Signals of Recombination From, Nucleotide Sequence Datasets', *Virus Evolution*, 7: veaa087.

——— et al. (2005) 'A Modified Bootscan Algorithm for Automated Identification of Recombinant Sequences and Recombination Breakpoints', *AIDS Research and Human Retroviruses*, 21: 98–102.

Martin, D., and Rybicki, E. (2000) 'RDP: Detection of Recombination Amongst Aligned Sequences', *Bioinformatics*, 16: 562–3.

Matsumura, T. et al. (1992) 'Epizootiological Aspects of Type 1 and Type 4 Equine Herpesvirus Infections among Horse Populations', *The Journal of Veterinary Medical Science/the Japanese Society of Veterinary Science*, 54: 207–11.

Maturana-Russel, P. et al. (2019) 'Model Selection and Parameter Inference in Phylogenetics Using Nested Sampling', *Systematic Biology*, 68: 219–33.

Maynard Smith, J. (1992) 'Analyzing the Mosaic Structure of Genes', *Journal of Molecular Evolution*, 34: 126–9.

Meade, B. J. (2012) 'The Transmission Dynamics of Equine Herpesvirus Type 1 (EHV-1) Infection in Outbreaks Characterised Predominately by Neurologic or Respiratory Illness'. University of Kentucky. PhD thesis.

Mekonnen, A., Eshetu, A., and Gizaw, D. (2017) 'Equine Herpesvirus 1 And/or 4 in Working Equids: Seroprevalence and Risk Factors in North Shewa Zone, Ethiopia', *Ethiopian Veterinary Journal*, 21: 28–39.

Minh, B. Q. et al. (2020) 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Molecular Biology and Evolution*, 37: 1530–4.

Morand, S., McIntyre, K. M., and Baylis, M. (2014) 'Domesticated Animals and Human Infectious Diseases of Zoonotic Origins: Domestication Time Matters', *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 24: 76–81.

Moustafa, A. et al. (2017) 'The Blood DNA Virome in 8,000 Humans', *PLOS Pathogens.*, 13: e1006292.

Murrell, B. et al. (2012) 'Detecting Individual Sites Subject to Episodic Diversifying Selection', *PLoS Genetics.*, 8: e1002764.

Narasimhan, V. M. et al. (2019) 'The Formation of Human Populations in South and Central Asia', *Science*, 365: eaat7487.

Nielsen, R. et al. (2017) 'Tracing the Peopling of the World through Genomics', *Nature*, 541: 302–10.

EFSA Panel on Animal Health and Welfare, (AHAW) et al. (2022) 'Assessment of Listing and Categorisation of Animal Diseases within the Framework of the Animal Health Law (Regulation (EU) No 2016/429): Infection with Equine Herpesvirus-1', *EFSA Journal*, 20: e07036.

Oda, S. et al. (2016) 'The Interaction between Herpes Simplex Virus 1 Tegument Proteins UL51 and UL14 and Its Role in Virion Morphogenesis', *Journal of Virology.*, 90: 8754–67.

Orlando, L. et al. (2021) 'Ancient DNA Analysis', *Nature Reviews Methods Primers*, 1: 1–26.

——— et al. (2013) 'Recalibrating Equus Evolution Using the Genome Sequence of an Early Middle Pleistocene Horse', *Nature*, 499: 74–8.

Osterrieder, N., and Van de Walle, G. R. (2010) 'Pathogenic Potential of Equine Alphaherpesviruses: The Importance of the Mononuclear Cell Compartment in Disease Outcome', *Veterinary Microbiology*, 143: 21–8.

Padidam, M., Sawyer, S., and Fauquet, C. M. (1999) 'Possible Emergence of New Geminiviruses by Frequent Recombination', *Virology*, 265: 218–25.

Paradis, E., and Schliep, K. (2019) 'Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R', *Bioinformatics*, 35: 526–8.

Patel, J. R., and Heldens, J. (2005) 'Equine Herpesviruses 1 (EHV-1) and 4 (Ehv-4)–epidemiology, Disease and Immunoprophylaxis: A Brief Review', *The Veterinary Journal*, 170: 14–23.

Pavulraj, S. et al. (2021) 'Equine Herpesvirus Type 4 (EHV-4) Outbreak in Germany: Virological, Serological, and Molecular Investigations', *Pathogens*, 10: 810.

Pfrengle, S. et al. (2021) 'Mycobacterium Leprae Diversity and Population Dynamics in Medieval Europe from Novel Ancient Genomes', *BMC Biology*, 19: 220.

Posada, D., and Crandall, K. A. (2001) 'Evaluation of Methods for Detecting Recombination from DNA Sequences: Computer Simulations', *Proceedings of the National Academy of Sciences of the United States of America.*, 98: 13757–62.

Pusterla, N. et al. (2005) 'Equine Herpesvirus-4 Kinetics in Peripheral Blood Leukocytes and Nasopharyngeal Secretions in Foals Using Quantitative Real-Time TaqMan PCR', *Journal of Veterinary Diagnostic Investigation: Official Publication of the American Association of Veterinary Laboratory Diagnosticians, Inc*, 17: 578–81.

Rambaut, A. et al. (2018) 'Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7', *Systematic Biology*, 67: 901–4.

Rascovan, N. et al. (2019) 'Emergence and Spread of Basal Lineages of Yersinia Pestis during the Neolithic Decline', *Cell*, 176: 295–305.e10.

Reimer, P. J. et al. (2020) 'The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 Cal kBP)', *Radiocarbon*, 62: 725–57.

Rohland, N. et al. (2015) 'Partial Uracil–DNA–Glycosylase Treatment for Screening of Ancient DNA', *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, 370: 20130624.

Roman-Binois, A. (2017) 'L'archéologie Des épizooties: Mise En évidence et Diagnostic Des Crises de Mortalité Chez Les Animaux D'élevage, Du Néolithique à Pasteur', Université Panthéon-Paris-Sorbonne I. PhD thesis.

Sanjuán, R. et al. (2010) 'Viral Mutation Rates', *Journal of Virology*, 84: 9733–48.

Schubert, M. et al. (2014) 'Characterization of Ancient and Modern Genomes by SNP Detection and Phylogenomic and Metagenomic Analysis Using PALEOMIX', *Nature Protocols*, 9: 1056–82.

Schuenemann, V. J. et al. (2018) 'Ancient Genomes Reveal a High Diversity of Mycobacterium Leprae in Medieval Europe', *PLOS Pathogens*, 14: e1006997.

——— et al. (2013) 'Genome-Wide Comparison of Medieval and Modern Mycobacterium Leprae', *Science*, 341: 179–83.

Sehnal, D. et al. (2021) 'Mol* Viewer: Modern Web App for 3D Visualization and Analysis of Large Biomolecular Structures', *Nucleic Acids Research*, 49: W431–37.

Sharrer, G. T. (1995) 'The Great Glanders Epizootic, 1861-1866: A Civil War Legacy', *Agricultural History*, 69: 79–97.

Skilling, J. (2006) 'Nested Sampling for General Bayesian Computation', *Bayesian Analysis*, 1: 833–60.

Spyrou, M. A. et al. (2019a) 'Ancient Pathogen Genomics as an Emerging Tool for Infectious Disease Research', *Nature Reviews Genetics*, 20: 323–40.

——— et al. (2022) 'The Source of the Black Death in Fourteenth-Century Central Eurasia', *Nature*, 606: 718–24.

——— et al. (2019b) 'Phylogeography of the Second Plague Pandemic Revealed through Analysis of Historical Yersinia Pestis Genomes', *Nature Communications*, 10: 4470.

Stasiak, K., Dunowska, M., and Rola, J. (2018) 'Prevalence and Sequence Analysis of Equid Herpesviruses from the Respiratory Tract of Polish Horses', *Virology Journal*, 15: 106.

Studdert, M. J. et al. (2003) 'Outbreak of Equine Herpesvirus Type 1 Myeloencephalitis: New Insights from Virus Identification by PCR and the Application of an EHV-1-specific Antibody Detection ELISA', *Veterinary Record*, 153: 417–23.

Taubenberger, J. K. et al. (2005) 'Characterization of the 1918 Influenza Virus Polymerase Genes', *Nature*, 437: 889–93.

Telford, E. A. et al. (1992) 'The DNA Sequence of Equine Herpesvirus-1', *Virology*, 189: 304–16.

——— et al. (1998) 'The DNA Sequence of Equine Herpesvirus-4', *The Journal of General Virology*, 79: 1197–203.

Valtueña, A. A. et al. (2022) 'Stone Age *Yersinia Pestis* Genomes Shed Light on the Early Evolution, Diversity, and Ecology of Plague', *Proceedings of the National Academy of Sciences*, 119: e2116722119.

Vandekerckhove, A. P. et al. (2011) 'Equine Alphaherpesviruses (EHV-1 and EHV-4) Differ in Their Efficiency to Infect Mononuclear Cells during Early Steps of Infection in Nasal Mucosal Explants', *Veterinary Microbiology*, 152: 21–8.

Vargas-Bermudez, D. et al. (2018) 'Detección de Los Herpesvirus Equinos 1 Y 4 Y Su Relación Con Transcriptos Asociados a la Latencia En Caballos Infectados Naturalmente En Colombia', *Revista MVZ Córdoba*, 23: 6826–37.

Vaz, P. K. et al. (2016) 'Evidence of Widespread Natural Recombination among Field Isolates of Equine Herpesvirus 4 but Not among Field Isolates of Equine Herpesvirus 1', *The Journal of General Virology*, 97: 747–55.

Vissani, M. A., Damiani, A. M., and Barrandeguy, M. E. (2021) 'Equine Coital Exanthema: New Insights on the Knowledge and Leading Perspectives for Treatment and Prevention', *Pathogens*, 10: 1055.

Weaver, S. et al. (2018) 'Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes', *Molecular Biology and Evolution*, 35: 773–7.

White, T. A., and Perkins, S. E. (2012) 'The Ecoimmunology of Invasive Species', *Functional Ecology*, 26: 1313–23.

Williams, A. J. (1924) 'Analogies between Influenza of Horses and Influenza of Man', *Proceedings of the Royal Society of Medicine*, 17: 47–58.

Wood, D. E., Lu, J., and Langmead, B. (2019) 'Improved Metagenomic Analysis with Kraken 2', *Genome Biology*, 20: 257.

Yamauchi, Y. et al. (2008) 'The UL14 Tegument Protein of Herpes Simplex Virus Type 1 Is Required for Efficient Nuclear Transport of the Alpha Transinducing Factor VP16 and Viral Capsids', *Journal of Virology*, 82: 1094–106.

——— et al. (2002) 'Herpes Simplex Virus Type 2 UL14 Gene Product Has Heat Shock Protein (Hsp)-like Functions', *Journal of Cell Science.*, 115: 2517–27.

Yang, Z. (2007) 'PAML 4: Phylogenetic Analysis by Maximum Likelihood', *Molecular Biology and Evolution*, 24: 1586–91.

Yang, J. et al. (2015) 'The I-TASSER Suite: Protein Structure and Function Prediction', *Nature Methods*, 12: 7–8.

Yeargan, M. R., Allen, G. P., and Bryans, J. T. (1985) 'Rapid Subtyping of Equine Herpesvirus 1 with Monoclonal Antibodies', *Journal of Clinical Microbiology*, 21: 694–7.

Virus Evolution, 2024, **10(1)**, 1–12
DOI: https://doi.org/10.1093/ve/vead087
Advance Access Publication 12 January 2024
Research Article