UNIVERSITY OF CALIFORNIA

SANTA CRUZ

EVOLUTIONARY HISTORY OF EARLY-MIDDLE AND LATE PLEISTOCENE EQUIDS, REVEALED BY ANALYSIS OF THEIR PALEOGENOMES

A dissertation submitted in partial satisfaction

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECOLOGY AND EVOLUTIONARY BIOLOGY

by

Alisa Vershinina

June 2020

The Dissertation of Alisa Vershinina is approved:

_____
Professor Beth Shapiro, Chair

_____
Professor Gant Pogson

_____
Professor Paul L. Koch

_____
Professor Russell Corbett-Detig

_____
Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

ABSTRACT

Evolutionary history of Early-Middle and Late Pleistocene equids, revealed by analysis of their paleogenomes

By

Alisa O. Vershinina

DNA from archaeological, paleontological, and museum samples (ancient, or aDNA) provides a unique opportunity to trace eco-evolutionary history of populations affected by environmental shifts on geological time scale. Yet it is still unclear how climate-driven environmental change and biogeographical barriers affect diversification, population size and population structure of large-bodied herbivores inhabiting northern regions of the Northern Hemisphere. The goal of this dissertation is to fill in this gap by utilizing ancient DNA techniques and population genetic analysis to reveal the demographic and population history of extinct and present-day equids, genus *Equus*, focusing on their key ancient dispersal corridor - the Bering Land Bridge. In the following chapters, I explore the links between paleoenvironments and population history of various equid groups using high coverage paleogenomes recovered from fossil horse specimens sampled across Beringia. In my first chapter, I use in-solution DNA capture enrichment and mitochondrial genome assembly to reconstruct a whole mitochondrial genome of a specimen found in Western Beringia and initially identified as *E. hemionus*, or an Asiatic wild ass. With molecular phylogenetic analysis I demonstrate that the specimen belongs to a group of caballoid horses, *E. ferus*, rather than stenonid wild asses. The results obtained in Chapter 1

highlight the utility of ancient DNA studies in identification of incomplete, juvenile, or otherwise problematic museum specimens. In my second chapter I discover that Beringia was a key contact zone for populations of Late Pleistocene caballoid horses, *E. ferus*. I use new high coverage nuclear and mitochondrial paleogenomes, isolated from fossils of caballoid horses sampled across the Northern Hemisphere to infer that North American and Eurasian caballoid horse populations diverged around 0.8-1 million years ago. With coalescent simulations and genome-wide adamixtude inference I show that evolution of caballoid horses after this divergence continued in the presence of cross-continental gene flow. My demographic inference suggests that disappearance of the Bering Land Bridge likely exacerbated an already ongoing extinction of Beringian caballoid horse populations. In the third chapter, I recover the ~700,000 year old paleogenome of a previously unknown stenonid horse species inhabiting Klondike, Canada's Yukon Territory - the oldest non-caballoid equid genome known to date. Using genotype likelihood approach on a dataset of present-day and ancient equid nuclear genomes, I show that the population of the newly discovered stenonid equid species was evolutionary close to the present-day zebras and Asiatic wild asses. I suggest that the new to genetics species likely represents an extinct branch of archaic stenonid ungulates that coexisted with "true", or caballoid equids in the Early-Middle Pleistocene Yukon. In the fourth chapter I expand my study system to another iconic Bringian megafauna species - steppe bison, *Bison priscus.* Using molecular phylogeny reconstructed from new high coverage mitochondrial genomes, I explore the phylogenetic diversity of steppe bison in Western Beringia. I confirm the

existence of the deeply divergent steppe bison clade and shed new light on the evolutionary history of bison during the Pleistocene to Holocene transition in ancient Siberia.

ACKNOWLEDGMENTS

I would like to thank my PhD adviser Beth Shapiro. Her guidance and mentorship have been essential to my upbringing as a scientist and a female academic. Beth taught me how to do things and how to do them right: how to see the big picture, focus on research goals, maintain and broaden scientific collaborations, and be on track with the planning of experiments and projects. Beth was generous in supporting (financially and morally) my international student status, which I am endlessly thankful for. I have been privileged to enjoy our collaboration and I hope it continues.

I am extremely grateful to my dissertation committee, Russ Corbett-Detig, Grant Pogson, and Paul Koch for fruitful discussions and insights, without which the completion of this dissertation would not be possible. I also thank Ed Green, who taught me how to be a creative thinker and also that making good graphs is not a waste of time.

An amazing network of collaborators, museum curators, and colleagues have been essential in accomplishing my research. In particular, I would like to acknowledge Ross MacPhee for encouraging me to face the wild world of wild horse conservation. I have never left our meetings without a boost of inspiration and motivation to excel in the scientific field and beyond. Ludovic Orlando has kindly provided sequencing data for several horse genomes, I am very thankful to him and his research team for making this data available for my project. I also would like to thank Pete Heinzman and Grant Zazula, whose energetic mentorship and immense knowledge guided me through the exciting field of equid evolutionary biology. I am privileged to work with Pamela

Groves, Irina Kirillova, Britta Jensen, Neda DeMayo and the Return to Freedom Foundation, Manda Kailmian and the CANA Foundation, Eric Scott, Daniel Mann, Duane Froese, Mathias Stiller, Dan Chang, Elizabeth Hall, Love Dalen, John Southon, Gennady Baryshnikov, Haowen Tong, and Mikhail Tiunov. Thank you for providing crucial assistance with all aspects of this dissertation, none of my research would be possible without your help.

Sample collection in the Yukon region was made possible thanks to the Klondike placer gold mining community, the Tr'ondëk Hwëch'in First Nation, and the Vuntut Gwitchin First Nation. I am grateful for their partnership and support for our ongoing research in the Canadian Arctic. Financial support of my dissertation was provided by NSF Office of Polar Programs Award #1417036, a UC Santa Cruz Chancellor Dissertation Year Fellowship, and generous anonymous donors. I am beyond grateful for the opportunity provided by their financial help, which was essential to accomplishing my research.

I am grateful for all the help the Paleogenomics lab tech crew provided in data generation for the project. Special thank you to Josh Kapp for teaching me everything about ancient DNA wizardry and for being a great truck companion in the Arctic. Molly Cassatt-Johnstone and Shelby Dunn, thank you for always having a time slot available for my horses. I am grateful to my ex-mentee now-friend Christian Lowson for helping with The Bison.

Shapiro\Green\Fehren-Schmitz lab became my foster family abroad. Sabrina, thank you for being a constant source of inspiration and helping me grow my witch

energy. Thank you, Jannine, for putting up with my “what am I doing with my life” rants, your friendship means so much to me. No words are enough to express how thankful I am to all my lab sisters, Neve, Chloe, Kim, Molly, Shelby, Megan, Katie, Merely, Nedda, Paloma, Kalina, and Naz, for being so supportive during my PhD journey. Thank you to all my present and past labmates for fruitful discussions, staircase lunches, and good laughs. I deeply appreciate my grad school cohort (a.k.a. “The Lost Cohort”). Thank you to the amazing UCSC EEB department for providing a true sense of community. Special thanks to the WiSE and the EEB peer-to-peer mentorship programs that I had a delightful time participating in.

My undergraduate adviser Vladimir Lukhtanov took me under his wing in the rocky world of Russian entomology and gave me the courage to embark on my PhD journey. I am extremely grateful for his support along the way.

My friends and family have been essential in helping me to follow my dream of becoming a scientist and to pursue a career abroad. To my parents, Oleg and Viktoriia Vershinin’y, and grandparents, Albina and Oleg Shutt, thank you for your support and encouragement that always keeps me moving forward, I would never have made it through without you. My dearest friends Lera Lerner, Natalya Pashkovskaya, and Anastasiya Kotyleva helped me get where I am today through the long St-Petersburg nights. Thank you for always being there for me.

Last but not least, I would like to acknowledge my partner Alex. Thank you for your kind care, warmth, and love that brings so much happiness and balance to my life. I look forward to continuing our journey together.

The text of this dissertation includes reprints of the following previously published material: (Chapter 1) Vershinina, A. O., Kapp, J. D., Baryshnikov, G. F., & Shapiro, B. (2019). The case of an arctic wild ass highlights the utility of ancient DNA for validating problematic identifications in museum collections. *Molecular Ecology Resources,* https://doi.org/10.1111/1755-0998.13130; (Chapter 4) Vershinina, A. O., Kapp, J. D., Rodrigues Soares, A. E., Heintzman, P. D., Lowson, C., Cassatt-Johnstone, M., Shidlovskiy, F. K., Kirillova, I.V., & Shapiro, B. (2019). Ancient DNA analysis of a Holocene bison from the Rauchua River, North Western Chukotka, and the existence of a deeply divergent mitochondrial clade *Zoologicheskiĭ zhurnal,* 98(10): 1091-1099. Beth Shapiro, the co-author listed on these publications, directed and supervised the research which forms the basis for the dissertation.

Statement of contribution. The published studies presented in Chapter 1 and Chapter 4 were conceived by A. Vershinina together with coauthors. A. Vershinina obtained the samples, extracted DNA from the samples, conducted analyses of the sequencing data, and wrote the initial draft of both manuscripts. All authors discussed the results and contributed to the final version of the manuscripts.

GENERAL INTRODUCTION

Insights from the ecosystems of the past are essential to the understanding of the present-day and future patterns of biodiversity change. Paleontology and paleoenvironmental science are the conventional approaches proven to be essential in revealing these insights. While being valuable tools, the conventional methods are unable to answer questions that are critical in the time of climate emergency and the Anthropocene mass extinction, when conservation efforts increasingly focus on implementing genomic approaches to policymaking (Supple & Shapiro, 2018). What is the genetic context of extinction in a broad spatial and temporal context? How do environmental change and population fragmentation impact the history of the populations undergoing decline and what signals do these factors leave in genomes of endangered species? Finally, do factors maintaining genetic diversity such as admixture and connectivity between populations mitigate extinction risk? It is possible to shed light on the outlined questions by a synthesis between conventional paleontological approaches and paleogenomic methods that allow us to precisely track the evolutionary history of extinct and surviving populations with DNA recovered from fossil, museum, and archival specimens (ancient DNA, aDNA). The goal of my dissertation is to understand the role of biogeographical barriers and large-scale ecosystem changes in the evolution and diversification of large-bodied mammals by synthesizing classical paleontology with a paleogenomic approach. I focus on extinct and present-day horses (genus *Equus*) - members of the iconic megafauna community (animals weighing >45kg), that formerly were abundant in the Pleistocene Holarctic,

but experienced a dramatic decline in the last 50 thousand years narrowly surviving the environmental shift of the Pleistocene-Holocene transition 12-10 thousand years ago (kya) (Barnosky et al., 2004). As a study system, I use Pleistocene Beringia - a now closed dispersal corridor between Eurasia and North America. During major glaciations, Beringia exposed a Bering Land Bridge that connected Eurasia with North America and served as a biogeographical dispersal corridor for equids and other Arctic megafauna species, such as bison and mammoth (Elias & Crocker, 2008; Mann et al., 2015). Furthermore, the vegetation of the Bering Land Bridge cycled through grassland and forest or tundra disrupting the spatial continuity of the northern steppe belt during glacial/interglacial transitions (Mann et al., 2018). I use aDNA methods to recover nuclear and mitochondrial genomes of fossil Middle and Late Pleistocene horses inhabiting Beringia. I use paleogenomics to discover a new equid species that was present in the Klondike. Using the newly-recovered and previously published high-coverage paleogenomes I utilize coalescent modeling and phylogenetic reconstruction to infer population history of the genus *Equus* in the last 5 million years. I discuss the results of the population genetic analysis within the framework of current knowledge about Pleistocene paleoenvironments.

Present-day members of the horse family, Equidae belong to the genus *Equus* (Wilson and Reeder, 2005), the earliest members of which diverged from other equids in North America 4-5 Myr ago (MacFadden, 2005). Present-day equids are split into two major clades: the "caballoid" group (*E. ferus*, includes the domestic and Przewalski's horse*s*), and the stenonid group, unifying all other recent members of the

genus (wild asses, donkeys, kiangs, and zebras) (Heintzman et al., 2017; MacFadden, 1994; Orlando, 2015; Vilstrup et al., 2013). Small populations of feral caballoid horses persist in Central Asia (Przewalski's horse) and the American West (mustangs, originating from reintroduced domestic horses) as well as elsewhere, but all appear to be derived from lineages domesticated during the Holocene (Fages et al., 2019).

Caballoid horses trace their evolutionary history to North America, where they were a widespread group, occupying many parts of the New World during the Pleistocene (MacFadden, 2005). Soon after their divergence, caballoids dispersed to adjacent Eurasia across the Bering Land Bridge and to South America across the Isthmus of Panama (MacFadden, 1994). Fossil evidence indicates that while horse populations were markedly successful on all three continents, their populations declined to the point of extinction concurrently with Pleistocene-Holocene transition (12-10 kyr bp), when all New World equid populations (including caballoid *Equus*, *Hippidion*, and potentially *Haringtonhippus*) completely collapsed (Heintzman et al., 2017). By contrast, Eurasian equids survived although extinction occurred in at least one non-caballoid species, *E. ovodovi*, during the transition or thereafter (Yuan et al., 2019), at the same time as caballoid horses experienced a sharp decline before domestication around 5.5 kyr ago (Fages et al., 2019). It is debated whether human over-hunting, climatically driven environmental change, cosmic impact, or a combination of these factors forced the North American extinction of equids (Guthrie, 2006; Koch & Barnosky, 2006; Mann et al., 2018; Moore et al., 2020).

Horse species provide an ideal study system for exploring the interplay between population history and environmental perturbations for several reasons. First, equids are large-bodied mammals that have long generation time (5-8 years) and low fecundity and thus have limited adaptability to rapid shifts in selective pressure (Baskin & Danell, 2003; Rosenheim & Tabashnik, 1991). Formerly abundant in the Pleistocene Holarctic, caballoid, or "true horses", narrowly survived the environmental shift of the Pleistocene-Holocene transition 12-10 thousand years ago (kya) (Barnosky et al., 2004). Previous work on mitochondrial DNA of extinct caballoid horses revealed that physical and genetic isolation between their populations increased, as habitat patchiness increased over time which preceded complete extirpation of horses in North America and Asia across a formally wide range (Lorenzen et al., 2011). Thus, equids provide a good proxy for forecasting impacts of environmental changes in the present-day deteriorating Arctic ecosystems on other cold-adapted high latitude species that have similar life history traits, such as polar bears, muskox, and caribou (Berteaux et al., 2004).

Dissertation outline

Chapter 1. The case of an arctic wild ass highlights the utility of ancient DNA for validating problematic identifications in museum collections.

Museum collections are an essential tool in reconstructing and understanding past biodiversity (Johnson et al., 2011). However, many specimens are incomplete, which impedes taxonomic identification (Lyman, 2019). Inaccurate identification in

turn can lead to false biogeographic reconstructions, with cascading influence on paleontological and paleoecological research. The situation is particularly challenging in the case of extinct equids, which led to taxonomic oversplitting due to species description based on dubious fossil remains. Incidentally, Beringian equid fossils analysed here have been attributed to a variety of nominal species (*E. ferus*, *E. scotti*, *E. lambei*, *E. lenensis*), but it is widely acknowledged that the species-level systematics of Holarctic caballoids suffers from excessive splitting (Orlando, 2019).

In **Chapter 1** I use ancient DNA and a mitochondrial hybridization capture approach to correct a previous misidentification of a horse mandible in the collection at the Zoological Institute of St. Petersburg, Russia. Based on morphological data, the incomplete specimen that was found on Begichev Islands, Russian Far North, had been identified as a wild ass, *Equus hemionus.* Another stenonid equid, *E. hydruntinus* was identified in the middle latitudes of Eurasia in the Pleistocene, thus identification of the fossil as *E. hemionus* is disputable. Furthermore, the taxonomic identification led to a conclusion that ancestors of the present-day asiatic wild asses inhabited far northern regions of Eurasia during the Late Pleistocene - the territory far outside of the proposed Pleistocene range of *E.hemionus* and *E. hydruntinus*. I show instead that the specimen belongs to the caballoid horses, *E. caballus*, the lineage ancestral to the present-day domestic horse. My study demonstrates the power of ancient DNA research to identify the taxonomic position of incomplete or problematic museum specimens and contributes to the understanding of the evolutionary and biogeographic history of Asiatic wild asses. I discuss the finding in the context of museum science and the utility

of paleogenomics in unlocking the "biomolecular biobanks" of the natural history museums.

Chapter 2. Paleogenomes reveal Pleistocene cross-continental dispersals and admixture in caballine horses.

Regional-scale shifts in habitat productivity, availability, and accessibility have potentially broad evolutionary consequences by both creating and eliminating barriers to dispersal for many species. The impact of climate-driven regional shifts is particularly acute for the Arctic mammals, which are highly vulnerable to the effects of climate-associated fluctuations on habitats (Berteaux et al., 2004). In Chapter **2** I examine the influence of Beringia on the evolution and diversification of caballoid horses, including *Equus ferus,* during the Pleistocene. I am specifically focusing on caballoid horse populations inhabiting the Bering Land Bridge.

The unique Arctic ecosystem of the Bering Land Bridge acted as a biogeographical corridor for dispersal of Old World megafauna species into North America and vice versa during the glaciations when the sea level dropped and exposed a land tract connecting the continents (DeChaine, 2008; Elias & Crocker, 2008; Meiri et al., 2014). The role of Beringia in the evolution of the caballoid horses has never been thoroughly assessed, despite their extremely well-preserved fossil record. To tackle this gap, I analyze a panel of previously published and newly-recovered high coverage nuclear and mitochondrial genomes from Eurasian and North American fossil caballoid horses. The latter are the first high coverage New World caballoid horse genomes of the Late Pleistocene recovered so far. Using coalescent simulations on a

set of neutral regions of the nuclear genomes, I infer that the common ancestors of North American and Eurasian caballoid horses diverged around 1-0.8 million years ago. The population history of caballoid horses after the split on major continental populations continued in the presence of gene flow, as suggested by the signal of migration and admixture in the coalescent analysis and D-statstic estimates. In addition, I reconstruct 79 complete mitochondrial genomes, acquired from processing a collection of >300 Holarctic horse fossil remains. Analysis of mitochondrial genome phylogeny supports that after the divergence 1-0.8 Myr ago, the evolution of horses on both continents continued in the presence of gene flow across the Bering Land Bridge. My results show that the disappearing of the Bering Land Bridge during Pleistocene-Holocene 11 kyr ago ceased gene flow between continental caballoid horse populations. In the discussion section of **Chapter 2** I provide new insights into the critical role Beringia played in shaping the genetic diversity for Holarctic mammals during the Quaternary.

Chapter 3. A high coverage paleogenome of an Early-Middle Pleistocene Equid from northwest Canada.

One of the specimens processed for **Chapter 2**, an equid cranium found in the Paradise Hill gold mine (Klondike, Canada), was associated with a layer of Gold Run volcanic ash dated to ~0.69-0.76 Myr ago (Buryak et al., 2019). When placed on mitochondrial phylogeny reconstructed in Chapter 2, the specimen fell into the clade of stenonid horses close to zebras and Asiatic wild asses with high statistical support. Motivated by the unusual phylogenetic position of the Paradise Hill equid and its

extremely old age, in the **Chapter 3** I take a deeper look at the specimen's taxonomic and phylogenetic position using a comprehensive genomic approach. I extract DNA from the specimen's petrosal part of the temporal bone. High DNA preservation in this part of the cranium allowed me to sequence the nuclear genome of the Paradise Hill equid to 12-fold coverage and mitochondrial genome to 75-fold coverage. I use genotype likelihood approach to compare the nuclear genome of the Paradise Hill with ancient caballoid horse genomes recovered for Chapter 2 as well as with previously published present-day genomes of donkeys, wild asses, and zebras. I augment the nuclear genome analysis with the tip-dated mitochondrial genome phylogeny. Analysis of nuclear polymorphisms and topology of mitochondrial genome phylogeny indicated that Paradise Hill horse belongs to a new to genetics species of stenonid equids, evolutionary distinct from previously-known caballoid equids inhabiting Yukon Klondike. I discuss the results in a context of the evolutionary history of Middle Pleistocene equids and outline future strategies for taxonomic identification of the specimen.

Chapter 4. Ancient DNA analysis of a Holocene bison from the Rauchua river, Northwestern Chukotka, and the existence of a deeply divergent mitochondrial clade.

In order to understand a broad context of genetic diversity and diversification of Beringian megafauna, I analyzed mitochondrial genomes of another representative member of the Pleistocene Bering Land Bridge species community - steppe bison, *Bison priscus*. As in equids, the evolutionary history of several steppe bison populations remains unclear. Of particular interest is the Rauchua River bison lineage

(Kirillova et al., 2015). Similar to extinct caballoid horses from the New Siberian Islands, which according to the **Chapter 2** results represent a population bearing a North American haplotype, Rauchua bison is a case of a population with unclear genetic affinity.

The steppe bison found on the Rauchua River is dated to 9497 ± 92 years before present and is the single specimen representing the deeply-divergent bison clade found in North-Western Chukotka, Russia (Kirillova et al., 2015). When the mitochondrial genome of the Rauchua bison was compared to data generated from a large dataset of extant and extinct *Bison* lineages from across the Northern Hemisphere this sample was clustered outside, but sister to the cluster, representing known steppe bison maternal genetic diversity. Following the bioinformatic pipeline developed in **Chapter 1** and the discovery of hidden mitochondrial genetic diversity of equids as in **Chapter 2**, I applied mitochondrial DNA capture and molecular phylogenetic approach hoping to identify more bison fossils representing the Rauchua clade in order to trace the population history of the Rauchua population.

I had two goals in **Chapter 4**. First, because the Rauchua bison haplotype was so divergent from other bison haplotypes, I aimed to replicate and therefore confirm the results generated previously from the Rauchua bison using additional bone elements that were shipped and processed separately. Second, I aimed to build on the previous results to provide new insights into bison taxonomy and evolution by generating and analyzing DNA from additional ancient bison remains. Specifically, I isolated mitochondrial DNA from other bison samples found in the region of the Rachua River,

as well as from newly collected samples from Chukotka and other locations in the Russian Far East. New mitochondrial genomes captured from previously unprocessed fragments of the original Rauchua specimen confirmed the existence of the deeply divergent Rauchua clade, however, surprisingly, none of the 26 new specimens tested in the study had the Rauchua bison haplotype. Using molecular phylogeny reconstructed from high coverage mitochondrial genomes of these specimens, I explored the phylogenetic diversity of bison in Western Beringia during the last Ice Age and at the beginning of Holocene warming. I discussed the results of the study with a particular focus on the divergent Rauchua lineage and its distribution in the Pleistocene.

BIBLIOGRAPHY

Barnosky, A. D., Koch, P. L., Feranec, R. S., Wing, S. L., & Shabel, A. B. (2004). Assessing the causes of late Pleistocene extinctions on the continents. *Science*, *306*(5693), 70–75.

Baskin, L., & Danell, K. (2003). *Ecology of Ungulates: A Handbook of Species in Eastern Europe and Northern and Central Asia*. Springer, Berlin, Heidelberg.

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., & Courchamp, F. (2012). Impacts of climate change on the future of biodiversity. *Ecology Letters*, *15*(4), 365–377.

Berteaux, D., Réale, D., McAdam, A. G., & Boutin, S. (2004). Keeping pace with fast climate change: can arctic life count on evolution? *Integrative and Comparative Biology*, *44*(2), 140–151.

Buryak, S., Reyes, A., Froese, D., & Jensen, B. (2019). Zircon U-Pb dates for Gold Run tephra confirm the antiquity of ancient permafrost and the world's oldest recovered genome. *20th Congress of the International Union for Quaternary Research (INQUA)*. INQUA, Dublin, Ireland. https://app.oxfordabstracts.com/events/574/program-app/submission/94741

DeChaine, E. G. (2008). A bridge or a barrier? Beringia's influence on the distribution and diversity of tundra plants. *Plant Ecology & Diversity*, *1*(2), 197–207.

Elias, S. A., & Crocker, B. (2008). The Bering Land Bridge: a moisture barrier to the dispersal of steppe–tundra biota? *Quaternary Science Reviews*, *27*(27), 2473–2483.

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K. A. S., Albizuri, S., Alfarhan, A. H., Allentoft, M., Alquraishi, S., Anthony, D., Baimukhanov, N., Barrett, J. H., Bayarsaikhan, J., Benecke, N., Bernáldez-Sánchez, E., … Orlando, L. (2019). Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell*, *177*(6), 1419–1435.e31.

Guthrie, R. D. (2006). New carbon dates link climatic change with human colonization and Pleistocene extinctions. *Nature*, *441*(7090), 207–209.

Heintzman, P. D., Zazula, G. D., MacPhee, R. D. E., Scott, E., Cahill, J. A., McHorse, B. K., Kapp, J. D., Stiller, M., Wooller, M. J., Orlando, L., Southon, J., Froese, D. G., & Shapiro, B. (2017). A new genus of horse from Pleistocene North America. *eLife*, *6*. https://doi.org/10.7554/eLife.29944

Johnson, K. G., Brooks, S. J., Fenberg, P. B., Glover, A. G., James, K. E., Lister, A. M., Michel, E., Spencer, M., Todd, J. A., Valsami-Jones, E., Young, J. R., & Stewart, J. R. (2011). Climate Change and Biosphere Response: Unlocking the Collections Vault. *Bioscience*, *61*(2), 147–153.

Kirillova, I. V., Zanina, O. G., Chernova, O. F., Lapteva, E. G., Trofimova, S. S., Lebedev, V. S., Tiunov, A. V., Soares, A. E. R., Shidlovskiy, F. K., & Shapiro, B. (2015). An ancient bison from the mouth of the Rauchua River (Chukotka, Russia). *Quaternary Research*, *84*(2), 232–245.

Koch, P. L., & Barnosky, A. D. (2006). Late Quaternary Extinctions: State of the Debate. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 215–250.

Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., Ugan, A., Borregaard, M. K., Gilbert, M. T. P., Nielsen, R., Ho, S. Y. W., Goebel, T., Graf, K. E., Byers, D., Stenderup, J. T., Rasmussen, M., Campos, P. F., Leonard, J. A., Koepfli, K.-P., … Willerslev, E. (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, *479*(7373), 359–364.

Lyman, R. L. (2019). Assumptions and Protocol of the Taxonomic Identification of Faunal Remains in Zooarchaeology: a North American Perspective. *Journal of Archaeological Method and Theory*. https://doi.org/10.1007/s10816-019-09414-0

MacFadden, B. J. (1994). *Fossil Horses: Systematics, Paleobiology, and Evolution of the Family Equidae*. Cambridge University Press.

MacFadden, B. J. (2005). Fossil horses - evidence for evolution. *Science*, *307*(5716), 1728–1730.

Mann, D. H., Groves, P., Gaglioti, B. V., & Shapiro, B. A. (2018). Climate-driven ecological stability as a globally shared cause of Late Quaternary megafaunal extinctions: the Plaids and Stripes Hypothesis. *Biological Reviews of the Cambridge Philosophical Society*. https://doi.org/10.1111/brv.12456

Mann, D. H., Groves, P., Reanier, R. E., Gaglioti, B. V., Kunz, M. L., & Shapiro, B. (2015). Life and extinction of megafauna in the ice-age Arctic. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(46), 14301–14306.

Meiri, M., Lister, A. M., Collins, M. J., Tuross, N., Goebel, T., Blockley, S., Zazula, G. D., van Doorn, N., Dale Guthrie, R., Boeskorov, G. G., Baryshnikov, G. F., Sher, A., & Barnes, I. (2014). Faunal record identifies Bering isthmus conditions as constraint to end-Pleistocene migration to the New World. *Proceedings. Biological Sciences / The Royal Society*, *281*(1776), 20132167.

Moore, A. M. T., Kennett, J. P., Napier, W. M., Bunch, T. E., Weaver, J. C., LeCompte, M., Adedeji, A. V., Hackley, P., Kletetschka, G., Hermes, R. E., Wittke, J. H., Razink, J. J., Gaultois, M. W., & West, A. (2020). Evidence of Cosmic Impact at Abu Hureyra, Syria at the Younger Dryas Onset (~12.8 ka): High-temperature melting at >2200 °C. *Scientific Reports*, *10*(1), 4185.

Orlando, L. (2015). Equids. *Current Biology: CB*, *25*(20), R973–R978.

Orlando, L. (2019). An Ancient DNA Perspective on Horse Evolution. In C. Lindqvist & O. P. Rajora (Eds.), *Paleogenomics: Genome-Scale Analysis of Ancient DNA* (pp. 325–351). Springer International Publishing.

Rosenheim, J. A., & Tabashnik, B. E. (1991). Influence of Generation Time on the Rate of Response to Selection. *The American Naturalist*, *137*(4), 527–541.

Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology*, *19*(1), 131.

Vilstrup, J. T., Seguin-Orlando, A., Stiller, M., Ginolhac, A., Raghavan, M., Nielsen, S. C. A., Weinstock, J., Froese, D., Vasiliev, S. K., Ovodov, N. D., Clary, J., Helgen, K. M., Fleischer, R. C., Cooper, A., Shapiro, B., & Orlando, L. (2013). Mitochondrial phylogenomics of modern and ancient equids. *PloS One*, *8*(2), e55950.

Yuan, J.-X., Hou, X.-D., Barlow, A., Preick, M., Taron, U. H., Alberti, F., Basler, N., Deng, T., Lai, X.-L., Hofreiter, M., & Sheng, G.-L. (2019). Molecular identification of late and terminal Pleistocene Equus ovodovi from northeastern China. *PloS One*, *14*(5), e0216883.

Chapter 1

The case of an arctic wild ass highlights the utility of ancient DNA for validating problematic identifications in museum collections

Alisa Vershinina[1], Joshua D. Kapp[1], Gennady Baryshnikov[2], Beth Shapiro[1,3]

1. Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA 95064, USA

2. Laboratory of Theriology, Zoological Institute of Russian Academy of Sciences, 199034 St. Petersburg, Russia

3. Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA 95064, USA

Molecular Ecology Resources (2020), 00: 1– 9.

ABSTRACT

Museum collections are essential for reconstructing and understanding past biodiversity. Many museum specimens are, however, challenging to identify. Museum samples may be incomplete, have an unusual morphology, or represent juvenile individuals, all of which complicate accurate identification. In some cases, inaccurate identification can lead to false biogeographic reconstructions with cascading impacts on paleontological and paleoecological research. Here we analyze an unusual Equid mandible found in the Far North of the Taymyr peninsula that was identified morphologically as *Equus hemionus*, an ancestor of present-day Asiatic wild asses. If correct, this identification represents the only finding of a putative Late Pleistocene hemione in the Arctic region, and is therefore critical to understanding wild ass evolution and paleoecology. To confirm the accuracy of this specimen's taxonomic assignment, we used ancient DNA and mitochondrial hybridization capture to identify and place this specimen in the larger equid phylogeny. We find that the specimen is actually a member of *E. caballus*, the ancestor of domestic horses. Our study demonstrates the utility of ancient DNA to validate morphological identification, in particular of incomplete, otherwise problematic, or taxonomically unusual museum specimens.

INTRODUCTION

Natural history museums are biomolecular biobanks. Museum collections maintain fossil and biomolecular archives of the evolutionary history of life, including species that are rare, threatened with extinction, or already extinct (Johnson et al.,

2011). Museum collections can be used, for example, to reconstruct evolutionary relationships and geographic distribution of species (Foote et al., 2013; White et al., 2018). Historically, specimens archived in museums have been identified based on their morphological characters. Recently, however, advances in biomolecular techniques including ancient DNA, RNA, and protein sequencing have provided other sources of information that can be recovered from museum specimens (Cappellini et al., 2019; Keller et al., 2017; Meyer et al., 2016). These approaches have expanded the potential utility of museum collections by making it possible to provide taxonomic assignments to fragmentary and otherwise difficult to identify specimens (Brown et al., 2016). In some instances, ancient DNA data have contradicted morphological identifications, leading to both correction of accidental misidentifications and discovery of unknown taxonomic groups (Heintzman et al., 2017; Krause et al., 2010).

Among the most widely studied taxonomic groups using museum-preserved remains is the family Equidae. Equids, whose living members include horses, zebras, donkeys, and asses, were among the first taxonomic groups for which evolutionary history was inferred through examination of archived fossils and bones (MacFadden, 1994). Thanks in part to their abundance and history of living in colder climates, horse evolution has also been a common theme in paleogenomics research. In 1984, the first ancient DNA sequences were recovered from the skin of a museum-preserved subspecies of zebra, the quagga (Higuchi et al., 1984), and the oldest genome yet sequenced is from an early Middle Pleistocene horse from Canada's Yukon (Orlando et al., 2013). The abundance and diversity of equid fossils has also made their taxonomy

contentious, with genetic and morphological analyses often leading to taxonomic reassignments and redesignations (Barron-Ortiz et al., 2019). For example, Przewalski's horse, *E. ferus przewalskii*, was considered for many decades to be the only remaining truly wild horse (Der Sarkissian et al., 2015), but has recently been linked to the lineage of Botai horses that were tamed and herded five thousand years ago (Gaunitz et al. (2018). In another example, DNA from North American horse remains dating to the late Pleistocene was recently used to name a new genus, *Haringtonhippus francisci,* the New World stilt-legged horse (Heintzman et al., 2017). While this lineage was known, its phylogenetic placement within the equids was uncertain, with different authors assigning it to at least five different species (Eisenmann et al., 2008; Heintzman et al., 2017; Weinstock et al., 2005).

Living horses can be broadly subdivided into two major lineages that diverged 4-4.5 million years ago (Orlando et al., 2013): caballoid horses, which includes domestic horses (most often referred to as either *E. ferus caballus* or *E. caballus*), Przewalski's horses (*E. f. przewalski* or *E. c. przewalski*); and non-caballoid horses, which includes hemiones (Asiatic wild asses *E. hemionus* and kiangs *E. kiang*), zebras (*E. zebra*), and donkeys (*E. asinus*). While caballoid horses have been well characterized using ancient DNA (Fages et al., 2019; Gaunitz et al., 2018; Heintzman et al., 2017; Orlando et al., 2013), less is known about the geographic distribution and evolutionary relationships among extinct and extant hemiones. Morphologically, hemiones have both caballoid horse features, such as gracile and slender bodies, and features similar to donkeys, such as a relatively small body size and a large head.

Today, kiangs are found across Tibetan plateau (St-Louis & Côté, 2009) and Asiatic wild asses are found in the deserts and arid steppes of southern Mongolia, Kazakhstan, Iran, and India (IUCN, 2015). While fossils of kiang-like wild asses are known from Pleistocene deposits in Alaska (Harington, 1980), the Pleistocene range of kiangs remains unknown. Wild asses during the Pleistocene spanned from present-day France, where they were known as the European wild ass (*E. hydruntinus*), to China (Figure 1.1A) (E. A. Bennett et al., 2017) in regions south of 50˚N.

During the 1980s, a complete mandible, sample ZIN-35608 (the Zoological Institute of St. Petersburg, Russia), was discovered on the Begichev Islands in the Laptev sea to the east of the Taymyr peninsula, Russia (Figure 1.1B). The mandible, which was identified as belonging to an equid, was small, with both the length of the mandible bone and the lower tooth row 5 cm smaller than what is typical of extinct caballoid horses. Although the other measurements of the mandible were inconclusive, it had a curved lower jaw ridge and a V-shaped linguaflexid of the lower teeth, both of which are characteristics of hemionid horses (Cucchi et al., 2017). Based on these morphological characters, the mandible was assigned to *E. hemionus*, or Asiatic wild ass (Kuzmina, 1997). The discovery of a wild ass in the Taymyr peninsula was paleontologically significant, as it expands the range of wild asses far to the north and adds them to the list of arctic Siberian fauna that lived contemporaneously with mammoths (Figure 1.1A) (Markova et al., 1995).

There are several reasons to suspect that the taxonomic identification of sample ZIN-35608 as a wild ass may have been inaccurate. First, the location where the

specimen was recovered is far outside the known range of Late Pleistocene hemiones. Although wild asses were widespread during the Late Pleistocene, no other wild ass samples are known from northerly regions of Eurasia. To date, the only equid known from the Pleistocene of Siberian Far North is the caballoid horse, thus assigning the specimen in question to caballoids is an alternative hypothesis. Second, the identification of equid fossils is challenging, given their extensive morphological variation both within and between taxonomic groups (D. K. Bennett, 1980; Forsten, 1998; Geigl & Grange, 2012; Orlando et al., 2009; Twiss et al., 2017). Taxonomic identification is particularly complicated if the specimen is a subadult, as suggested for ZIN-35608 based on its small size, as diagnostic features may have not yet formed.

To confirm the identification of a Pleistocene wild ass in the Russian Far North, we extracted and captured ancient mitochondrial DNA and placed sample ZIN-35608 in a phylogenetic tree of the Equidae. Our analyses revealed the specimen not to be a wild ass, but instead a caballoid horse, the ancestor of the domestic horse. The unusually small size of the animal, combined with the standard challenges of morphological identification in equids, likely led to its misinterpretation as a member of the smaller species, *E. hemionus*. Our results underscore the utility of ancient DNA as a paleontological tool, in particular when specimens are challenging to identify morphologically, and highlight the important role that museum collections play in understanding evolutionary and biogeographic processes.

MATERIALS AND METHODS

ZIN-35608 is a partial mandible found on the Begichev Islands, Russia, during the 1980s and currently held in the collection of the Zoological Institute of St. Petersburg (Figure 1.1). To estimate the age of the individual at time of death, we followed the protocol outlined in Hillson (2005). We then collected ~1g of bone surrounding the M2 tooth socket of the ZIN-35608 mandible, which we subdivided for radiocarbon dating at the Keck Radiocarbon facility at UC Irvine and ancient DNA analysis in the purpose-built, sterile, ancient DNA facility at the University of California Santa Cruz Paleogenomics Laboratory. We performed DNA extraction and subsequent processing following standard protocols for working with degraded DNA (Fulton & Shapiro, 2019). Briefly, the sterile lab is located in the building isolated from other molecular research laboratories. Laboratory personnel are instructed to not enter any areas of campus that have a risk of PCR contamination prior to entering the aDNA facility. Once inside, they wear coverall suits with hoods, face masks, and a double layer of gloves. All equipment and surfaces are washed with bleach and ethanol.

We extracted DNA following the Dabney et al (2013) protocol with modifications for the recovery of short molecules (Campos et al., 2012), and a sodium hypochlorite pretreatment (Boessenkool et al., 2017) to reduce the amount of contaminating DNA potentially adhered to the surface of the bone. Following extraction, we prepared the extract into Illumina DNA sequencing libraries following Meyer and Kircher (2010), using Sera-Mag SPRI SpeedBeads (ThermoScientific) in 18% PEG-8000 between each step for library clean-up. To enrich libraries for mitochondrial DNA, we performed in-solution hybridization capture using MyBaits

Mito *E. caballus* RNA bait set (Arbor Biosciences, Ann Arbor, MI), following the manufacturer's protocol version 3.01. We incubated the hybridization reactions for 36h at 65 °C, and then isolated DNA from the probes using Dynabeads magnetic streptavidin-coated beads. We amplified captured libraries with KAPA HiFi 2X master mix using IS5 and IS6 primers, and purified the enriched libraries with SPRI beads as above. We then pooled and sequenced pre- and post-capture libraries on two Illumina MiSeq runs (v3 chemistry, 75bp paired end reads).

We used in-house scripts to process the recovered data, map reads to the reference genomes, and assemble a mitochondrial genome (https://github.com/Paleogenomics/DNA-Post-Processing/blob/master/mito_assembly_pipeline.sh). We removed adapters and merged reads with a minimum overlap of 15bp and minimum length of 27bp using SeqPrep (http://github.com/jstjohn/SeqPrep). After verifying the presence of ancient DNA damage at the end of the reads using MapDamage2 (Jónsson et al., 2013), we mapped merged reads to previously published genomes of *E. caballus* (EquCab3, GeneBank accession GCF_002863925, mitochondrial NC_001640), *E. asinus* (nuclear GCA_003033725), and *E. hemionus* (mitochondrial NC_016061). For alignments to nuclear genomes, we used bwa aln with seed disabled (Li & Durbin, 2009). To assemble the mitochondrial genome, we used mia (https://github.com/mpieva/mapping-iterative-assembler), calling bases with at least 10X independent read coverage and >90% consensus among reads, so as to avoid false nucleotide calls due to ancient DNA damage. We imported the BAM file of the

mitochondrial alignment into Geneious R11 (Biomatters, NZ) to create a consensus nucleotide sequence, which we deposited in GenBank as accession number MN503280.

To create a data set for comparison to ZIN-35608, we added its mitochondrial genome to a previously published alignment of 30 equid mitochondrial genomes (Table 1) (Heintzman et al., 2017). We aligned the assembled mitochondrial genome to this dataset using MAFFT v.7 (Nakamura et al., 2018), and manually checked the alignment for discrepancies. Using the annotation of *E. caballus* mitochondrial genome (GenBank ID JN398421), we subdivided the mitochondrial alignment into six partitions (1st, 2nd, and 3rd codon positions for protein coding genes, concatenated tRNAs, rRNA genes, and the control region). We estimated the appropriate evolutionary models for each partition using jModeltest2 (Darriba et al., 2012).

We estimated phylogenetic trees describing the relationships among the 32 equids in our data set using both a Maximum Likelihood (ML) and a Bayesian approach. For the ML reconstruction, we specified the Malasyian tapir, *Tapirus indicus* as the outgroup lineage (GenBank ID NC_023838) and ran three instances of RAxML v.8.2.4 (Stamatakis, 2014) with the GTRGAMMAI nucleotide model on each partition. We then selected the best ML tree and estimated branch supports using 1000 rapid bootstrap iterations. Bayesian inference (BI) does not require outgroup, therefore we did not include the tapir sequence in this analysis. Following Heinzman et al. (2017), we excluded the 3rd position of the codon and the control region as to account for possible convergent mutations in rapidly evolving positions of the mitochondrial

genome. For the 1st codon, 2nd codon, and the rRNA partitions, we specified the TN93+I+G nucleotide model, and for the tRNA partition we specified HKY+I. We ran two iterations of BEAST 1.8.4 for 100 million iterations per run, sampling model parameters and trees every 1000 iterations (Drummond & Rambaut, 2007). We assumed the uncorrelated relaxed clock model and calibrated the molecular clock using the radiocarbon dates of ancient samples as priors , with ages of the present-day samples set to 0, and a divergence of the crown caballoid group of 4–4.5 Ma (normal prior, mean 4.25M, SD: 1.5M; Orlando et al., 2013). We calibrated radiocarbon ages of the Late Pleistocene samples reported in Table 1 using the IntCal13 curve (Reimer et al., 2013) and OxCal v4.2 (Ramsey, 2009), and assigned the median calibrated age of each sample as prior information. We used the birth-death process with serial samples as a tree prior (Drummond et al., 2006; Stadler, 2010). We discarded the first 25% of MCMC iterations from each run as burn-in, and analyzed parameters for convergence in Tracer (Rambaut et al., 2018). All parameters reached an effective sample size >1000. We then combined trees from the two BEAST runs in logCombiner and calculated the maximum clade credibility (MCC) tree in Tree Annotator (Rambaut & Drummond, 2010). We visualized the BEAST MCC tree and the estimated RAxML phylogeny in Figtree v1.4.2 (Rambaut, 2014).

RESULTS

DNA extraction and mitochondrial enrichment were both successful for specimen ZIN-35608. We sequenced the unenriched library to a depth of 841,010 reads. When these reads were mapped to the nuclear references genomes of *E. caballus*

and *E. asinus,* 68,276 (8.1%) of recovered reads mapped to the *E. caballus* reference nuclear genome, and 63,916 (7.6%) of reads mapped *E. asinus*. Although this number of reads is small and is not indicative of a final taxonomic assignment, it suggests a closer relationship between ZIN-35608 and caballoid horses than donkeys. The median length of aligned DNA sequences was 55bp, and we observed an elevated frequency of G>A and C>T substitutions at the ends of molecules, consistent with degraded DNA. After mitochondrial capture, we sequenced the captured library to a depth of 249,469 reads. Of these, 35,271 (14.1%) unique reads mapped to the *E. hemionus* mitochondrial genome, and 44,270 (17.8%) unique reads mapped to *E. caballus*, resulting in mitochondrial genomes with an average coverage of 206x using *E. hemionus* as the starting reference for mapping and 217x when assembling on *E. caballus*. The final assemblies were identical except for three sections of the control region that were not assembled with *E. hemionus* as the starting reference. We therefore chose the assembly seeded with *E. caballus* for phylogenetic reconstruction.

Radiocarbon dating, which was performed at the Keck facility at UC Irvine, provided an uncalibrated age of 19470±70 years before present (UCIAMS-199226). To incorporate this age into the Bayesian analysis described above, we calibrated it using the IntCal13 radiocarbon curve as implemented in OxCal 4.3 (Ramsey, 2009; Reimer et al., 2013). This provided a median age of 23,457 calibrated years before present (Cal BP 23697-23138, 95.4% probability range). Based on the state of tooth eruption and microwear analysis (Hillson, 2005), we estimate the age of the individual to be 4-4.5 years at the time of death.

The mitochondrial phylogenies reconstructed with Maximum Likelihood and Bayesian approaches were topologically concordant with respect to the major clades within the Equidae family, although branching order slightly differed within the non-caballoid clade (Figure 1.2A,B). The major difference is that in the ML phylogeny, *E. ovodovi* falls outside of the diversity of zebras, donkeys, and asses, whereas in the Bayesian phylogeny it is sister to *E. asinus,* although with low statistical support. In both ML and BEAST analysis, ZIN-35608 falls within the clade of caballoid horses with strong statistical support (Figure 1.2B). Based on these results, we conclude that ZIN-35608 is a caballoid horse.

DISCUSSION

Our ancient mitochondrial DNA data indicate that ZIN-35608 is a member of the caballoid clade of ancient horses, rather than a hemione as assigned based on morphology. Although we generated too few nuclear reads for a confident taxonomic assignment based on nuclear genomic data, a greater proportion of reads from the shotgun library mapped to the *E. caballus* genome than to the *E. asinus* genome, suggesting that the former is a closer evolutionary match. Both ML and BEAST phylogenies reconstructed with complete high coverage mitochondrial genomes place ZIN-35608 confidently within the past and present diversity of caballoid horses (Figure 1.2).

Our results do not support the proposed expansion of wild asses to the North of Eastern Siberia, but instead indicate that caballoid horses were present in the Begichev Islands during the Late Pleistocene (Figure 1.1A). With a calibrated age of 23,457 years

ago, ZIN-35608 lived during the peak cold interval of the Last Glacial Maximum. At this time, the sea level was significantly lower than today, which would have allowed the mainland horse population to expand to what is today the Begichev Islands. Adjacent to the Barents-Kara Ice Sheet, the region was at the arctic edge of the dry Mammoth steppe, dominated by graminoid vegetation (Binney et al., 2017; Mangerud et al., 2004; Tarasov et al., 2000). Such habitats were known to be widely populated by caballoid horses at this time (Zimov et al., 2012).

The erroneous taxonomic assignment of ZIN-35608 highlights the challenge of generating an accurate taxonomic identification from some paleontological remains, which can be problematic when, for example, specimens are fragmentary or come from animals that have not reach full adult size. In this and other cases, ancient DNA has proven to be a useful tool for resolving such questions. Barbanera et al. (2016), for example, amplified the mitochondrial cytochrome *b* gene from specimens from three museum collections that had been identified as smooth-coated otters (*Lutrogale perspicillata*). Ancient DNA revealed these to belong to three different species, none of which were *L. perspicillata* (Barbanera et al., 2016). Similarly, Cappellini et al. (2014) used a combination of ancient DNA and proteomics to correct the taxonomic assignment of the original type material of the Asian elephant *Elephas maximus.* The biomolecular data revealed that this sample, a complete ethanol-preserved fetus originally described by Linnaeus, was in fact an African elephant, *Loxodonta sp.* (Cappellini et al., 2014).

In addition to resolving incorrect taxonomic identifications, paleogenomic data can augment the contribution of museum collections to our understanding of evolutionary history, paleoecology, and conservation. Samples from organisms that lived before environmental shifts or periods of population decline can be used to estimate evolutionary changes that occur as a consequence of those events. For example, ancient DNA from museum preserved specimens collected within the last few centuries has revealed a dramatic reduction in genetic diversity of the critically endangered Western Australian woylie and Eastern gorilla, both of which are associated with anthropogenic habitat changes (Pacioni et al., 2015; van der Valk et al., 2018). Ancient DNA from much older museum specimens, coupled with radiocarbon dating and stable isotope analysis, has also been used to reconstruct ecological changes during the early Holocene megafaunal mass extinction event (Lorenzen et al., 2011; Shapiro et al., 2004). Finally, museum-preserved specimens have tremendous potential value for conservation. Recently, museum specimens of Eurasian beaver, *Castor fiber*, from the last 10,000 years helped to identify potential source populations for this species' reintroduction to Britain (Marr et al., 2018). In the European beaver study, museum specimens allowed the reconstruction of past diversity and ancient dispersal events across geographic locations, which is essential for finding a suitable source population for controlled genetic rescue (Dietl & Flessa, 2011; Leonard, 2008). Although the preservation of ZIN-35608 is poor, future advances in DNA recovery efficiency may allow a complete genome sequence to be isolated from this specimen, which would help to reveal whether the Begichev populations were genetically isolated

from the mainland Yakutiya and how horse population structure changed with the separation of the Begichev Islands from the continent. While the present study of ZIN-35608 is a single example, it highlights the potential power of museum specimens, combined with increasingly sophisticated biomolecular approaches, to reveal the pattern and process of biodiversity change over time.

While not all museum specimens retain DNA, advances in paleogenomics approaches continue to expand the range of material from which DNA can be recovered, increasing the value of museum specimens. New methods have been developed to extract DNA from samples fixed in formalin (Hykin et al., 2015; Ruane & Austin, 2017) and ethanol (McGuire et al., 2018) and to decontaminate samples with sodium hypochlorite prior to extraction (Korlević et al., 2015) thereby increasing the fraction of useful DNA recovered. In the current study, we used in-solution hybridization capture, which is an approach developed to recover short fragments of a targeted region of the genome efficiently. This method allows recovery of sufficient quantities of data for population genetic analyses even when samples are poorly preserved. Such data can be generated US$50-$100 per sample, although costs vary depending on sample preservation, experimental approach to DNA extraction and library preparation, and local costs of consumables and sequencing. Other strategies to reduce cost and/or generate new biological information from ancient specimens include DNA barcoding and metabarcoding of bulk-extracted bone fragments, which enables taxonomic identification of a co-recovered community of organisms even when remains are too fragmentary to attempt morphological analysis (Grealy et al., 2015).

Finally, while ancient DNA recovered from museum specimens has broad utility in ecological and evolutionary analyses, this approach is most powerful when used in combination with other methods, including morphological analysis, radiocarbon dating, stable isotopic analysis, and other techniques. Many of these approaches are inherently destructive, and it is therefore important to consider the long-term impact on the collections when deciding what approaches are best for any particular sample.

CONCLUSION

Our results reveal that a museum specimen recovered in the Russian Far North and identified based on morphological characters as a wild ass is actually a caballoid horse. This incorrect identification is probably a consequence of the unusually small size of the horse combined with problematic teeth characteristics. Our finding therefore disputes the hypothesis that Pleistocene wild asses expanded to the North of Eurasia during the Pleistocene. Our study demonstrates how ancient DNA can be used to validate the taxonomic identity of problematic museum specimens. The growing diversity of approaches that can be used to analyze the preserved remains of organisms highlights the important role of museum collections in advancing our understanding of evolutionary history, paleoecology, and conservation.

ACKNOWLEDGEMENTS

This work was funded by National Science Foundation grant NSF ANS 1417036 and Institute of Museum and Library Services grant MG-30-17-0045-17. GB is supported by the Program of the Russian Academy of Sciences Presidium and the Russian

Ministry of Education and Science "Evolution of the organic world. The role and significance of planetary processes" (2019).

AUTHOR CONTRIBUTIONS

G.B., A.V. and B.S. conceived the study; A.V. obtained the sample; J.K. extracted DNA, prepared sequencing library and performed hybridization capture; A.V. conducted analyses of the sequencing data; A.V. and B.S. wrote the manuscript. All authors discussed the results and contributed to the final version of the manuscript.

BIBLIOGRAPHY

Barbanera, F., Moretti, B., Guerrini, M., Al-Sheikhly, O. F., & Forcina, G. (2016). Investigation of ancient DNA to enhance natural history museum collections: Misidentification of smooth-coated otter (Lutrogale perspicillata) specimens across multiple museums. *Belgian Journal of Zoology*, *146*(2), 101–112.

Barron-Ortiz, C., Avilla, L., Jass, C., Bravo-Cuevas, V., Machado, H., & Mothé, D. (2019). What is Equus? Reconciling taxonomy and phylogenetic analyses. *Frontiers in Ecology and Evolution*, *7*, 343.

Bennett, D. K. (1980). Stripes do not a Zebra Make, Part I: A Cladistic Analysis of Equus. *Systematic Zoology*, *29*(3), 272–287.

Bennett, E. A., Champlot, S., Peters, J., Arbuckle, B. S., Guimaraes, S., Pruvost, M., Bar-David, S., Davis, S. J. M., Gautier, M., Kaczensky, P., Kuehn, R., Mashkour, M., Morales-Muñiz, A., Pucher, E., Tournepiche, J.-F., Uerpmann, H.-P., Bălăşescu, A., Germonpré, M., Gündem, C. Y., … Geigl, E.-M. (2017). Taming the late Quaternary phylogeography of the Eurasiatic wild ass through ancient and modern DNA. *PloS One*, *12*(4), e0174216.

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., Andreev, A., Bezrukova, E., Blyakharchuk, T., Jankovska, V., Khazina, I., Krivonogov, S., Kremenetski, K., Nield, J., Novenko, E., Ryabogina, N., Solovieva, N., Willis, K., & Zernitskaya, V. (2017). Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. *Quaternary Science Reviews*, *157*, 80–97.

Boessenkool, S., Hanghøj, K., Nistelberger, H. M., Der Sarkissian, C., Gondek, A. T., Orlando, L., Barrett, J. H., & Star, B. (2017). Combining bleach and mild predigestion improves ancient DNA recovery from bones. *Molecular Ecology Resources*, *17*(4), 742–751.

Brown, S., Higham, T., Slon, V., Pääbo, S., Meyer, M., Douka, K., Brock, F., Comeskey, D., Procopio, N., Shunkov, M., Derevianko, A., & Buckley, M. (2016). Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Scientific Reports*, *6*, 23559.

Campos, P. F., Craig, O. E., Turner-Walker, G., Peacock, E., Willerslev, E., & Gilbert, M. T. P. (2012). DNA in ancient bone - where is it located and how should we extract it? *Annals of Anatomy = Anatomischer Anzeiger: Official Organ of the Anatomische Gesellschaft*, *194*(1), 7–16.

Cappellini, E., Gentry, A., Palkopoulou, E., Ishida, Y., Cram, D., Roos, A.-M., Watson, M., Johansson, U. S., Fernholm, B., Agnelli, P., Barbagli, F., Littlewood, D. T. J., Kelstrup, C. D., Olsen, J. V., Lister, A. M., Roca, A. L.,

Dalén, L., & Gilbert, M. T. P. (2014). Resolution of the type material of the Asian elephant, Elephas maximus Linnaeus, 1758 (Proboscidea, Elephantidae). *Zoological Journal of the Linnean Society*, *170*(1), 222–232.

Cappellini, E., Welker, F., Pandolfi, L., Ramos-Madrigal, J., Samodova, D., Rüther, P. L., Fotakis, A. K., Lyon, D., Víctor Moreno-Mayar, J., Bukhsianidze, M., Jersie-Christensen, R. R., Mackie, M., Ginolhac, A., Ferring, R., Tappen, M., Palkopoulou, E., Dickinson, M. R., Stafford, T. W., Chan, Y. L., … Willerslev, E. (2019). Early Pleistocene enamel proteome from Dmanisi resolves Stephanorhinus phylogeny. *Nature*, 1–5.

Cucchi, T., Mohaseb, A., Peigné, S., Debue, K., Orlando, L., & Mashkour, M. (2017). Detecting taxonomic and phylogenetic signals in equid cheek teeth: towards new palaeontological and archaeological proxies. *Royal Society Open Science*, *4*(4), 160997.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., & Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(39), 15758–15763.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772.

Der Sarkissian, C., Ermini, L., Schubert, M., Yang, M. A., Librado, P., Fumagalli, M., Jónsson, H., Bar-Gal, G. K., Albrechtsen, A., Vieira, F. G., Petersen, B., Ginolhac, A., Seguin-Orlando, A., Magnussen, K., Fages, A., Gamba, C., Lorente-Galdos, B., Polani, S., Steiner, C., … Orlando, L. (2015). Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Current Biology: CB*, *25*(19), 2577–2583.

Dietl, G. P., & Flessa, K. W. (2011). Conservation paleobiology: putting the dead to work. *Trends in Ecology & Evolution*, *26*(1), 30–37.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, *4*(5), e88.

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, *7*, 214.

Eisenmann, V., Howe, J., & Pichardo, M. (2008). Old world hemiones and new world slender species (Mammalia, Equidae). *Palaeovertebrata*, *36*(1-4). https://doi.org/10.18563/pv.36.1-4.159-233

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K. A. S., Albizuri, S., Alfarhan,

A. H., Allentoft, M., Alquraishi, S., Anthony, D., Baimukhanov, N., Barrett, J. H., Bayarsaikhan, J., Benecke, N., Bernáldez-Sánchez, E., … Orlando, L. (2019). Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell*, *177*(6), 1419–1435.e31.

Foote, A. D., Kaschner, K., Schultze, S. E., Garilao, C., Ho, S. Y. W., Post, K., Higham, T. F. G., Stokowska, C., van der Es, H., Embling, C. B., Gregersen, K., Johansson, F., Willerslev, E., & Gilbert, M. T. P. (2013). Ancient DNA reveals that bowhead whale lineages survived Late Pleistocene climate change and habitat shifts. *Nature Communications*, *4*, 1677.

Forsten, A. (1998). The Late Pleistocene-Holocene horses,not asses, from Japan (Mammalia, Perissodactyla, Equus). *Geobios . Memoire Special*, *31*(4), 545–548.

Fulton, T. L., & Shapiro, B. (2019). Setting Up an Ancient DNA Laboratory. In B. Shapiro, A. Barlow, P. D. Heintzman, M. Hofreiter, J. L. A. Paijmans, & A. E. R. Soares (Eds.), *Ancient DNA: Methods and Protocols* (pp. 1–13). Springer New York.

Gaunitz, C., Fages, A., Hanghøj, K., Albrechtsen, A., Khan, N., Schubert, M., Seguin-Orlando, A., Owens, I. J., Felkel, S., Bignon-Lau, O., de Barros Damgaard, P., Mittnik, A., Mohaseb, A. F., Davoudi, H., Alquraishi, S., Alfarhan, A. H., Al-Rasheid, K. A. S., Crubézy, E., Benecke, N., … Orlando, L. (2018). Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science*, *360*(6384), 111–114.

Geigl, E.-M., & Grange, T. (2012). Eurasian wild asses in time and space: morphological versus genetic diversity. *Annals of Anatomy = Anatomischer Anzeiger: Official Organ of the Anatomische Gesellschaft*, *194*(1), 88–102.

Grealy, A. C., McDowell, M. C., Scofield, P., Murray, D. C., Fusco, D. A., Haile, J., Prideaux, G. J., & Bunce, M. (2015). A critical evaluation of how ancient DNA bulk bone metabarcoding complements traditional morphological analysis of fossil assemblages. *Quaternary Science Reviews*, *128*, 37–47.

Harington, C. R. (1980). Pleistocene mammals from Lost Chicken Creek, Alaska. *Canadian Journal of Earth Sciences*, *17*(2), 168–198.

Heintzman, P. D., Zazula, G. D., MacPhee, R. D. E., Scott, E., Cahill, J. A., McHorse, B. K., Kapp, J. D., Stiller, M., Wooller, M. J., Orlando, L., Southon, J., Froese, D. G., & Shapiro, B. (2017). A new genus of horse from Pleistocene North America. *eLife*, *6*. https://doi.org/10.7554/eLife.29944

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A., & Wilson, A. C. (1984). DNA sequences from the quagga, an extinct member of the horse family. *Nature*, *312*(5991), 282–284.

Hillson, S. (2005). *Teeth by Simon Hillson*. Cambridge University Press.

Hykin, S. M., Bi, K., & McGuire, J. A. (2015). Fixing Formalin: A Method to Recover Genomic-Scale DNA Sequence Data from Formalin-Fixed Museum Specimens Using High-Throughput Sequencing. *PloS One*, *10*(10), e0141579.

IUCN. (2015). *Equus hemionus: Kaczensky, P., Lkhagvasuren, B., Pereladova, O., Hemami, M. & Bouskila, A.: The IUCN Red List of Threatened Species 2015: e.T7951A45171204* [Data set]. https://doi.org/10.2305/IUCN.UK.2015-4.RLTS.T7951A45171204.en

Johnson, K. G., Brooks, S. J., Fenberg, P. B., Glover, A. G., James, K. E., Lister, A. M., Michel, E., Spencer, M., Todd, J. A., Valsami-Jones, E., Young, J. R., & Stewart, J. R. (2011). Climate Change and Biosphere Response: Unlocking the Collections Vault. *Bioscience*, *61*(2), 147–153.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684.

Keller, A., Kreis, S., Leidinger, P., Maixner, F., Ludwig, N., Backes, C., Galata, V., Guerriero, G., Fehlmann, T., Franke, A., Meder, B., Zink, A., & Meese, E. (2017). miRNAs in Ancient Tissue Specimens of the Tyrolean Iceman. *Molecular Biology and Evolution*, *34*(4), 793–801.

Korlević, P., Gerber, T., Gansauge, M.-T., Hajdinjak, M., Nagel, S., Aximu-Petri, A., & Meyer, M. (2015). Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *BioTechniques*, *59*(2), 87–93.

Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P., & Pääbo, S. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature*, *464*(7290), 894–897.

Kuzmina, I. E. (1997). *Horses of North Eurasia from The Pliocene Till The Present Time* (N. K. Vereschagin (ed.); Vol. 273). Russian Academy of Sciences.

Leonard, J. A. (2008). Ancient DNA applications for wildlife conservation. *Molecular Ecology*, *17*(19), 4186–4196.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., Ugan, A., Borregaard, M. K., Gilbert, M. T. P., Nielsen, R., Ho, S. Y. W., Goebel, T., Graf, K. E., Byers, D., Stenderup, J. T., Rasmussen, M., Campos, P. F., Leonard, J. A., Koepfli, K.-P., … Willerslev, E. (2011). Species-specific

responses of Late Quaternary megafauna to climate and humans. *Nature*, *479*(7373), 359–364.

MacFadden, B. J. (1994). *Fossil Horses: Systematics, Paleobiology, and Evolution of the Family Equidae*. Cambridge University Press.

Mangerud, J., Jakobsson, M., Alexanderson, H., Astakhov, V., Clarke, G. K. C., Henriksen, M., Hjort, C., Krinner, G., Lunkka, J.-P., Möller, P., Murray, A., Nikolskaya, O., Saarnisto, M., & Svendsen, J. I. (2004). Ice-dammed lakes and rerouting of the drainage of northern Eurasia during the Last Glaciation. *Quaternary Science Reviews*, *23*(11), 1313–1332.

Markova, A. K., Smirnov, N. G., Kozharinov, A. V., Kazantseva, N. E., & Kitaev, L. M. (1995). *Late Pleistocene distribution and diversity of mammals in Northern Eurasia (PALEOFAUNA Database)* (Vols. 28–29). Paleontologia i Evolucio, t.28-29, pp. 5-143.

Marr, M. M., Brace, S., Schreve, D. C., & Barnes, I. (2018). Identifying source populations for the reintroduction of the Eurasian beaver, Castor fiber L. 1758, into Britain: evidence from ancient DNA. *Scientific Reports*, *8*(1), 2708.

McGuire, J. A., Cotoras, D. D., O'Connell, B., Lawalata, S. Z. S., Wang-Claypool, C. Y., Stubbs, A., Huang, X., Wogan, G. O. U., Hykin, S. M., Reilly, S. B., Bi, K., Riyanto, A., Arida, E., Smith, L. L., Milne, H., Streicher, J. W., & Iskandar, D. T. (2018). Squeezing water from a stone: high-throughput sequencing from a 145-year old holotype resolves (barely) a cryptic species problem in flying lizards. *PeerJ*, *6*, e4470.

Meyer, M., Arsuaga, J.-L., de Filippo, C., Nagel, S., Aximu-Petri, A., Nickel, B., Martínez, I., Gracia, A., Bermúdez de Castro, J. M., Carbonell, E., Viola, B., Kelso, J., Prüfer, K., & Pääbo, S. (2016). Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*, *531*(7595), 504–507.

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, *2010*(6), db.prot5448.

Nakamura, T., Yamada, K. D., Tomii, K., & Katoh, K. (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* , *34*(14), 2490–2492.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P. L. F., Fumagalli, M., Vilstrup, J. T., Raghavan, M., Korneliussen, T., Malaspinas, A.-S., Vogt, J., Szklarczyk, D., Kelstrup, C. D., … Willerslev, E. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, *499*(7456), 74–78.

Orlando, L., Metcalf, J. L., Alberdi, M. T., Telles-Antunes, M., Bonjean, D., Otte, M., Martin, F., Eisenmann, V., Mashkour, M., Morello, F., Prado, J. L., Salas-Gismondi, R., Shockey, B. J., Wrinn, P. J., Vasil'ev, S. K., Ovodov, N. D., Cherry, M. I., Hopwood, B., Male, D., … Cooper, A. (2009). Revising the recent evolutionary history of equids using ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(51), 21754–21759.

Pacioni, C., Hunt, H., Allentoft, M. E., Vaughan, T. G., Wayne, A. F., Baynes, A., Haouchar, D., Dortch, J., & Bunce, M. (2015). Genetic diversity loss in a biodiversity hotspot: ancient DNA quantifies genetic decline and former connectivity in a critically endangered marsupial. *Molecular Ecology*, *24*(23), 5813–5828.

Rambaut, A. (2014). FigTree 1.4. 2 software. *Institute of Evolutionary Biology, Univ. Edinburgh*. https://scholar.google.ca/scholar?cluster=5129645832770813008&hl=en&as_sdt=0,5&sciodt=0,5

Rambaut, A., & Drummond, A. J. (2010). TreeAnnotator version 1.6. 1. *University of Edinburgh, Edinburgh, UK*. https://scholar.google.ca/scholar?cluster=16034997356612446945&hl=en&as_sdt=0,5&sciodt=0,5

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904.

Ramsey, C. B. (2009). Bayesian Analysis of Radiocarbon Dates. *Radiocarbon*, *51*(1), 337–360.

Reimer, P. J., Bard, E., Bayliss, A., Warren Beck, J., Blackwell, P. G., Ramsey, C. B., Buck, C. E., Cheng, H., Lawrence Edwards, R., Friedrich, M., Grootes, P. M., Guilderson, T. P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T. J., Hoffmann, D. L., Hogg, A. G., Hughen, K. A., … van der Plicht, J. (2013). IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon*, *55*(4), 1869–1887.

Ruane, S., & Austin, C. C. (2017). Phylogenomics using formalin-fixed and 100+ year-old intractable natural history specimens. *Molecular Ecology Resources*, *17*(5), 1003–1008.

Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., Gilbert, M. T. P., Barnes, I., Binladen, J., Willerslev, E., Hansen, A. J., Baryshnikov, G. F., Burns, J. A., Davydov, S., Driver, J. C., Froese, D. G., Harington, C. R., Keddie, G., … Cooper, A. (2004). Rise and fall of the Beringian steppe bison. *Science*, *306*(5701), 1561–1565.

Stadler, T. (2010). Sampling-through-time in birth-death trees. *Journal of Theoretical Biology*, *267*(3), 396–404.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* , *30*(9), 1312–1313.

St-Louis, A., & Côté, S. D. (2009). Equus kiang (Perissodactyla: Equidae). *Mammalian Species*, *835*, 1–11.

Tarasov, P. E., Volkova, V. S., Webb, T., Guiot, J., Andreev, A. A., Bezusko, L. G., Bezusko, T. V., Bykova, G. V., Dorofeyuk, N. I., Kvavadze, E. V., Osipova, I. M., Panova, N. K., & Sevastyanov, D. V. (2000). Last glacial maximum biomes reconstructed from pollen and plant macrofossil data from northern Eurasia. *Journal of Biogeography*, *27*(3), 609–620.

Twiss, K. C., Wolfhagen, J., Madgwick, R., Foster, H., Demirergi, G. A., Russell, N., Everhart, J. L., Pearson, J., & Mulville, J. (2017). Horses, Hemiones, Hydruntines? Assessing the Reliability of Dental Criteria for Assigning Species to Southwest Asian Equid Remains: Assessing the Reliability of SW Asian Equid Dental Criteria. *International Journal of Osteoarchaeology*, *27*(2), 298–304.

van der Valk, T., Sandoval-Castellanos, E., Caillaud, D., Ngobobo, U., Binyinyi, E., Nishuli, R., Stoinski, T., Gilissen, E., Sonet, G., Semal, P., Kalthoff, D. C., Dalén, L., & Guschanski, K. (2018). Significant loss of mitochondrial diversity within the last century due to extinction of peripheral populations in eastern gorillas. *Scientific Reports*, *8*(1), 6551.

Weinstock, J., Willerslev, E., Sher, A., Tong, W., Ho, S. Y. W., Rubenstein, D., Storer, J., Burns, J., Martin, L., Bravi, C., Prieto, A., Froese, D., Scott, E., Xulong, L., & Cooper, A. (2005). Evolution, systematics, and phylogeography of pleistocene horses in the new world: a molecular perspective. *PLoS Biology*, *3*(8), e241.

White, L. C., Mitchell, K. J., & Austin, J. J. (2018). Ancient mitochondrial genomes reveal the demographic history and phylogeography of the extinct, enigmatic thylacine (Thylacinus cynocephalus). *Journal of Biogeography*, *45*(1), 1–13.

Zimov, S. A., Zimov, N. S., Tikhonov, A. N., & Chapin, F. S. (2012). Mammoth steppe: a high-productivity phenomenon. *Quaternary Science Reviews*, *57*, 26–45.

Table 1. Samples used in the current study, their NCBI numbers, taxonomic position, uncalibrated radiocarbon age, and calendar dates used for BEAST. The star (*) marks specimen ages estimated by stratigraphic dating. All other non-present day specimens are radiocarbon dated.

GenBank ID	Species	Age (uncalibrated)	Calendar date BP (IntCal13 median)
Non-caballoid equids			
Hemiones			
NC016061	*Equus hemionus*	Present day	0
NC018782	*E. hemionus*	Present day	0
JX312732	*E. kiang*	Present day	0
Donkeys			
KM881681	*E. asinus somalicus*	Present day	0
NC001788	*E. asinus*	Present day	0
Zebras			
NC018780	*E. zebra hartmannae*	Present day	0
NC020476	*E. zebra*	Present day	0
NC020432	*E. grevyi*	Present day	0
JX312729	*E. burchellii chapmani*	Present day	0
KM881680	*E. burchellii quagga*	Present day	0
Extinct non-caballoid equids			
NC018783	*E. ovodovi*	45,000*	45,000*
Caballoid equids			
MN503280	*E. caballus* - ZIN35608	19,470±70	23,457
KT757749	*E. caballus*	28,800±1100	32,918
KT757757	*E. caballus*	34,460±240	38,961
KT757759	*E. caballus*	13,940±55	16,908
KT757761	*E. caballus*	Present day	0
NC001640	*E. caballus*	Present day	0
KT757763	*E. scotti*	560,000-780,000*	650,000*

KT168318	*E. lambei*	33,760±400	38,138
KT168322	*E. lambei*	21,420±80	25,749
	New World Stilt Legged horses		
KT168320	*Haringtonhippus francisci*	28,740±570	32,767
KT168326	*H. francisci*	46,500±1900	47,770
KT168332	*H. francisci*	33,400±430	37,664
KT168333	*H. francisci*	33,560±440	37,858
KT168335	*H. francisci*	14,450±90	17,598
KT168336	*H. francisci*	28,390±240	32,311
	Extinct South American horses		
KM881671	*Hippidion saldiasi*	13,990±150	16,980
KM881672	*H. saldiasi*	11,900±60	13,709
KM881673	*H. saldiasi*	13,890±60	16,830
KM881675	*H. saldiasi*	10,680±40	12,653
KM881677	*H. sp*	13,275±30	15,961

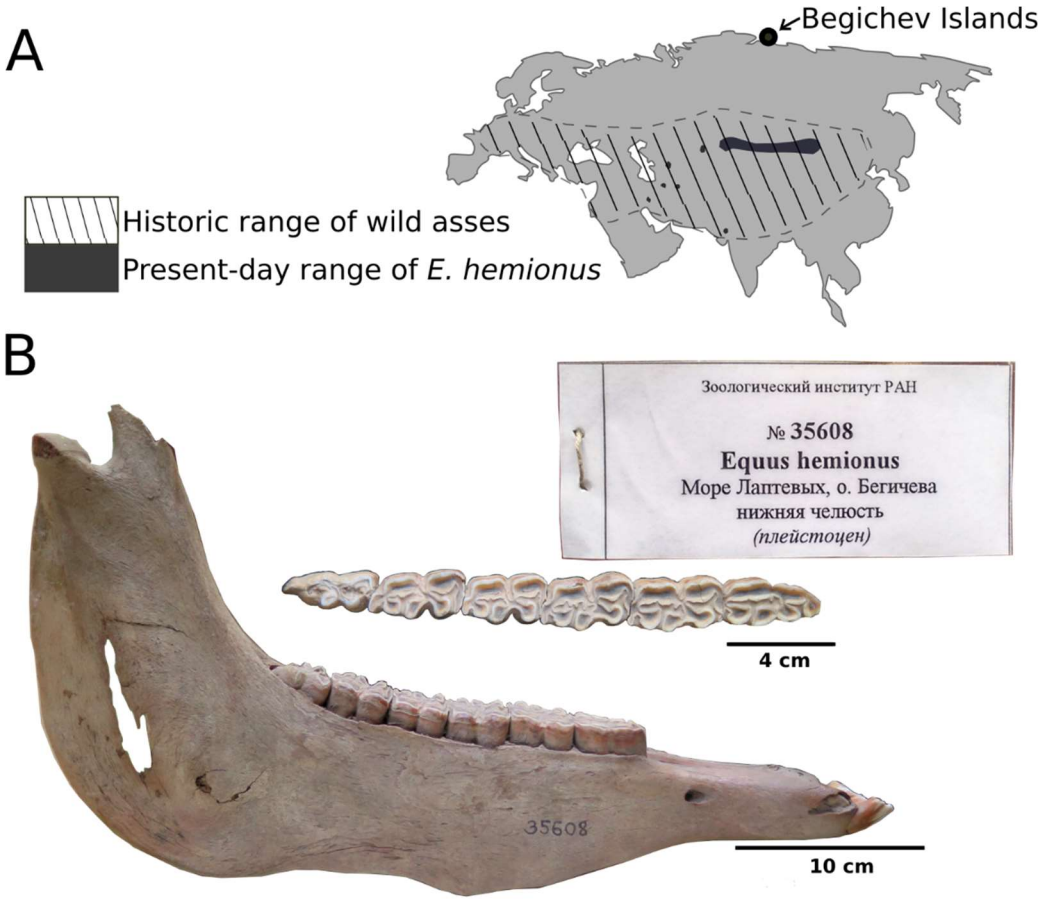


Figure 1.1. A. A map of Eurasia highlighting the distribution of extinct European (*E. hydruntinus)* and present-day Asiatic (*E. hemionus*) wild asses. The black dot shows the location of the Begichev Islands where ZIN-35608 was found. **B.** A mandible found on Begichev Islands (ZIN-35608) and identified as *E. hemionus*.

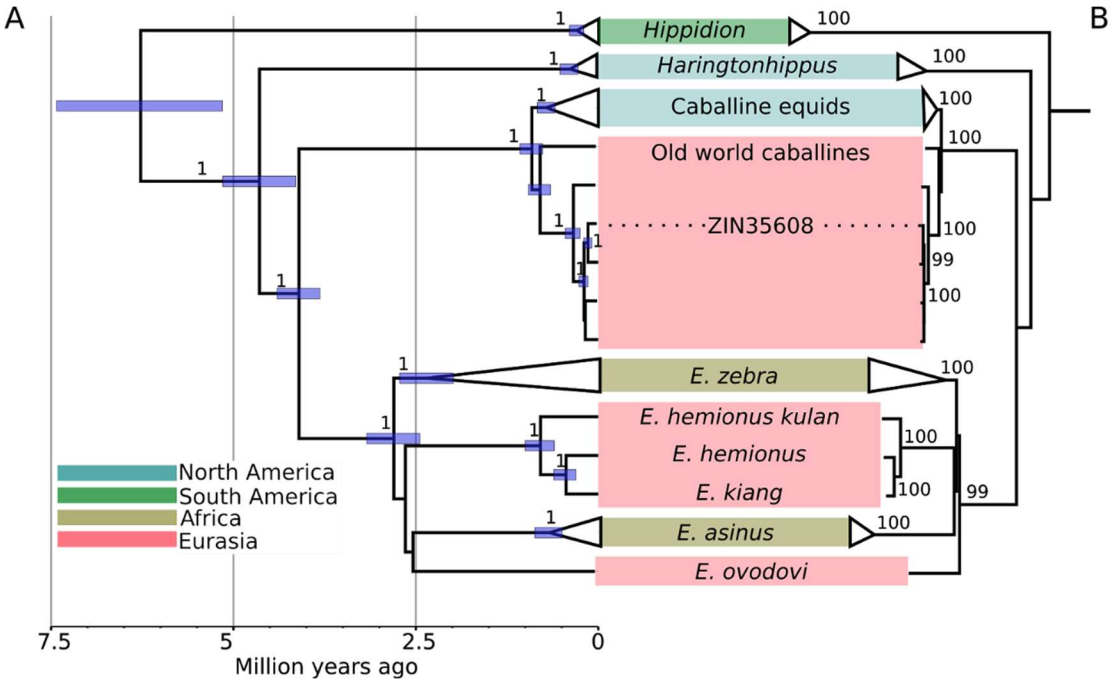


Figure 1.2. Molecular phylogenies of 31 mitochondrial genomes of various Equid groups sampled worldwide. Color-coding corresponds to different continents. **A.** Maximum clade credibility tree reconstructed with BEAST. Each node has a bar corresponding to a 95% HPD height interval. **B.** Maximum likelihood phylogeny estimated using tapir as an outgroup (the outgroup is not shown). Posterior probabilities and bootstrap supports higher than 0.95 and 95 respectively are indicated with numbers above and below branches.

Chapter 2

Paleogenomes reveal Pleistocene cross-continental dispersals and admixture in caballine horses

ABSTRACT

Regional scale shifts in habitat productivity, availability and accessibility have potentially broad evolutionary consequences by both creating and eliminating barriers to dispersal for many species. Here, we examine the influence of Beringia on the evolution and diversification of caballoid horses, including *Equus caballus,* during the Pleistocene. We analyse two high coverage nuclear paleogenomes, recovered from fossils of caballoid horses found in Yukon permafrost, northwest Canada, and a panel of mitochondrial genomes, acquired from a collection of Holarctic horse fossils. Phylogenetic analysis of mitochondrial genomes indicated a signal of phylogeographic population structure in caballoid horses inhabiting Western and Eastern Beringia. We infer, however, a signal of extensive population dispersal as individuals with maternal haplogroup corresponding to Eurasia were found in Alaska while the North American haplogroup was present in Russian Far North. Using coalescent simulations on a set of neutral regions ascertained from two new North American and previously published ancient Eurasian caballoid horse genomes, we infer that ancestors of two continental caballoid horse populations diverged around 0.8-1 million years ago. Estimated cross-continental migration rates (2-4%) and signal of admixture inferred with D-statistic show that after this divergence, the evolution of caballoid horses continued in the

presence of gene flow across the Bering Land Bridge. Our findings show that Beringia acted as a critical contact zone for various caballoid horse populations, suggesting its important role in shaping the genetic diversity for Holarctic mammals during the Quaternary.

INTRODUCTION

Present-day Beringia is the region spanning eastern Russian Taymyr, Chukotka, Alaska, and Yukon with the Bering Strait separating Eurasian and North American continents. The heart of Beringia is the Bering Land Bridge (BLB) - an ephemeral tract of land that connected Chukotka and Alaska in the past during glaciations when the sea level lowered (S. A. Elias & Brigham-Grette, 2013). The BLB sporadically submerged under water during rapid cold-to-warm climate transitions and finally inundated around 11 thousand years before present (kyr BP) (Jakobsson et al., 2017), ceasing cross-continental gene flow and dispersal of animals inhabiting the Ice Age Beringian Arctic. The disappearance of this corridor critically disrupted the continuity of the Mammoth Steppe grassland and impacted large-bodied grazers by fragmenting their habitats (Lorenzen et al., 2011; Zimov et al., 2012).

The final inundation of the BLB happened roughly at the same time as the North American extinction of its iconic mammal - the Pleistocene caballoid horse (CH, *Equus* sp.) (Guthrie, 2003). Caballoid, or "true" horses, are a group encompassing domestic horses, Przewalski's horse, and their extinct Pleistocene ancestors, which together with their non-caballoid sister clade of wild asses and donkeys belong to the genus *Equus*

(Barron-Ortiz et al., 2019). Now surviving in the wild only in Asia, caballoid horses trace their evolutionary history to North America where they have been an extremely successful group thriving during the dynamic environment of the last Ice Age (MacFadden, 2005). CH population density began to decline concurrently with the climatic change of the Pleistocene-Holocene transition and the mass extinction of other megamammals (Barnosky et al., 2004; Guthrie, 2003; Lorenzen et al., 2011). While all horses in North America eventually became extinct, the Eurasian CH population survived the climatic transition, although they experienced a sharp decline before their domestication around 5 kyr ago (Fages et al., 2019). It is still unclear whether human impacts, environmental change, or a combination of these factors were the leading cause of CH extirpation in their native continent (Guthrie, 2006; Koch & Barnosky, 2006; Mann et al., 2018).

Caballoid horses hold the record for the oldest paleogenome known to date (Orlando et al., 2013). The genome of the Thistle Creek horse, found in Klondike, Yukon Territory and dated to 560–780 kyr BP, was the first and only nuclear genome of a New World CH until the present study. The analysis of the Thistle Creek horse genome clarified the age of the most recent ancestor (MRCA) of *Equus* to be 4.0-4.5 million years before present (Myr BP) and gave an estimated divergence time between the New and Old World CH of 1 Myr BP. This estimate places their split into the end of the Early Pleistocene, which agrees with the first findings of CH in Eurasia around 700 kyr BP (Eisenmann, 1992). The Thistle Creek horse genome is close to the age of the MRCA of the CH and predates the Holocene BLB inundation; it is thus not

informative for demographic reconstructions of the Late Pleistocene New World caballoid horses. Moreover, the genome has been recovered to only 1-fold coverage, which is too shallow for reliable estimates of post-divergence gene flow and migration rates.

Although fossil deposits of Pleistocene CH are extremely abundant in the paleontological record, the relationship between the Old and New World populations after their split remains unclear. Previous genetic research has hinted that North American and Eurasian CH lineages were not isolated (Barrón-Ortiz et al., 2017; Vilstrup et al., 2013; Weinstock et al., 2005). The phylogenetic analysis of a short mitochondrial control region identified two major mitochondrial clades corresponding to the Eastern and Western Beringia, with a few Alaskan specimens falling into the Eurasian clade, suggesting a broad distribution of CH populations in Beringia (Weinstock et al., 2005). Another study by Barrón-Ortiz et al. (2017) failed to identify clear groups within mitochondrial lineages, although they did find differences in CH tooth morphology indicative of geographic variation (Barrón-Ortiz et al., 2017; Weinstock et al., 2005). This result, however, is based on a short region of a mitochondrial genome which has a limited resolution for determining the population structure of a recently diverged group. Only a few complete mitochondrial genomes have been reconstructed for the New World CH, which is hardly representative of the vast North American horse population (Heintzman et al., 2017).

Here, we present two high coverage nuclear genomes of North American Equids - the first genomic data for the Late Pleistocene New World caballoid horses.

Using these and previously published high coverage genomes of extinct and present-day horses, we apply coalescent approaches to reconstruct and compare the demography of Eurasian and North American CH. We estimate migration rates, admixture signals, and divergence times between Western and Eastern BLB horse populations. In addition to the nuclear genomes, we reconstruct new complete mitochondrial genomes for several dozen bone subfossil specimens from the Northern Hemisphere and resolve phylogenetic relationships within a large diversity of CH maternal lineages. Finally, we synthesize the results into a comprehensive framework of caballoid horse evolution in the last 5 million years and discuss the role of Beringia in horse evolutionary history.

MATERIALS AND METHODS

Sample collection. We assembled a collection of bone fossils from Russian Far North and Far East, Eastern Asia, Alaska, Yukon and Jawdropper Cave, Idaho (Figure 2.1A, Table 2.1). Sources of high coverage nuclear genomes: Klondike30k (UP10.MS258/YG303.325) and Klondike33k (CGG_1_011850/YT03-40) were found on Irish Gulch Klondike, Yukon Territory. DNA was recovered mostly from fossil metapodial bones, teeth, and petrosal bones, but also from morphologically non-diagnostic bone fragments. For each specimen, after cleaning the surface of the bone with a Dremel, we collected ~1g of bone powder and subdivided it for radiocarbon and genetic analyses. Following standard protocols for working with degraded DNA (Fulton & Shapiro, 2019), we conducted all procedures prior to PCR amplification in a sterile facility at the University of California, Santa Cruz Paleogenomics Laboratory.

Processing of the Klondike33k was done at the Center for GeoGenetics – University of Copenhagen. We sent the remaining bone powder to Lawrence Livermore National Laboratory and KECK AMS facilities for radiocarbon dating. All dates reported in the current study were calibrated with IntCal13 99.7% probability range and given as calibrated years before present (cal BP) ± sigma error (Ramsey, 2009; Reimer et al., 2013).

High coverage genomes. We made DNA extracts from ~0.12g of bone powder following (Dabney & Meyer, 2019; Korlević & Meyer, 2019), constructed dual-stranded DNA libraries from these extracts following Meyer & Kircher (2010), and sequenced the libraries on the Illumina HiSeq-2500 platform using paired-end 2x100 and 2x150 technologies. Illumina reads were trimmed of adapter sequences and collapsed when at least 10 nucleotides overlapped using SeqPrep2 (http://github.com/jstjohn/SeqPrep). Using the same program, we filtered out reads that were shorter than 30bp after collapsing and adapter trimming. Following trimming, sequencing reads were mapped with BWA aln algorithm against the horse reference genome EquCab2.0 (GenBank GCA_000002305.1 (Wade et al., 2009)), with a minimum read mapping quality of 20 and seed disabled (Li & Durbin, 2009; Schubert et al., 2012) (Appendix figure S1). Duplicated reads were removed with samtools rmdup (Li, 2011). We realigned reads around indels using GATK IndelRealigner and recalibrated quality scores with MapDamage (Jónsson et al., 2013; McKenna et al., 2010). Other high coverage genomes analyzed in this study were published previously: Batagai (Batagai5k) in (Librado et al., 2015); CGG10022 (Taymyr43k), CGG10023

and Przewalski's horse in (Der Sarkissian et al., 2015; Orlando et al., 2013; Schubert et al., 2014); Twilight (Thoroughbred) in (Wade et al., 2009); Donkey (Willy) in (Orlando et al., 2013).

Mitochondrial enrichment, assembly, and phylogenetic inference. For specimens used to reconstruct mitochondrial genomes, the same DNA recovery steps were performed as outlined in the previous section, except for ancient horse fossils from Eastern Asia that we processed using single-stranded DNA library method similar to Troll et al. (2019), but modified for ancient DNA (Vershinina, Kapp et al., in prep, Appendix table S1). All specimens were sequenced with 1-2 million Illumina reads to assess the quality of DNA preservation. Mitochondrial genomes of Klondike30k and Klondike33k were assembled using shotgun reads that mapped to a mitochondrial genome of a domestic horse GenBank ID NC_001640 (Xu & Arnason, 1994). All other whole mitochondrial genomes were reconstructed using in-solution hybridization enrichment with Arbor Biosciences MYTObaits RNA-oligo kit following manufacturer's protocol v. 4 (previously MYcroarray protocols v. 2 and 3). DNA was hybridized with probes at 65°C for 36 hours following (Vershinina et al., 2019). Enriched libraries were sequenced on Illumina MiSeq 2x75 runs and processed using mito_assembly_pipeline.sh script from Vershinina et al. (2019). To assemble the mitochondrial genome, we used MIA (https://github.com/mpieva/mapping-iterative-assembler) to call bases with at least 3X independent read coverage and > 60% consensus among reads to avoid false nucleotide calls due to ancient DNA damage. We removed samples from the analysis if the assembled mitochondrial genome had >20%

missing nucleotides (Appendix table S1). To construct the alignment, we added ancient and present-day horse mitochondrial genomes previously published in GenBank to the new dataset (Appendix table S2). We converted mitochondrial genomes published in (Fages et al., 2019) as BAM files back to reads with bedtools bamtofastq (Quinlan & Hall, 2010), which we then re-assembled into mitochondrial genomes using MIA (Appendix table S2). We aligned the final mitochondrial dataset with MAFFT v.7 (Nakamura et al., 2018) and manually inspected the alignment for discrepancies. Since the control region of the horse mitochondrial genome is highly variable and difficult to assemble (Xu & Arnason, 1994), mitochondrial genome assemblies in this region were enriched in errors and missing data. We explored different filtering schemes for the control region of the alignment and performed downstream analysis excluding the fragment stretching from 16,167bp to 16,475bp. For Maximum Likelihood (ML) reconstructions, we used donkey *E. asinus* GenBank ID NC_001788 (Xu et al., 1996) as an outgroup and ran three instances of RAxML v.8.2.4 (Stamatakis, 2014) using the GTRGAMMAI nucleotide model. We selected the best ML tree and estimated branch supports using 500 bootstrap iterations. We visualized the final tree in the iTOL online tree viewer (Letunic & Bork, 2007).

Genotyping. We used three variant callers to genotype high coverage genomes: bcftools v. 1.1 (Li, 2011), AntCaller v. 1.1 (Zhou et al., 2017) and GATK v. 3.7 HaplotypeCaller (McKenna et al., 2010). Each variant callset was filtered for a minimum of 20 phred base and mapping quality cut-offs. We intersected three callsets with vcftools vcf-isec and used only variants that were called by all three programs for

downstream analysis (Danecek et al., 2011). Each final variant set was additionally filtered: SNPs within 5bp of indels were removed, a minimum depth of coverage was set to 5 reads, and a maximum depth of coverage was set to 0.995 quantiles of coverage distribution for each individual genome.

Putatively neutral loci for G-PhoCS analysis. Given the overlapping ranges of both Old and New World caballoid horse populations in Beringia, and signal of migrations in their maternal lineages (Weinstock et al., 2005), we attempted to use their high coverage genomes to infer demographic history of caballoid horses in the presence of gene flow. We used the Bayesian coalescent approach implemented in G-PhoCS on a panel of North American and previously published domestic, feral, and ancient Siberian horse genomes to estimate divergence times and migration rates between the continents (Table 2.1, Figure 2.2). We pseudohaplodized each genome BAM alignment using ANGSD -doFasta incorporating IUPAC codes for heterozygous sites, which allows G-PhoCS to take diploidy into account and iterate over all possible phases of the genome (Gronau et al., 2011; Korneliussen et al., 2014). G-PhoCS reconstructs the coalescent process for each neutral locus assuming that recombination and selection within them is negligible (Gronau et al., 2011). To ascertain putatively neutral autosomal loci we took advantage of publicly available annotations of the EquCab2 genome: Ensembl (Build 92), UCSC Genome Browser, Duplicated Genes and RepeatMasker Databases, and model-based CpG islands (Ouedraogo et al., 2012; Smit et al., 2013-2015; Wu et al., 2010). Assuming that regions of the genome prone to false-positive allele calls have a high SNP density, we used the created variant callset to

detect genomic intervals with 2 or more SNPs near each other within 5, 10 and 50bp windows. Using bedtools, we created a coordinate "filter" file that consisted of genome coordinates: exons, genome assembly gaps, tandem and simple repeats, paralogs, CpG islands, and clustered SNP that were excluded from the analysis (Quinlan & Hall, 2010). All excluded regions were extended to 500bp on both sides to account for selection and genome mis-assemblies near them (Vitti et al., 2013). Using a custom python script, we ascertained a random set of 1kb loci that would satisfy three criteria: (1) no overlap with the "filter" file coordinates; (2) a minimum of 30kb interlocus distance to avoid linkage disequilibrium; (3) maximum 10% of missing nucleotides in each pseudohaploid locus. This resulted in 4215 putatively neutral loci scattered across 31 horse autosomes. All scripts, as well as filtering criteria that were used to ascertain the set of loci, are published on github https://github.com/avershinina/BLB_horses. The final "filter" file is available on Data Dryad (Reviewer'sURL https://datadryad.org/stash/share/ULWUXqNoB1oqqgwYblVaitZTSFzKi605QiWlktpu5F8). To verify if the ascertained set of loci have phylogenetic signal, we concatenated them into an alignment of 7 pseudohaploidized genomes including the donkey outgroup and reconstructed neighbor-joining phylogeny with HKY model in Geneious Prime® 2020.0.5 using 500 bootstrap-replicates (Appendix figure S2).

Demographic inference. We ran G-PhoCS on a set of putatively neutral loci ascertained from 7 genomes including the outgroup. Since the program is extremely computationally intensive, we limited each model to only four samples: an outgroup donkey genome and a combination of ancient and present-day genomes (1 present-day

horse, 1 ancient Eurasian and 1 ancient North American horse genome per run, Figure 2.2A). This resulted in 8 models of the tree topology. Each model was run through the MCMC process three times to ensure a sufficient effective sampling size. In each run, we modeled three bi-directional migration bands: m1 - gene flow after the split into the North American and Eurasian populations; m2 - gene flow between North American cabaloid horse and present-day horses; m3 - gene flow between North American horses and extinct ancient Eurasian lineage. We ran each MCMC chain for 10 million steps sampled every 10 iterations skipping the first 5% steps as burn-in and allowed automatic fine-tuning of the parameters for the first 1000 steps. We incorporated the following priors into the analysis in the form of a gamma-distribution with β = 20,000: ages (τ) of the present-day samples were set to 0, ages of the ancient samples were set to their radiocarbon dates, the age of the root was calibrated to the MRCA of *Equus* 4.5 Myr BP years ago, the MRCA of caballoid horses was set to 1.0 Myr BP (Heintzman et al., 2017; Orlando et al., 2013), and the split between Eurasian populations was set to the younger boundary of 0.127 Myr BP following Schubert et al. (2014). Posterior distributions of the parameters were re-calculated to divergence times in units of years (T_{div}) and effective population sizes (N_e) by scaling with a per-generation mutation rate (μ) of $7.24*10^{-9}$ and a generation time (g) of 8 years (Orlando et al., 2013). We followed these equations to scale the parameters: $N_e=\theta/4\mu$, $T_{div}= \tau g/\mu$. The G-PhoCS migration band is modeled as a migration rate between two lineages over the period when both lineages existed. Thus we calculated the total migration rate m_{total} by scaling the per generation migration rate $m_{A\rightarrow B}$ by the duration of the corresponding migration bands $\tau_{A\rightarrow B}$

(the length of the entire branch when both A and B populations existed). The m_{total} estimate is free of the assumption about the mutation rate (Gronau et al., 2011). Estimates of m_{total} approximate the probability that a locus sampled in a target population originated in the source population. To account for uncertainty in μ and g, we modeled uncertainty around their reported values and randomly subsampled these parameters from their gamma distributions while calibrating the G-PhoCS results (Appendix figure S4). We monitored convergence of the MCMC chains with Tracer v. 1.7.1 (Rambaut et al., 2018). For PSMC analysis, we generated a pseudohaploid consensus sequence of autosomes with mpileup using -EA options and -C50, following fq2psmcfa -q20 (Li et al., 2009). We ran PSMC with default parameters, -N25 -t15 -r5 -p "4+25*2+4+6", and repeated the analysis with 100 bootstrap replicates (Li & Durbin, 2011).

Admixture estimates using D-statistics. We used the ABBA\BABA test, also known as D statistic (Durand et al., 2011; Green et al., 2010) to estimate the excess of shared sites between one of the two closely related lineages (P1 and P2) and a possible introgressor (P3). In a tree topology ((P1,P2),P3), outgroup) where an ancestral allele is A and a derived allele is B, an excess of BABA sites would indicate allele sharing between P1 and a P3 while an excess of ABBA sites would indicate allele sharing between P2 and P3. We used ANGSD -doAbbababa in 5Mb chromosome blocks to perform this analysis on autosomes (Figure 2.3, Appendix figure S6) and the chromosome X (Appendix figure S7). We limited this test to positions with phred-scaled mapping and sequencing quality above 30. We also trimmed 5bp from both ends of the reads due to extremely high DNA damage in these regions. To evaluate

significance of the D-statistic, we used Z-scores estimated as D-statistic values divided by its standard error.

RESULTS

We successfully recovered DNA from 79 paleontological fossil specimens from 15 locations across the Holarctic (Figure 2.1A, Appendix table S1). Two fossils of Yukon Klondike caballoid horses contained about 70% endogenous DNA enabling us to sequence their complete genomes with up to 25-fold coverage (Appendix figures S1, S2). These are the first genomes from North American Pleistocene caballoid horses sequenced to high coverage. Other fossil specimens with endogenous DNA content higher than 5-10% were subjected to whole mitochondrial genome enrichment and assembly, resulting in 77 enriched DNA libraries and average mitochondrial genome coverage of 133.4-fold per sample (4.2-2287.0, Appendix table S1). All sequenced libraries had characteristic post-mortem aDNA damage: excesses of C>T and G>A transitions at the ends of the molecules and short DNA fragment length (<100bp). This panel of fossils samples represents the largest dataset of sequenced North American caballoid horses.

We compared newly sequenced mitochondrial genomes with 110 present-day and ancient mtDNA genomes published previously (Figure 2.1B, C, Appendix tables S1, S2, for the extended mtDNA phylogeny see Appendix figure S3). Maximum likelihood inference showed geographic phylogenetic structure, splitting all caballoid horses into North American and Eurasian clades. Interestingly, lineages of reciprocal migrants are found in both clusters: the Eurasian clade contains caballoid horses that

were found in Alaska North Slope, Interior Alaska, and Yukon Klondike areas while the North American major clade diverges into a local Eastern Beringian population and caballoid horses from Russian Far East and Eastern Asia (Appendix figure S3). The ancient samples-sources of high-coverage genomes fall into the Eurasian clade (Taymyr43k, CGG10023, and Batagai) and the North American clade (Klondike30k and Klondike33k).

Following the model presented in Figure 2.2A, we inferred the divergence time between the caballoid *Equus* and their sister species *E. asinus*. When assuming the range of mutation rates centered around 7.242×10^{-9} per site per generation, the calibration resulted into an average 4.4±1 Myr BP divergence time (Figure 2.2B) which is in line with previous estimates of 4.0-4.5 Myr BP (Heintzman et al., 2017; Jónsson et al., 2014; Orlando et al., 2013). We found that the branching of the Old and New World caballoid horse populations occurred around 0.818±0.187 Myr BP, which is slightly younger than the previous estimates of 1.062 Myr (Orlando et al., 2013) but consistent with 0.7-1 Myr BP estimate of the mitochondrial genome divergence in (Heintzman et al., 2017). For the split between the ancestors of extinct and present-day Eurasian caballoid horses, we observed an older divergence estimate when the Przewalski's horse genome was used to represent the present-day caballoid horse population (416 ±0.9 kyr BP) and a younger estimate when the Thoroughbred horse Twilight was used (346±0.8 kyr BP). The divergence time between the Eurasian lineages is in line with the 384 kyr BP estimate from (Schubert et al., 2014) and does not account for post-divergence gene flow between the extinct Eurasian horses and

ancestors of the present-day domestic horses. Since diversification of lineages is a gradual process that occurs in the presence of gene flow immediately after the split, our estimates provide the upper boundary to the MRCA of the caballoid horse populations (Edwards & Beerli, 2000).

We evaluated bi-directional gene flow across the Bering Land Bridge between the ancestors of present-day and ancient Eurasian and Yukon Klondike caballoid horses (Figure 2.2C). We estimate the first migration band (m1) connecting the two continental horse populations after their split around 818 kyr ago to be 0.04±0.8 which corresponds to around 3-4% migration probability in the Middle Pleistocene. We found very similar signals of gene flow between Old and New World caballoid horses after the Eurasian split: m2 was 1.89±0.55% and 1.91±0.60% into and from ancestors of the present-day horse lineage accordingly while m3 probability was 2.06±0.45% and 2.05±0.46% into and from the ancient Siberian horses.

With the D-statistic (ABBA\BABA test), we investigated whether the extinct New World horses (P3) share polymorphisms with Eurasian horse lineages (P1 and P2) using donkey as an outgroup (Figure 2.3, Appendix figures S6, S7). We found a significant signal of derived allele sharing between the Late Pleistocene Siberian horses Taymyr43k and CGG10023 and both Klondike caballoid horses (Figure 2.3). Among the two ancient Yukon Klondike caballoid horses, Klondike30k appears to be less admixed than Klondike 33k. In all tested configurations of P1, P2 and P3, the Batagai specimen shows the least amount of sites shared with Yukon Klondike caballoid horse genomes (Appendix figure S6). This is supported by the non-significant D-statistic

value when the ((Klondike33k, Klondike30k), Batagai5k) Donkey) topology is tested on the dataset that excludes transitions to account for ancient DNA deamination (Appendix figure S6). We found significant support for an excess of shared polymorphisms between North American horses and present-day feral and domestic horses: both Thoroughbred and Przewalski's horses have admixture with both horses from Klondike (Figure 2.3). Remarkably, a signal of admixture was more pronounced on X chromosomes than on autosomes (Appendix figure S7): the D-statistic for configuration (((Batagai5k,Taymyr43k), Klondike33k),Donkey) on the X-chromosome was D=0.23 (Z-score=8.3), while on autosomes, D=0.06 (Z-score=13.5). We found the same discrepancy across all other P1, P2, and P3 combinations. In comparison to the autosomes, the proportion of X chromosome D-statistic values having $|Z| < 3$ was higher. This is mainly due to the short size of the X chromosome compared to autosomes while using a 5Mb block size to estimate Z-scores (Cahill et al., 2015). Following these results, we conclude that the bi-directional admixture between Eurasian and North American caballoid horses occurred after their split. The D-statistic measurements are in line with the analysis of mitochondrial genomes and migration estimates of G-PhoCS.

PSMC profiles reported virtually identical trajectories of effective population size change for Klondike and ancient Siberian caballoid horses up until the last 80,000 years (Figure 2.4A,B, Appendix figure S8). After the divergence around 1 Myr BP, close to the 0.818 Myr BP split estimated in G-PhoCS, both Klondike caballoid horses as well as ancient Taymyr and Batagai horses show evidence of a population size

decline (Appendix figure S5). The PSMC trace at this divergence shows effective population size to be around $6*10^4$ ($N_e=(5.6\pm0.2)*10^4$ estimated by G-PhoCS, appendix figure S5). Subsequently, effective population sizes on both sides of the BLB increased, with the Eurasian horses showing a higher peak than the Yukon Klondike horses (Figure 2.4A,B, Appendix figure S8). Following the peak, North American caballoid horse populations significantly declined around 0.1 Myr BP while the Eurasian caballoid horses recovered during this time. G-PhoCS estimates of effective population sizes support this trend: both Klondike30k ($N_e=(3.2\pm0.1)*10^4$) and Klondike33k ($N_e=(3.0\pm0.1)*10^4$) have lower effective population size than the Taymyr horse ($N_e=(4.8\pm0.2)*10^4$). The Holocene aged Batagai5k is the most recent caballoid horse in our panel of ancient genomes; Batagai5k has the lowest effective population size, as estimated in G-PhoCS ($(1.8\pm0.1)*10^4$), in accordance with a sharp decline of its PSMC trace during the last 10,000 years and previous data of pre-domestication population bottleneck (Librado et al., 2015). We found similar estimates in present-day Przewalski and Thoroughbred horses: $(1.9\pm0.1)*10^4$ and $(1.8\pm0.1)*10^4$ accordingly.

DISCUSSION

Paleogenetic research of ancient caballoid horses was mainly focused on the domestication process, while few studies have thoroughly examined ancient North American populations from the Pleistocene or have adequately resolved their phylogenetic relationship with present-day domestic breeds (Heintzman et al., 2017; Orlando et al., 2013). Previously reconstructed phylogenies of a short fragment of mitochondrial DNA were unable to resolve the relationships between maternal lineages

with high confidence (Barrón-Ortiz et al., 2017; Vilstrup et al., 2013; Weinstock et al., 2005). Since the caballoid horse group diverged fairly recently, the short mitochondrial genome fragment may not be appropriate for the analysis of phylogenetic relationships between them. Our panel of 79 new high coverage complete mitochondrial genomes significantly improved the phylogenetic inference and yielded high support to geographic phylogenetic structure with two major clades of Eurasian (including domestic) and New World caballoid horses (Figure 2.1B,C). Using two new ancient high coverage genomes of North American horses, we further examined admixture signals between the two major clades (figures 2A,B, 3A,B). We found a signal of cross-continental gene flow and wide population dispersal of the Late Pleistocene ancestors of present-day and ancient North American caballoid horses.

We observed geographic subdivision between the Pleistocene caballoid horse populations inhabiting Eurasia and North America, however, both lineages were not isolated (Figure 2.1B). We found evidence of gene flow between Yukon Klondike caballoid horses and ancestors of present-day domestic and ancient Eurasian caballoid horses after their divergence 0.8-1 Myr ago (Figure 2.2B). Gene flow was likely bi-directional at roughly 4% total, then decreased to 2% in the Middle Pleistocene, suggesting gradual separation between the lineages despite at least the intermittent presence of the Bering Land Bridge at that time (Figure 2.2C). We note, however, that the values of bidirectional migration in m2 and m3 had extremely similar posterior distributions and further investigation revealed autocorrelation between m2 and m3 estimates. This is likely a result of the limited number of genomes used in our analysis

(only 4 per run) which we were unable to increase due to the highly-demanding computational process of G-PhoCS. The autocorrelation problem, however, does not nullify the evidence of gene flow obtained in the analysis.

We investigated whether the signal of migrations in demographic inference is supported by signatures of admixture in nuclear genomes (Figure 2.3). Incomplete lineage sorting (ILS) and admixture can lead to a derived allele being shared between two non-sister lineages. It is expected that random ILS will result in the same amount of shared derived alleles between North American and Eurasian caballoid horses (D=0). In the presence of admixture, however, D-statistics are expected to deviate from 0. We detected statistically significant signals of admixture on autosomes and the chromosome X (Figure 2.3, Appendix figures S6, S7), which further supports post-divergence gene flow between Yukon Klondike and Eurasian caballoid horses. Signal of gene flow even exists in the present-day genomes of domesticated Thoroughbred and feral Przewalski's horses.

Interestingly, we found a decrease in admixture signal with time. The Taymyr43k appears to be the most admixed genome in the panel, while CGG10023 dated to 16,099 ± 192 cal BP shares less alleles with the Klondike individuals, and a 5 kyr old Batagai does not show statistically significant sharing of derived alleles with the Klondike. Such decrease of the admixture signal can result from a negative selection against the introgressed blocks of the North American ancestry. However, according to the G-PhoCS inference, the estimated rates of gene flow are lower in tree topologies testing the ancient Taymyr horse, in comparison to higher *m* estimates rather

in Batagai. The discrepancy can be explained by an influx of alleles from an unsampled lineage into the ancestors of the Batagai horse. Future research can further explore both scenarios by fitting the site frequency spectrum of North American and Eurasian caballoid horse populations into the outlined models.

We found a signal of enriched North American caballoid horse ancestry on the X chromosome of Eurasian caballoid horses (Appendix figures S6, S7) suggesting unequal contribution of male and female individuals into the introgression process and a complex hybridization history between the two continental horse populations. This may have been caused by the biased dispersal strategies of mares compared to stallions. However, a local population density and a complex behavioral and ecological scenario may underlie this finding (Marjamäki et al., 2013). A similar sex-biased dispersal is observed in other large mammals that disperse over long distances (Cahill et al., 2015).

The relationship between Old and New World caballoid horses

The Old and New World horse clades contain individuals found on the side of the Bering Land Bridge opposite to the side hosting their ancestral populations. We found members of the Eurasian mitochondrial lineage in Alaska and members of the more archaic North American clade, including the Middle Pleistocene Thistle Creek horse, in the Russian Far East. Two scenarios could explain this distribution. The first scenario assumes the existence of two large continental populations that overlapped in Beringia. The second one is explained by the existence of the single large Holarctic meta-population. We discuss both scenarios below.

Hypothesis (1): Two partially sympatric, but genetically isolated caballine horse populations. Several observations support this hypothesis. Specimens AV109 (the Eurasian red clade) and AV193 (the North American blue clade), dated to 19,916±103 and 23,350±134 cal years BP accordingly, are found in interior Alaska near Cripple Creek, Fairbanks area (Appendix figure S3). Their close radiocarbon ages suggest not only spatial, but also temporal overlap of populations they belonged to during the Last Glacial Maximum. An analogous example of sympatry in Eurasia is found in the North of Yakutiya, on the North Siberian Islands: JW28/MS298 (dated to 32,918±1140 cal years BP) and JW26/MS300 (dated to 20,272±213 cal years BP) (Appendix figure S3). Both samples came from the same locality but belong to separate mitochondrial clades. Remarkably, the latter observation explains the signal of a "ghost" genetic ancestry found in Fages et al. (2019). None of the North American samples were taken into the analysis of horse domestication in Fages et al. (2019), therefore, these same specimens from the New Siberian Islands had unclear genetic affinities. On the Eurasian continent, caballoid horses from East Asian localities, such as Jilin and Primorye, as well as horses from Chukotka (Alazeya and Indigirka rivers), compose a clade phylogenetically close to the archaic North American haplogroup, which indicates a possible second expansion "Out of North America" that happened after the initial split of the Old World caballoid horses 0.8-1 Myr ago. As a result of this dispersal, the North American caballoid horse populations stretched farther to the west and coexisted with the members of the Eurasian clade, at least in the Indigirka river region. Furthermore, values of D-statistic and migration rates show little allele

sharing between Yukon Klondike and Siberian horses which is on par with estimates of the cross-species gene flow between other equids (Jónsson et al., 2014). We note, however, that our D-statistic analyses omits deaminated sites due to ancient DNA damage and we were unable to recover complete nuclear genomes of the samples with haplogroup\locality discrepancy. Further research should be focused on extending the dataset of the New World horse lineage to fully test the described hypothesis.

Hypothesis (2): Large admixed meta-population. A second hypothesis that could explain the reciprocal genetic affinities of the Pleistocene caballoid horses is the existence of a large meta-population that stretched across the Northern Hemisphere, enabling constant gene flow and creating a continuum of ancestry across both sides of the Bering Land Bridge. In this case, the split of caballoid horses into "North American" and "Eurasian" is artificial since the meta-population was spread cross-continentally. This is supported by previous morphometric analysis of the Late Pleistocene caballoid horse cheek teeth which showed little differences among various local North American horse populations likely explained by local ecological adaptations (Barrón-Ortiz et al., 2017). Given this hypothesis, either hybridization or incomplete lineage sorting between the local subpopulations would explain what appears to be a bi-directional exchange of maternal lineages. In the case of the mitochondrial introgression via several hybridization events, the North American haplotype would be acquired by the ancestor of populations inhabiting Primorye and Eastern Asia, while horses from Alaska would possess the Siberian maternal ancestry. Both hybridization and ILS would lead to a discrepancy between patterns of nuclear

and mitochondrial ancestry (Nichols, 2001). A similar discrepancy is found in European bison, *Bison bonasus*, whose mitochondrial genome is close to cattle, *Bos taurus*, while its nuclear genome phylogeny places it as a sister species to the American bison, *B. bison,* with whom it is cross-fertile (Massilani et al., 2016; Soubrier et al., 2016; Wang et al., 2018). Given the recent and rapid diversification of caballoid horses, ILS may be a more likely explanation of the observed pattern than mitochondrial introgression. Due to the lack of nuclear markers sampled from extinct horse populations inhabiting the entire Holarctic region, our analyses have limited power to resolve the question of continuity. A broader population genetic study combined with morphological investigation of East Asian and Alaskan specimens will enable a more thorough test of the proposed hypothesis.

CONCLUSION

The role of Beringia in evolution of Late Pleistocene caballoid horses

Our results emphasize the role of Beringia in cross-continental dispersals during the Pleistocene. Beringia was a large landmass during Pleistocene glacial periods that facilitated various biotic exchanges by linking Eurasia and North America (Elias & Crocker, 2008; Zimov et al., 1995). When present, the BLB played a key role as a Quaternary dispersal corridor for bison, mammoth, and humans, whose populations dispersed from Eastern Siberia into the New World and for horse populations expanding in the opposite westward direction (Debruyne et al., 2008; Scott A. Elias & Crocker, 2008; Meiri et al., 2014; Pedersen et al., 2016; Shapiro et al., 2004). Genetically different caballoid horse populations coexisted in Beringia and gene flow

between them demonstrates that the BLB acted as a critical contact zone (figures 1B, 4B). The results of the current study demonstrate how the vast region of Beringia was not solely a biotic dispersal corridor between continents, but rather, a critical ecoregion where significant evolutionary processes unfolded for many Holarctic cold-adapted taxa.

Horses thrived on the Mammoth steppe, the widespread steppe-tundra biome which characterized much of the Holarctic, during cold periods of the Pleistocene (Guthrie, 1982; Zimov et al., 2012). The productive habitats spanning Beringia likely allowed them to expand into the Old World and served as a ground for rapid lineage diversification and dispersal in the last 1 Myr years. We found that a reduction in effective population size of Yukon Klondike horses was more dramatic than that in Taymyr (figures 4A,B, Appendix figure S5), further supporting a sharp demographic decline that was likely exacerbated by the BLB inundation 11 kyr ago and transformation of the Mammoth steppe into shrub-tundra, which was less habitable for horses than the steppe-tundra ecosystems (Mann et al., 2015, 2018; Oswald et al., 2003). Present-day global climatic change and human impact fragment and transform habitats across the globe. Lessons learned from the critical role of the contact zones and dispersal corridors in the Pleistocene ecosystems have a potential to greatly improve future conservation strategies and should be explored in the future research.

BIBLIOGRAPHY

Barnosky, A. D., Koch, P. L., Feranec, R. S., Wing, S. L., & Shabel, A. B. (2004). Assessing the causes of late Pleistocene extinctions on the continents. *Science*, *306*(5693), 70–75.

Barron-Ortiz, C., Avilla, L., Jass, C., Bravo-Cuevas, V., Machado, H., & Mothé, D. (2019). What is Equus? Reconciling taxonomy and phylogenetic analyses. *Frontiers in Ecology and Evolution*, *7*, 343.

Barrón-Ortiz, C. I., Rodrigues, A. T., Theodor, J. M., Kooyman, B. P., Yang, D. Y., & Speller, C. F. (2017). Cheek tooth morphology and ancient mitochondrial DNA of late Pleistocene horses from the western interior of North America: Implications for the taxonomy of North American Late Pleistocene Equus. *PloS One*, *12*(8), e0183045.

Cahill, J. A., Stirling, I., Kistler, L., Salamzade, R., Ersmark, E., Fulton, T. L., Stiller, M., Green, R. E., & Shapiro, B. (2015). Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Molecular Ecology*, *24*(6), 1205–1217.

Dabney, J., & Meyer, M. (2019). Extraction of Highly Degraded DNA from Ancient Bones and Teeth. In B. Shapiro, A. Barlow, P. D. Heintzman, M. Hofreiter, J. L. A. Paijmans, & A. E. R. Soares (Eds.), *Ancient DNA: Methods and Protocols* (pp. 25–29). Springer New York.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics* , *27*(15), 2156–2158.

Debruyne, R., Chu, G., King, C. E., Bos, K., Kuch, M., Schwarz, C., Szpak, P., Gröcke, D. R., Matheus, P., Zazula, G., Guthrie, D., Froese, D., Buigues, B., de Marliave, C., Flemming, C., Poinar, D., Fisher, D., Southon, J., Tikhonov, A. N., … Poinar, H. N. (2008). Out of America: ancient DNA evidence for a new world origin of late quaternary woolly mammoths. *Current Biology: CB*, *18*(17), 1320–1326.

Der Sarkissian, C., Ermini, L., Schubert, M., Yang, M. A., Librado, P., Fumagalli, M., Jónsson, H., Bar-Gal, G. K., Albrechtsen, A., Vieira, F. G., Petersen, B., Ginolhac, A., Seguin-Orlando, A., Magnussen, K., Fages, A., Gamba, C., Lorente-Galdos, B., Polani, S., Steiner, C., … Orlando, L. (2015). Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Current Biology: CB*, *25*(19), 2577–2583.

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*(8), 2239–2252.

Edwards, S. V., & Beerli, P. (2000). Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution; International Journal of Organic Evolution*, *54*(6), 1839–1854.

Eisenmann, V. (1992). Origins, dispersals, and migrations of Equus (Mammalia, Perissodactyla). *Courier Forschungsinstitut Senckenberg*, *153*, 161–170.

Elias, S. A., & Brigham-Grette, J. (2013). Late Pleistocene Glacial Events in Beringia. In S. A. Elias & C. J. Mock (Eds.), *Encyclopedia of Quaternary Science (Second Edition)* (pp. 191–201). Elsevier.

Elias, S. A., & Crocker, B. (2008). The Bering Land Bridge: a moisture barrier to the dispersal of steppe–tundra biota? *Quaternary Science Reviews*, *27*(27), 2473–2483.

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K. A. S., Albizuri, S., Alfarhan, A. H., Allentoft, M., Alquraishi, S., Anthony, D., Baimukhanov, N., Barrett, J. H., Bayarsaikhan, J., Benecke, N., Bernáldez-Sánchez, E., … Orlando, L. (2019). Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell*, *177*(6), 1419–1435.e31.

Fulton, T. L., & Shapiro, B. (2019). Setting Up an Ancient DNA Laboratory. In B. Shapiro, A. Barlow, P. D. Heintzman, M. Hofreiter, J. L. A. Paijmans, & A. E. R. Soares (Eds.), *Ancient DNA: Methods and Protocols* (pp. 1–13). Springer New York.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*(5979), 710–722.

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, *43*(10), 1031–1034.

Guthrie, R. D. (1982). Mammals of the mammoth steppe as paleoenvironmental indicators. In D. M. Hopkins, J. V. Matthews, C. E. Schweger, & S. B. Young (Eds.), *Paleoecology of Beringia* (pp. 307–326). Academic Press.

Guthrie, R. D. (2003). Rapid body size decline in Alaskan Pleistocene horses before extinction. *Nature*, *426*(6963), 169–171.

Guthrie, R. D. (2006). New carbon dates link climatic change with human colonization and Pleistocene extinctions. *Nature*, *441*(7090), 207–209.

Heintzman, P. D., Zazula, G. D., MacPhee, R. D. E., Scott, E., Cahill, J. A., McHorse, B. K., Kapp, J. D., Stiller, M., Wooller, M. J., Orlando, L., Southon, J., Froese, D. G., & Shapiro, B. (2017). A new genus of horse from Pleistocene North America. *eLife*, *6*. https://doi.org/10.7554/eLife.29944

Jakobsson, M., Pearce, C., Cronin, T. M., Backman, J., Anderson, L. G., Barrientos, N., Björk, G., Coxall, H., Boer, A. de, Mayer, L. A., Mörth, C.-M., Nilsson, J., Rattray, J. E., Stranne, C., Semiletov, I., & O'Regan, M. (2017). Post-glacial flooding of the Bering Land Bridge dated to 11 cal ka BP based on new geophysical and sediment records. *Climate of the Past*, *13*(8), 991–1005.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* , *29*(13), 1682–1684.

Jónsson, H., Schubert, M., Seguin-Orlando, A., Ginolhac, A., Petersen, L., Fumagalli, M., Albrechtsen, A., Petersen, B., Korneliussen, T. S., Vilstrup, J. T., Lear, T., Myka, J. L., Lundquist, J., Miller, D. C., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A. S., Stagegaard, J., Strauss, G., … Orlando, L. (2014). Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), 18655–18660.

Koch, P. L., & Barnosky, A. D. (2006). Late Quaternary Extinctions: State of the Debate. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 215–250.

Korlević, P., & Meyer, M. (2019). Pretreatment: Removing DNA Contamination from Ancient Bones and Teeth Using Sodium Hypochlorite and Phosphate. In B. Shapiro, A. Barlow, P. D. Heintzman, M. Hofreiter, J. L. A. Paijmans, & A. E. R. Soares (Eds.), *Ancient DNA: Methods and Protocols* (pp. 15–19). Springer New York.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*, 356.

Letunic, I., & Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* , *23*(1), 127–128.

Librado, P., Der Sarkissian, C., Ermini, L., Schubert, M., Jónsson, H., Albrechtsen, A., Fumagalli, M., Yang, M. A., Gamba, C., Seguin-Orlando, A., Mortensen, C. D., Petersen, B., Hoover, C. A., Lorente-Galdos, B., Nedoluzhko, A., Boulygina, E., Tsygankova, S., Neuditschko, M., Jagannathan, V., … Orlando, L. (2015). Tracking the origins of Yakutian horses and the genetic basis for their fast

adaptation to subarctic environments. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(50), E6889–E6897.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* , *27*(21), 2987–2993.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079.

Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., Ugan, A., Borregaard, M. K., Gilbert, M. T. P., Nielsen, R., Ho, S. Y. W., Goebel, T., Graf, K. E., Byers, D., Stenderup, J. T., Rasmussen, M., Campos, P. F., Leonard, J. A., Koepfli, K.-P., … Willerslev, E. (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, *479*(7373), 359–364.

MacFadden, B. J. (2005). Fossil horses - evidence for evolution. *Science*, *307*(5716), 1728–1730.

Mann, D. H., Groves, P., Gaglioti, B. V., & Shapiro, B. A. (2018). Climate-driven ecological stability as a globally shared cause of Late Quaternary megafaunal extinctions: the Plaids and Stripes Hypothesis: A hypothesis about megafaunal extinctions. *Biological Reviews*, *400*, 740.

Mann, D. H., Groves, P., Reanier, R. E., Gaglioti, B. V., Kunz, M. L., & Shapiro, B. (2015). Life and extinction of megafauna in the ice-age Arctic. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(46), 14301–14306.

Marjamäki, P. H., Contasti, A. L., Coulson, T. N., & McLoughlin, P. D. (2013). Local density and group size interacts with age and sex to determine direction and rate of social dispersal in a polygynous mammal. *Ecology and Evolution*, *3*(9), 3073–3082.

Massilani, D., Guimaraes, S., Brugal, J.-P., Bennett, E. A., Tokarska, M., Arbogast, R.-M., Baryshnikov, G., Boeskorov, G., Castel, J.-C., Davydov, S., Madelaine, S., Putelat, O., Spasskaya, N. N., Uerpmann, H.-P., Grange, T., & Geigl, E.-M.

(2016). Past climate changes, population dynamics and the origin of Bison in Europe. *BMC Biology*, *14*(1), 93.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

Meiri, M., Lister, A. M., Collins, M. J., Tuross, N., Goebel, T., Blockley, S., Zazula, G. D., van Doorn, N., Dale Guthrie, R., Boeskorov, G. G., Baryshnikov, G. F., Sher, A., & Barnes, I. (2014). Faunal record identifies Bering isthmus conditions as constraint to end-Pleistocene migration to the New World. *Proceedings. Biological Sciences / The Royal Society*, *281*(1776), 20132167.

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, *2010*(6), db.prot5448.

Nakamura, T., Yamada, K. D., Tomii, K., & Katoh, K. (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* , *34*(14), 2490–2492.

Nichols, R. (2001). Gene trees and species trees are not the same. *Trends in Ecology & Evolution*, *16*(7), 358–364.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P. L. F., Fumagalli, M., Vilstrup, J. T., Raghavan, M., Korneliussen, T., Malaspinas, A.-S., Vogt, J., Szklarczyk, D., Kelstrup, C. D., … Willerslev, E. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, *499*(7456), 74–78.

Oswald, W. W., Brubaker, L. B., Hu, F. S., & Gavin, D. G. (2003). Pollen-vegetation calibration for tundra communities in the Arctic Foothills, northern Alaska: Pollen-vegetation calibration for arctic tundra. *The Journal of Ecology*, *91*(6), 1022–1033.

Ouedraogo, M., Bettembourg, C., Bretaudeau, A., Sallou, O., Diot, C., Demeure, O., & Lecerf, F. (2012). The duplicated genes database: identification and functional annotation of co-localised duplicated genes across genomes. *PloS One*, *7*(11), e50653.

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., Mendoza, M. L. Z., Beaudoin, A. B., Zutter, C., Larsen, N. K., Potter, B. A., Nielsen, R., Rainville, R. A., Orlando, L., Meltzer, D. J., Kjær, K. H., & Willerslev, E. (2016). Postglacial viability and colonization in North America's ice-free corridor. *Nature*, *537*(7618), 45–49.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* , *26*(6), 841–842.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904.

Ramsey, C. B. (2009). Bayesian Analysis of Radiocarbon Dates. *Radiocarbon*, *51*(1), 337–360.

Reimer, P. J., Bard, E., Bayliss, A., Warren Beck, J., Blackwell, P. G., Ramsey, C. B., Buck, C. E., Cheng, H., Lawrence Edwards, R., Friedrich, M., Grootes, P. M., Guilderson, T. P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T. J., Hoffmann, D. L., Hogg, A. G., Hughen, K. A., … van der Plicht, J. (2013). IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon*, *55*(4), 1869–1887.

Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J. F., Al-Rasheid, K. A. S., Willerslev, E., Krogh, A., & Orlando, L. (2012). Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics*, *13*, 178.

Schubert, M., Jónsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., Albrechtsen, A., Dupanloup, I., Foucal, A., Petersen, B., Fumagalli, M., Raghavan, M., Seguin-Orlando, A., Korneliussen, T. S., Velazquez, A. M. V., Stenderup, J., Hoover, C. A., Rubin, C.-J., Alfarhan, A. H., … Orlando, L. (2014). Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), E5661–E5669.

Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., Gilbert, M. T. P., Barnes, I., Binladen, J., Willerslev, E., Hansen, A. J., Baryshnikov, G. F., Burns, J. A., Davydov, S., Driver, J. C., Froese, D. G., Harington, C. R., Keddie, G., … Cooper, A. (2004). Rise and fall of the Beringian steppe bison. *Science*, *306*(5701), 1561–1565.

Smit, A. F. A., Hubley, R., & Green, P. (2013-2015). *RepeatMasker Open-4.0*. http://www.repeatmasker.org

Soubrier, J., Gower, G., Chen, K., Richards, S. M., Llamas, B., Mitchell, K. J., Ho, S. Y. W., Kosintsev, P., Lee, M. S. Y., Baryshnikov, G., Bollongino, R., Bover, P., Burger, J., Chivall, D., Crégut-Bonnoure, E., Decker, J. E., Doronichev, V. B., Douka, K., Fordham, D. A., … Cooper, A. (2016). Early cave art and ancient DNA record the origin of European bison. *Nature Communications*, *7*, 13158.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* , *30*(9), 1312–1313.

Troll, C. J., Kapp, J., Rao, V., Harkins, K. M., Cole, C., Naughton, C., Morgan, J. M., Shapiro, B., & Green, R. E. (2019). A ligation-based single-stranded library preparation method to analyze cell-free DNA and synthetic oligos. *BMC Genomics*, *20*(1), 1023.

Vershinina, A. O., Kapp, J. D., Baryshnikov, G. F., & Shapiro, B. (2019). The case of an arctic wild ass highlights the utility of ancient DNA for validating problematic identifications in museum collections. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13130

Vilstrup, J. T., Seguin-Orlando, A., Stiller, M., Ginolhac, A., Raghavan, M., Nielsen, S. C. A., Weinstock, J., Froese, D., Vasiliev, S. K., Ovodov, N. D., Clary, J., Helgen, K. M., Fleischer, R. C., Cooper, A., Shapiro, B., & Orlando, L. (2013). Mitochondrial phylogenomics of modern and ancient equids. *PloS One*, *8*(2), e55950.

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annual Review of Genetics*, *47*, 97–120.

Wade, C. M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T. L., Adelson, D. L., Bailey, E., Bellone, R. R., Blöcker, H., Distl, O., Edgar, R. C., Garber, M., Leeb, T., Mauceli, E., MacLeod, J. N., Penedo, M. C. T., Raison, J. M., … Lindblad-Toh, K. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science*, *326*(5954), 865–867.

Wang, K., Lenstra, J. A., Liu, L., Hu, Q., Ma, T., Qiu, Q., & Liu, J. (2018). Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. *Communications Biology*, *1*, 169.

Weinstock, J., Willerslev, E., Sher, A., Tong, W., Ho, S. Y. W., Rubenstein, D., Storer, J., Burns, J., Martin, L., Bravi, C., Prieto, A., Froese, D., Scott, E., Xulong, L., & Cooper, A. (2005). Evolution, systematics, and phylogeography of pleistocene horses in the new world: a molecular perspective. *PLoS Biology*, *3*(8), e241.

Wu, H., Caffo, B., Jaffee, H. A., Irizarry, R. A., & Feinberg, A. P. (2010). Redefining CpG islands using hidden Markov models. *Biostatistics* , *11*(3), 499–514.

Xu, X., & Arnason, U. (1994). The complete mitochondrial DNA sequence of the horse, Equus caballus: extensive heteroplasmy of the control region. *Gene*, *148*(2), 357–362.

Xu, X., Gullberg, A., & Arnason, U. (1996). The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA comparisons among four closely related mammalian species-pairs. *Journal of Molecular Evolution*, *43*(5), 438–446.

Zhou, B., Wen, S., Wang, L., Jin, L., Li, H., & Zhang, H. (2017). AntCaller: an accurate variant caller incorporating ancient DNA damage. *Molecular Genetics and Genomics: MGG*, *292*(6), 1419–1430.

Zimov, S. A., Chuprynin, V. I., Oreshko, A. P., Chapin, F. S., Reynolds, J. F., & Chapin, M. C. (1995). Steppe-Tundra Transition: A Herbivore-Driven Biome Shift at the End of the Pleistocene. *The American Naturalist*, *146*(5), 765–794.

Zimov, S. A., Zimov, N. S., Tikhonov, A. N., & Chapin, F. S. (2012). Mammoth steppe: a high-productivity phenomenon. *Quaternary Science Reviews*, *57*, 26–45.

Table 2.1. Localities, dates, and mapping statistics of high coverage ancient genomes used for demographic reconstruction. Radiocarbon dates are calibrated with IntCal13 as implemented in OxCal 4.3 (Ramsey, 2009; Reimer et al., 2013). Median and sigma values for 99.7 calibration probability range are reported.

Specimen ID	**Sample Name as used in the current study**	**Nuclear Genome Coverage (fold)**	**Calibrated radiocarbon age (cal BP)**	**Reference**
Batagai	Batagai5k	18.3	5,104 ± 109	Librado et al., 2015
CGG10022	Taymyr43k	24.3	42,692 ± 891	Schubert et al., 2014
YG303.325	Klondike30k	24.8	30,318 ± 245	This study
CGG_1_011850 (YT03-40)	Klondike33k	25.1	32,703 ± 334	This study

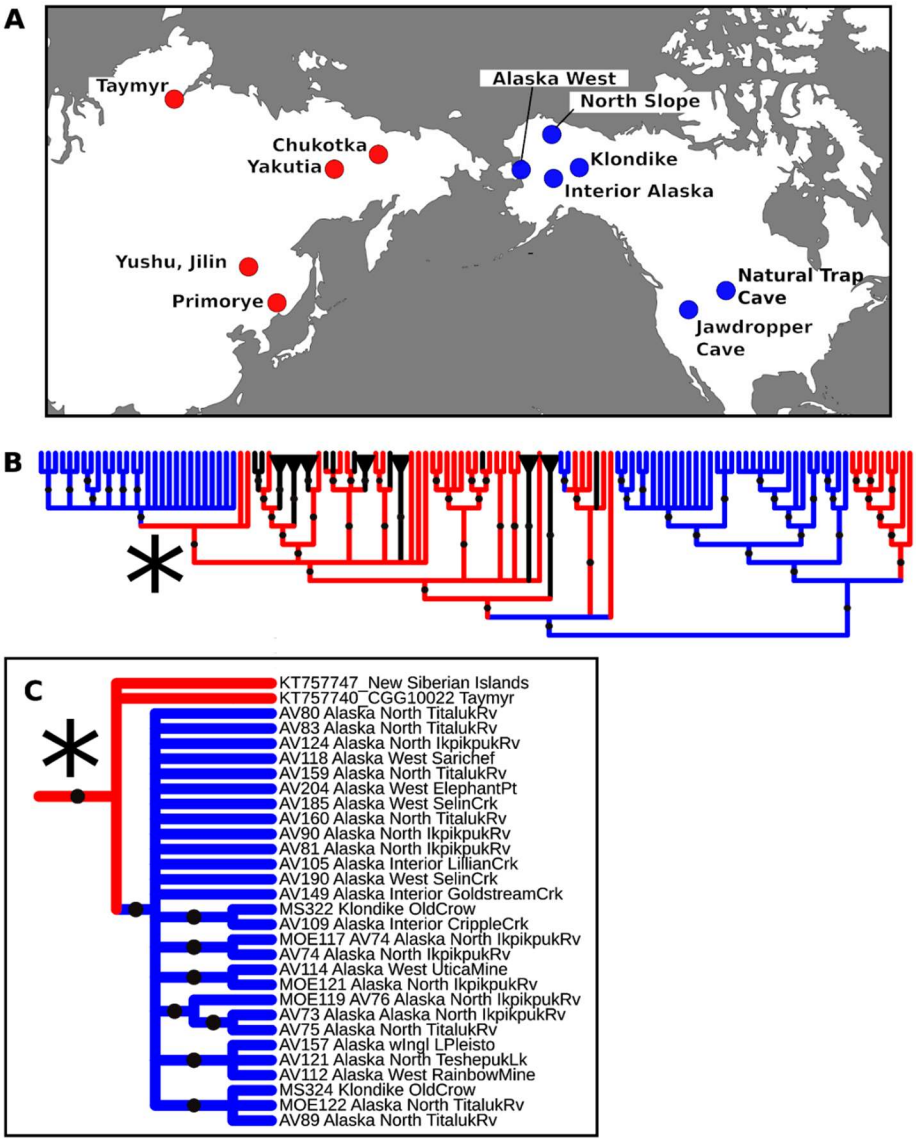


Figure 2.1. A. Sampling locations across Beringia. **B.** Molecular phylogeny of 79 new and 110 previously-published complete mitochondrial genomes reconstructed using RAxML. Black dots mark branches with bootstrap support >80, all clades with lower support are collapsed into polytomies. Branch lengths are not to scale to improve the visibility of the tree. Domestic horse clades (samples dated to the last 5 thousand years, including present-day) are collapsed. Present-day and historic domestic horses are in black, Pleistocene Eurasian horses are in red, and the extinct North American horses are in blue. Star marks the clade consisting of Alaskan horses with Eurasian haplogroup. Outgroup *E.asinus* (donkey) is not shown. **C.** Inset showing the representative clade of Alaskan and Klondike horses sharing Eurasian mitochondrial haplogroup.

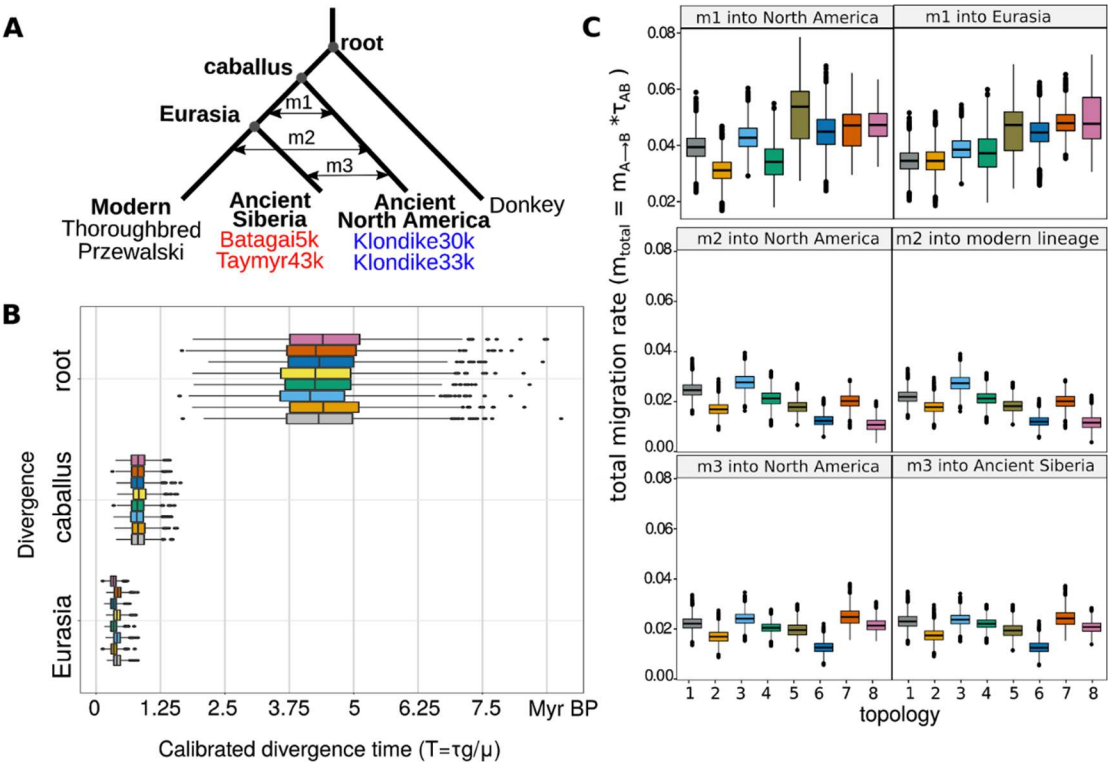


Figure 2.2. A schematic of the tested demographic scenario, estimates of the migration rates, and divergence times. In each G-PhoCS run only one genome was used as a representative of a population, producing eight possible configurations of the tree topology: (1) Przewalski - Batagai5k - Klondike30k; (2) Przewalski - Taymyr43k - Klondike30k; (3) Przewalski - Batagai5k - Klondike33k; (4) Przewalski - Taymyr43k - Klondike33k; (5) Thoroughbred - Batagai5k - Klondike30k; (6) Thoroughbred - Taymyr43k-Klondike30k; (7) Thoroughbred - Batagai5k - Klondike33k; (8) Thoroughbred - Taymyr43k- Klondike33k.
A. Four-lineage tree topology modeled in the demographic analysis. **B.** G-PhoCS estimates of the divergence times. **C.** Total migration rate determined by multiplying the number of migrants per generation with the time span of the migration band.

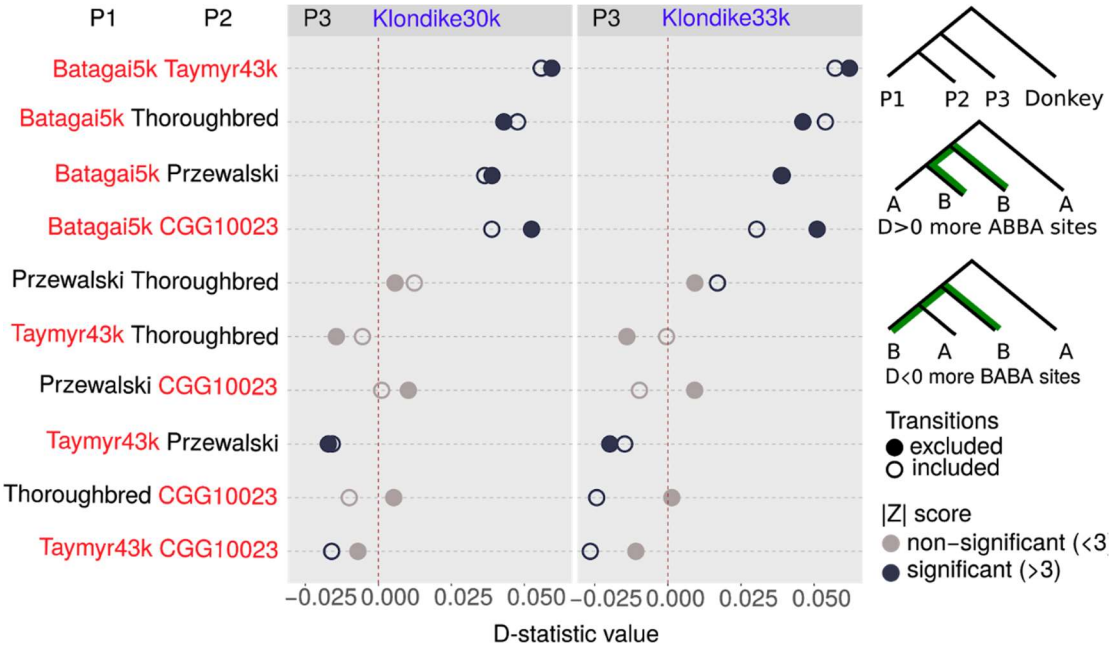


Figure 2.3. D-statistic test for admixture in autosomes of Pleistocene and present-day caballoid horses. The test in configuration (((P1,P2),P3), outgroup donkey) estimates whether the sample P3 shares alleles with P1 and P2 following the schematic on the figure. Removal of transitions accounts for sites of the genome damaged by post-mortem DNA decay. For all possible combinations of P1, P2, and P3 and X chromosome estimates see appendix figures S6, S7.

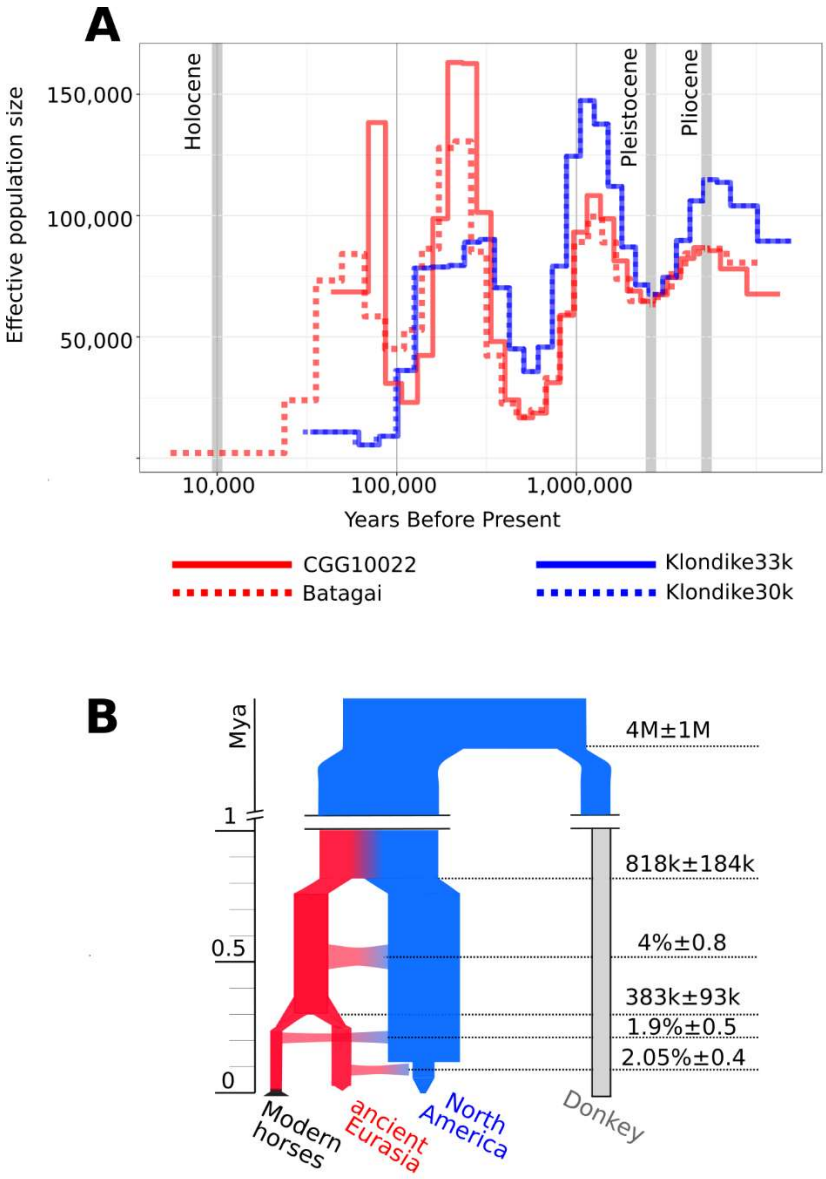


Figure 2.4. Demographic history of extinct and present-day caballoid horses. The ancient North American population is in blue (Klondike30k and Klondike33k) and ancient Eurasian horses are in red (Batagai5k and Taymyr43k). **A.** PSMC demographic profiles reconstructed with 100 bootstrap replicates (not shown) and scaled to calendar years using genome-wide substitution rate of 7.242e−9 per site per generation and a generation time of 8 years. **B.** Demographic history inferred by G-PhoCS. Widths of branches are proportional to effective population sizes. Horizontal dashed lines denote estimates for divergence times and total migration rates (mean values with standard deviation). Bands between the branches indicate the bidirectional gene flow. Parameters are scaled using the same mutation rate and generation time as PSMC.

APPENDICES

Table S1. Metadata for the new caballoid horse samples.

Abbreviations: YG – Government of Yukon Palaeontology Program; UAF – University of Alaska Fairbanks; ZIN RAS – Zoological Institute, Russian Academy of Sciences; IMNH - The Idaho Museum of Natural History; IVPP - The Institute of Vertebrate Paleontology and Paleoanthropology of China; IAM – Shidlovskiy's Ice Age Museum; KU – Kansas University; DV RAS - Federal Scientific Center of the East Asia Terrestrial Biodiversity Russian Academy of Sciences; RAM - Royal Alberta Museum ; UCIAMS – UC Irvine Keck Center; CAMS_LLNL - Center for Accelerator Mass Spectrometry (CAMS) at Lawrence Livermore National Laboratory.

Sample ID	Lab ID	Museum sample ID	Museum name	Radiocarbon lab	Date ID	Uncalibrated age (14C)	Age error (+\-)	Region	Locality	Mitochondrial genome coverage
Klondike30k	UP10.MS258	YG 303.325	YG	CAMS_LLNL	157454	26020	140	Klondike	Eldorado	13.3
Klondike33k	CGG_1_011850	YT03/40/ABC9779	YG	CAMS_LLNL	NA	28644	175	Klondike	Eldorado	320.2
AV103	SC16.AV049	UAM:ES:25586	UAF	UCIAMS	184441	19610	70	Alaska Interior	Fairbanks	51.4
AV105	SC16.AV084	UAM:ES:4614	UAF	UCIAMS	184446	31920	320	Alaska Interior	Lilian crk	4.7
AV106	SC16.AV046	UAM:ES:22097	UAF	UCIAMS	184442	>50300	Inf	Alaska Interior	Cripple crk	5.6
AV108	SC16.AV047	UAM:ES:22187	UAF	UCIAMS	184444	>48800	Inf	Alaska Interior	Fairbanks	10.9
AV109	SC16.AV048	UAM:ES:27107	UAF	UCIAMS	184445	16505	50	Alaska Interior	Fairbanks	21.5
AV112	SC16.AV058	UAM:ES:23663	UAF	UCIAMS	196059	16110	60	Alaska West	Rainbow Mine	763.8
AV114	SC16.AV087	UAM:ES:29667	UAF	UCIAMS	196061	17370	70	Alaska West	Utica Mine	229.1
AV118	SC16.AV075	UAM:ES:29799	UAF	UCIAMS	196063	20210	130	Alaska West	Sarichef I, Seward area	66.5
AV121	SC16.AV039	UAM:ES:n\a	UAF	UCIAMS	184437	11360	25	Alaska North	Teshepuk lake	26.1
AV122	SC16.AV027	UAM:ES:32679	UAF	UCIAMS	184438	34320	430	Alaska North	Kigalik riv	107.5
AV124	SC16.AV023	UAM:ES:34243	UAF	UCIAMS	184440	32880	360	Alaska North	Ikpikpuk riv	7.9
AV125	SC16.AV100	UAM:ES:7279	UAF	UCIAMS	196058	14165	50	Alaska Interior	Lost Chiken crk	183.9
AV149	SC16.AV013	UAM:ES:27275	UAF	UCIAMS	184449	13260	35	Alaska Interior	Goldstream crk	5.0
AV151	SC16.AV050	UAM:ES:22208	UAF	UCIAMS	184448	>50300	Inf	Alaska Interior	Fairbanks	9.9
AV154	SC16.AV117	ZIN 31607(81)	ZIN RAS	UCIAMS	196064	28050	280	Ural mtns	Perm oblast	4.7
AV157	SC16.AV064	UAM:ES:28590	UAF	UCIAMS	196060	14305	50	Alaska West	Inglutalik riv	95.4
AV159	SC16.AV037	UAM:ES:23918	UAF	GX	13939	21220	NA	Alaska North	Titaluk riv	46.8
AV160	SC16.AV104	UAM:ES:32985	UAF	Beta	362047	29510	170	Alaska North	Titaluk riv	24.0
AV181	SC16.AV083	UAM:ES:11068	UAF	UCIAMS	208140	39000	550	Alaska North	Ikpikpuk riv	4.4
AV185	SC16.AV007	UAM:ES:23547	UAF	UCIAMS	208129	17755	50	Alaska West	Selin crk	13.1
AV189	SC16.AV071	UAM:ES:27507	UAF	UCIAMS	208139	33650	290	Alaska Interior	Fairbanks	22.0
AV190	SC16.AV035	UAM:ES:23553	UAF	UCIAMS	208135	15595	40	Alaska West	Selin crk	5.9
AV191	SC16.AV070	UAM:ES:25745	UAF	UCIAMS	208138	>52200	Inf	Alaska Interior	Engineer crk	5.2
AV193	SC16.AV033	UAM:ES:22207	UAF	UCIAMS	208133	19390	60	Alaska Interior	Fairbanks	14.2
AV196	SC16.AV102	UAM:ES:7066	UAF	UCIAMS	208141	28020	150	Alaska Interior	Lost Chicken crk	11.0

AV204	SC16.AV008	UAM:ES:26694	UAF	UCIAMS	208130	30060	190	Alaska West	Elephant Point	4.2
AV208	UP.00.155	IMNH 71004/25895	IMNH	CAMS_LLNL	173978	295	25	Idaho	American Falls	114.5
AV236	SC17.AV091	V02173	IVPP	UCIAMS	218447	13090	35	China	Yushu, Jilin	27.7
AV237	SC17.AV089	V02172	IVPP	UCIAMS	218448	21470	90	China	Yushu, Jilin	17.4
AV238	SC17.AV087	V02196	IVPP	UCIAMS	218449	18610	70	China	Yushu, Jilin	22.1
AV241	SC17.AV092	V02178	IVPP	UCIAMS	218453	24290	130	China	Yushu, Jilin	17.1
AV34	SC15.PH171	F-1231	IAM	UCIAMS	196077	>49900	Inf	Yakutiya	Indigirka riv	4.2
AV36	SC15.PH158	F-2431	IAM	CAMS_LLNL	174388	40700	800	Yakutiya	Cherskiy	4.7
AV37	SC15.PH159	F-2442	IAM	UCIAMS	196482	31810	280	Chukotka	Maly Anyuy	10.6
AV38	SC15.PH160	F-2443	IAM	UCIAMS	196081	>49900	Inf	Chukotka	Maly Anyuy	6.2
AV39	SC15.PH161	F-2530	IAM	UCIAMS	196483	51900	3400	Yakutiya	Indigirka riv	11.0
AV40	SC15.PH164	F-2538	IAM	UCIAMS	196074	>49900	Inf	Yakutiya	Alazeya riv	9.5
AV42	SC15.PH157	F-255	IAM	CAMS_LLNL	174389	22210	90	Yakutiya	Bolshaya Kuropatochya riv	5.9
AV43	SC15.PH173	F-1674	IAM	UCIAMS	196076	>47800	inf	Yakutiya	Indigirka riv	5.9
AV5	UP08.BS.276	XA08-62	N/A	UBA	16489	29643	220	Taimyr	Kiyeng-Kyuel' lake	7.3
AV6	UP08.BS.282	N/A	N/A	CAMS_LLNL	174277	25290	200	Taimyr	N/A	5.8
AV7	UP08.BS.283	N/A	N/A	CAMS_LLNL	174278	44300	2100	Taimyr	N/A	10.9
AV73	SC16.AV089	UAM:ES:11953	UAF	UCIAMS	184431	12515	30	Alaska North	Ikpikpuk riv	285.5
AV74_MOE117	SC16.AV014	UAM:ES:29476	UAF	UCIAMS	184432	27280	190	Alaska North	Ikpikpuk riv	46.3
AV75_MOE118	SC16.AV096	UAM:ES:29465	UAF	UCIAMS	184433	13440	35	Alaska North	Titaluk riv	199.1
AV76_MOE119	SC16.AV056	UAM:ES:3299	UAF	CAMS_LLNL	120683	16885	45	Alaska North	Ikpikpuk	6.0
AV80	SC16.AV078	UAM:ES:3292	UAF	UCIAMS	184434	20250	80	Alaska North	Titaluk riv	22.6
AV81	SC16.AV090	UAM:ES:10644	UAF	CAMS_LLNL	91806	29700	200	Alaska North	Ikpikpuk	316.5
AV83	SC16.AV081	UAM:ES:29485	UAF	UCIAMS	184435	33010	360	Alaska North	Titaluk riv	53.8
AV90	SC16.AV093	UAM:ES:29495	UAF	Beta	331865	30560	160	Alaska North	Ikpikpuk riv	15.9
JK078	SC15.PH074	SW-11-10	N/A	UCIAMS	196065	35820	730	Chukotka	Kyttyk Peninsula	4.4
JK162	UP.00.153	IMNH 1136/11898	IMNH	CAMS_LLNL	175552	14225	40	Idaho	Jaw Dropper Cave	114.5
JK164	UP.00.155	IMNH 71004/25895	IMNH	CAMS_LLNL	173978	295	25	Idaho	American Falls	491.2
JK273	SC15.PH181	KU42626	KU	CAMS_LLNL	173976	20840	90	Wyoming	Natural Trap Cave	22.95
JK274	SC15.PH182	KU43413	KU	CAMS_LLNL	174029	20690	110	Wyoming	Natural Trap Cave	9.80
JK275	SC15.PH183	KU47538	KU	CAMS_LLNL	174093	21000	110	Wyoming	Natural Trap Cave	29.82
JK278	SC15.PH187	KU47519	KU	not dated	N/A	N/A	N/A	Wyoming	Natural Trap Cave	19.21
MOE079	SC17.AV073	Eq-Suh/2	DV RAS	UCIAMS	211762	51300	2500	Primorye	Sukhaya cave	60.5
MOE081	SC17.AV075	Eq-Suh/5	DV RAS	UCIAMS	211763	48400	1700	Primorye	Sukhaya cave	6.2
AV88_MOE121	SC16.AV063	UAM:ES:29474	UAF	not dated	N/A	N/A	N/A	Alaska North	Ikpikpuk	36.1
AV89_MOE122	SC16.AV106	TIT13-32	UAF	UCIAMS	184436	33280	370	Alaska North	Titaluk riv	41.0
MS273	N/A	QZ-10-237	N/A	CAMS_LLNL	157470	39740	720	Klondike	Quartz_Creek	60.8
MS309	UP10.MS240	YG 328.174	YG	CAMS_LLNL	157440	>53700	NA	Klondike	Hunker_Creek	81.6
MS313	UP10.MS244	YG 391.63	YG	CAMS_LLNL	157443	33350	340	Klondike	Quartz_Creek	35.7
MS317	UP10.MS248	YG 150.52	YG	CAMS_LLNL	157446	>53700	NA	Klondike	Thistle_Creek	32.5

MS322	UP10.MS399	YG 178.7	YG	CAMS_LLNL	157447	24380	120	Old Crow	OC_0.5km_from_Clijees	1187.3
MS324	UP10.MS401	YG 236.6	YG	CAMS_LLNL	157449	31040	250	Old Crow	OC_Pehria_Pt	11.1
MS338	UP10.MS259	YG 355.73	YG	CAMS_LLNL	157455	34690	390	Klondike	Green_Gulch	844.2
MS356	UP10.MS276	YG 328.94	YG	CAMS_LLNL	NA	>49400	NA	Klondike	Hunker_Creek	2287.0
MS365	UP10.MS285	YG 270.1	YG	CAMS_LLNL	157497	>52400	NA	Klondike	Lucky_Lady	51.3
MS371	UP10.MS290	YG 303.363	YG	CAMS_LLNL	157475	31400	260	Klondike	Eldorado	119.3
MS413	N/A	QZ-10-75	NA	CAMS_LLNL	161722	>44400	NA	Klondike	Quartz_Creek	87.0
MS440	N/A	QZ-10-393	NA	CAMS_LLNL	161546	36290	1460	Klondike	Quartz_Creek	31.4
MS451	UP10.MS316	YG 303.428	YG	CAMS_LLNL	161729	14340	90	Klondike	Eldorado	473.4
MS456	UP10.MS320	YG 126.55	YG	CAMS_LLNL	161730	24930	350	Klondike	Irish_Gilch	934.6
MS460	UP10.MS324	YG 126.33	YG	CAMS_LLNL	166299	29860	620	Klondike	Irish_Gilch	28.5
JW174_MS289	N/A	P98.21.1	RAM	OxA	14270	11200	90	Alberta	Grand Prairie	N/A

Table S2. Metadata for the previously published mitochondrial genomes used in the current study.

Sample ID	Reference	Equid group	Mitochondrial genome coverage (if re-assembled)
AM115_CGG_1_018579	Fages et al. 2019	Historic Domestic	192.9
Arz15_CGG_1_017084	Fages et al. 2019	Historic Domestic	9.2
Cap102_CGG_1_016984	Fages et al. 2019	Historic Domestic	18.2
CGG_1_018049	Fages et al. 2019	Historic Domestic	48.4
CGG_1_019246	Fages et al. 2019	Historic Domestic	109.3
CGG_1_019521	Fages et al. 2019	Historic Domestic	11.8
CGG_1_019994	Fages et al. 2019	Historic Domestic	34.0
Earb6_CGG_1_018495	Fages et al. 2019	Historic Domestic	133.4
Fen4_CGG_1_018396	Fages et al. 2019	Historic Domestic	105.4
Fr1_CGG_1_018151	Fages et al. 2019	Historic Domestic	377.9
GVA4_CGG_1_019262	Fages et al. 2019	Historic Domestic	72.2
HQ439441	GenBank	Modern Domestic (Akhal-Teke)	published_assembly
HQ439443	GenBank	Modern Domestic (Altai)	published_assembly
HQ439445	GenBank	Modern Domestic (Kladruber)	published_assembly
HQ439446	GenBank	Modern Domestic (Appaloosa)	published_assembly
HQ439447	GenBank	Modern Domestic (Arab)	published_assembly
HQ439450	GenBank	Modern Domestic (Ardennais)	published_assembly
HQ439451	GenBank	Modern Domestic (Bashkir Curly)	published_assembly
HQ439453	GenBank	Modern Domestic (Barb)	published_assembly
HQ439454	GenBank	Modern Domestic (Camargue)	published_assembly
HQ439455	GenBank	Modern Domestic (Clydesdale)	published_assembly
HQ439456	GenBank	Modern Domestic (German Sport Horse)	published_assembly
HQ439457	GenBank	Modern Domestic (Hanoverian)	published_assembly
HQ439458	GenBank	Modern Domestic (Holstein)	published_assembly
HQ439459	GenBank	Modern Domestic (Oldenburg)	published_assembly
HQ439460	GenBank	Modern Domestic (Westphalian)	published_assembly
HQ439461	GenBank	Modern Domestic (German Riding Pony)	published_assembly
HQ439466	GenBank	Modern Domestic (Icelandic)	published_assembly
HQ439467	GenBank	Modern Domestic (Yakut)	published_assembly
HQ439484	GenBank	Modern Feral (Equus przewalskii)	published_assembly
JN398380	GenBank	Modern Domestic (Middle East)	published_assembly
JN398457	GenBank	Modern Domestic (Middle East)	published_assembly
Kha2_t1_CGG_1_018909	Fages et al. 2019	Historic Domestic	56.3
KSH4_CGG_1_017098	Fages et al. 2019	Historic Domestic	76.3
KT168318 (YG13316)	Heintzman et al. 2017	Late Pleistocene (North America, Equus lambei)	published_assembly
KT168322 (YG32854)	Heintzman et al. 2017	Late Pleistocene (North America, Equus lambei)	published_assembly

KT368723.1 (Yukagir)	Librado et al. 2015	Late Pleistocene (Yakutia)	published_assembly
KT368724.1 (ODJ6)	Librado et al. 2015	Late Pleistocene (Yakutia)	published_assembly
KT368725.1 (Batagai)	Librado et al. 2015	Late Pleistocene (Yakutia)	published_assembly
KT368726.1 (CGG101397)	Librado et al. 2015	Late Pleistocene (Yakutia)	published_assembly
KT368727.1 (Horse1)	Librado et al. 2015	Modern Domestic (Yakutia)	published_assembly
KT368728.1 (Horse2)	Librado et al. 2015	Modern Domestic (Yakutia)	published_assembly
KT368729.1 (Horse3)	Librado et al. 2015	Modern Domestic (Yakutia)	published_assembly
KT368730.1 (Yak1)	Librado et al. 2015	Modern Domestic (Yakutia)	published_assembly
KT368731.1 (Yak2)	Librado et al. 2015	Modern Domestic (Yakutia)	published_assembly
KT368732.1 (Yak3)	Librado et al. 2015	Modern Domestic (Yakutia)	published_assembly
KT368733.1 (Yak4)	Librado et al. 2015	Modern Domestic (Yakutia)	published_assembly
KT368734.1 (Yak5)	Librado et al. 2015	Modern Domestic (Yakutia)	published_assembly
KT368735.1 (Yak6)	Librado et al. 2015	Modern Domestic (Yakutia)	published_assembly
KT368736.1 (Yak7)	Librado et al. 2015	Modern Domestic (Yakutia)	published_assembly
KT368737.1 (Yak8)	Librado et al. 2015	Modern Domestic (Yakutia)	published_assembly
KT368738.1 (Yak9)	Librado et al. 2015	Modern Domestic (Yakutia)	published_assembly
KT757740 (CGG10022)	Orlando et al. 2013	Late Pleistocene (Taymyr)	published_assembly
KT757741 (CGG10023)	Orlando et al. 2013	Late Pleistocene (Taymyr)	published_assembly
KT757742 (CGG10026)	Orlando et al. 2013	Late Pleistocene (Taymyr)	published_assembly
KT757743 (CGG10027)	Orlando et al. 2013	Late Pleistocene (Taymyr)	published_assembly
KT757744 (CGG10032)	Orlando et al. 2013	Late Pleistocene (Taymyr)	published_assembly
KT757746 (JW25/MS299)	Orlando et al. 2013	Holocene (Bol. Lyakhovsky Island, Russia)	published_assembly
KT757747 (JW26/MS300)	Orlando et al. 2013	Late Pleistocene (New Siberian Islands, Russia)	published_assembly
KT757748 (JW27/MS301)	Orlando et al. 2013	Late Pleistocene (New Siberian Islands, Russia)	published_assembly
KT757749 (JW28/MS298)	Orlando et al. 2013	Late Pleistocene (New Siberian Islands, Russia)	published_assembly
KT757750 (JW29/MS297)	Orlando et al. 2013	Late Pleistocene (New Siberian Islands, Russia)	published_assembly
KT757753 (JW307/MS295)	Orlando et al. 2013	Late Pleistocene (Ural Mt, Russia)	published_assembly
KT757754 (JW311/MS293)	Orlando et al. 2013	Late Pleistocene (Ural Mt, Russia)	published_assembly
KT757755 (JW313/MS305)	Orlando et al. 2013	Late Pleistocene (Ural Mt, Russia)	published_assembly
KT757756 (JW315/MS304)	Orlando et al. 2013	Late Pleistocene (Ural Mt, Russia)	published_assembly
KT757757 (JW341/MS296)	Orlando et al. 2013	Late Pleistocene (Ural Mt, Russia)	published_assembly
KT757758 (JW342/MS294)	Orlando et al. 2013	Late Pleistocene (Ural Mt, Russia)	published_assembly
KT757759 (JW374/MS303)	Orlando et al. 2013	Late Pleistocene (Ural Mt, Russia)	published_assembly
KT757761	Orlando et al. 2013	Modern Feral (Equus przewalskii)	published_assembly
KT757762	Orlando et al. 2013	Modern Domestic (Standardbred)	published_assembly
KT757763 (TC21c)	Orlando et al. 2013	Middle Pleistocene (North America, Thistle Creek)	published_assembly
KT757764 (Twilight)	Orlando et al. 2013	Modern Domestic (Thoroughbred)	published_assembly
KYRH8_CGG_1_018029	Fages et al. 2019	Historic Domestic	110.0
LOBOT_B_CGG_1_020182	Fages et al. 2019	Historic Domestic	76.4
Marvele18_CGG_1_019405	Fages et al. 2019	Historic Domestic	81.5

Mon87_CGG_1_018123	Fages et al. 2019	Historic Domestic	164.6
Mzr1_CGG_1_018150	Fages et al. 2019	Historic Domestic	89.2
NB_Ra_8_49_CGG_1_020509	Fages et al. 2019	Historic Domestic	11.2
NC001788	GenBank	Equus asinus (outgroup)	published_assembly
NewBotai10	Fages et al. 2019	Historic Domestic	281.7
NewBotai15_CGG_1_017007	Fages et al. 2019	Historic Domestic	41.7
NewBotai4_CGG_1_016996	Fages et al. 2019	Historic Domestic	59.6
NewBotai44_CGG_1_017036	Fages et al. 2019	Historic Domestic	66.2
NewBotai45_CGG_1_017037	Fages et al. 2019	Historic Domestic	27.2
NewBotai9_CGG_1_017001	Fages et al. 2019	Historic Domestic	34.2
NUSTAR5	Fages et al. 2019	Historic Domestic	116.4
OKG2_CGG_1_018398	Fages et al. 2019	Historic Domestic	115.5
Ote2_CGG_1_018473	Fages et al. 2019	Historic Domestic	111.3
Rid2_CGG_1_018469	Fages et al. 2019	Historic Domestic	174.9
Rus16_CGG_1_019166	Fages et al. 2019	Historic Domestic	292.8
Rus37_CGG_1_019185	Fages et al. 2019	Historic Domestic	187.6
Rus38_CGG_1_019186	Fages et al. 2019	Historic Domestic	154.3
Rus41_CGG_1_019189	Fages et al. 2019	Historic Domestic	145.4
Rus48_CGG_1_019195	Fages et al. 2019	Historic Domestic	14.2
Rus9_CGG_1_019160	Fages et al. 2019	Historic Domestic	266.7
Saa1_CGG_1_018474	Fages et al. 2019	Historic Domestic	89.9
SAG_S27_CGG_1_019559	Fages et al. 2019	Historic Domestic	46.8
Spain38_CGG_1_020484	Fages et al. 2019	Historic Domestic	337.7
Spain39_CGG_1_020485	Fages et al. 2019	Historic Domestic	151.1
Svi6_CGG_1_018375	Fages et al. 2019	Historic Domestic	28.9
Tur145_CGG_1_018711	Fages et al. 2019	Historic Domestic	300.9
UCIE2012_85	Fages et al. 2019	Historic Domestic	94.0
UE4618_CGG_1_020962	Fages et al. 2019	Historic Domestic	229.1
UK17_CGG_1_019444	Fages et al. 2019	Historic Domestic	74.6
Upps02_CGG_1_018490	Fages et al. 2019	Historic Domestic	46.3
Vert293_CGG_1_018522	Fages et al. 2019	Historic Domestic	19.9
Vert311_CGG_1_018540	Fages et al. 2019	Historic Domestic	22.2
VHR017_CGG_1_020952	Fages et al. 2019	Historic Domestic	40.4
YER28_CGG_1_020254	Fages et al. 2019	Historic Domestic	50.8

Figure S1. Depth of coverage for nuclear genomes of North American horses Klondike30k (UP10.MS258, YG303.325) and Klondike33k (CGG_1_011850).

Filtered and adapter-trimmed reads were aligned to EquCab2.0 horse genome with BWA aln with seed disabled.

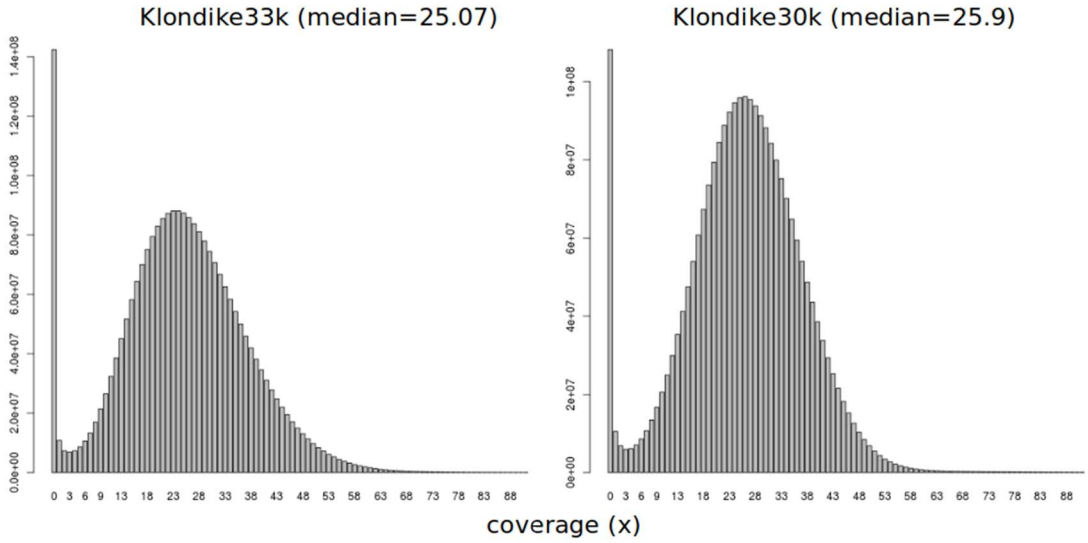


Figure S2. Neighbour-joining phylogenetic analysis of seven high coverage genomes built on 4215 putatively-neutral nuclear 1kb loci.

With the outgroup donkey genome excluded, there were 13646 variable sites in the source nucleotide alignment, 5240 of which were parsimony informative. Numbers at nodes show bootstrap support after 500 replicates.

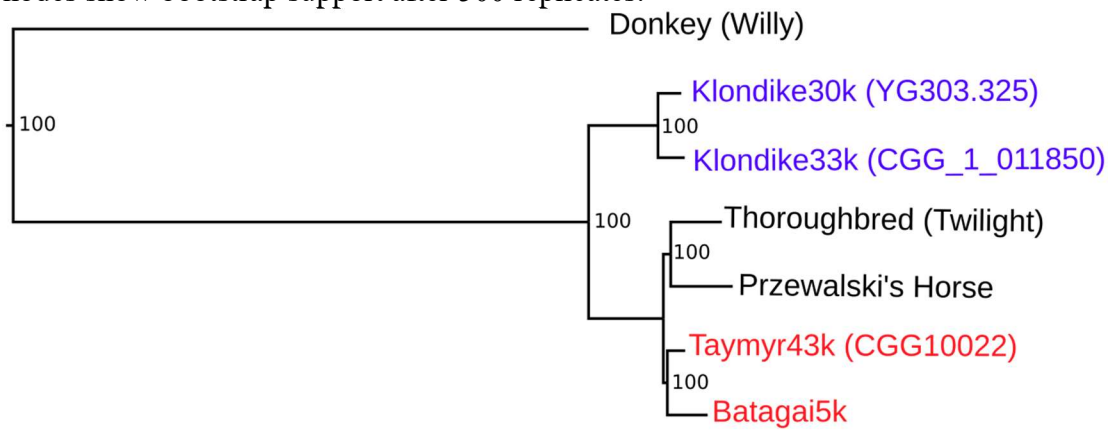


Figure S3. Maximum likelihood phylogenetic analysis based on complete mitochondrial genomes.

Black dots mark branches with bootstrap support >80, all clades with bootstrap lower than 80 are collapsed into polytomies. Branch lengths were omitted to improve tree visibility. Historic domestic horses (dated to the be 5 thousand years and younger) and modern domestic horses are in black, Ancient Eurasian horses are in red, extinct North American horses are in blue. Outgroup *E.asinus* is not shown.

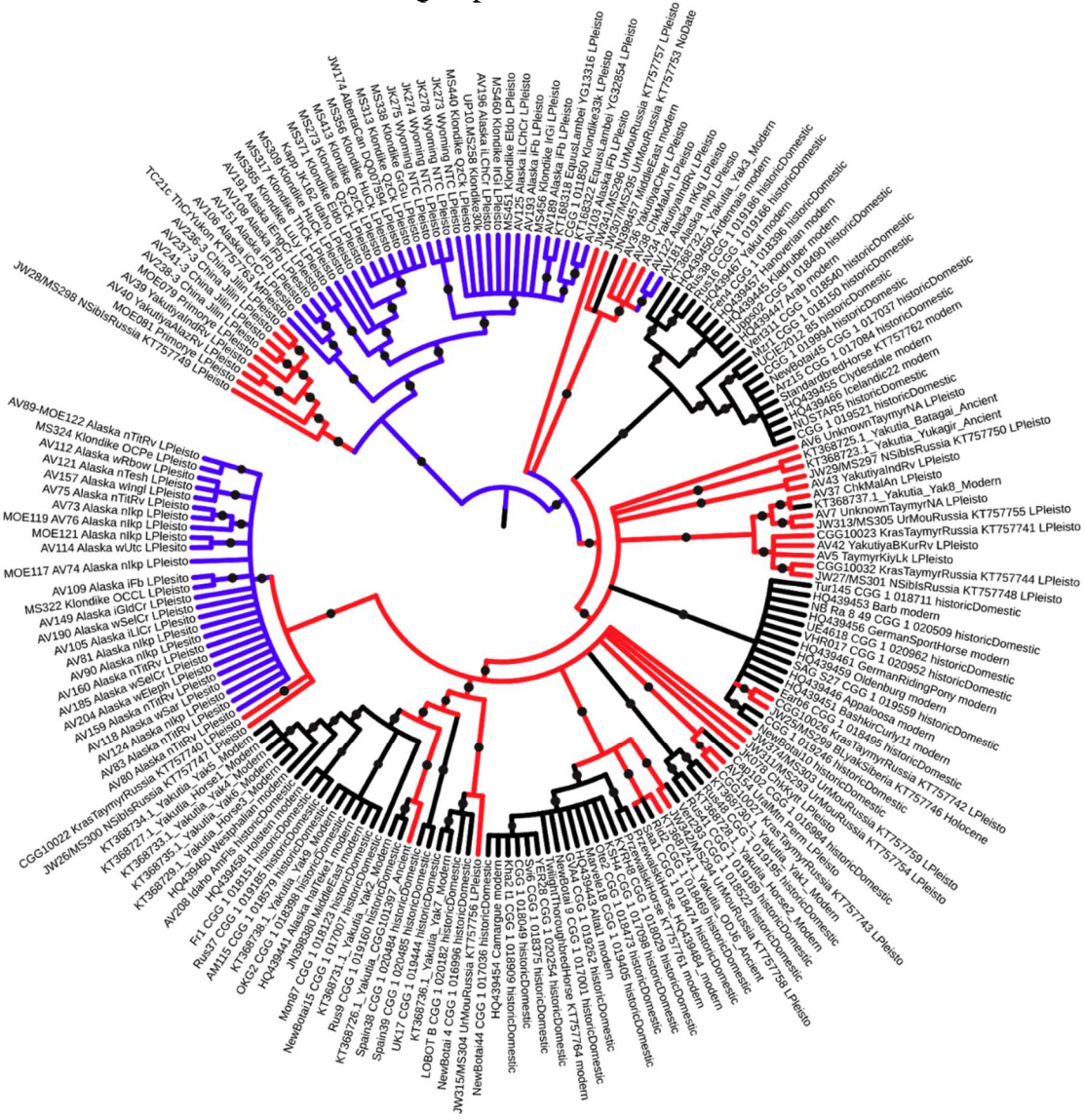


Figure S4. Genome-wide mutation rate (A) and generation time (B) used for re-scaling of G-PhoCS coalescent estimates into calendar years.
To account for uncertainty in estimates of the mutation rates and variable generation time we created gamma-distribution for each of these parameters. To calibrate G-PhoCS runs we randomly sampled values from these distributions.

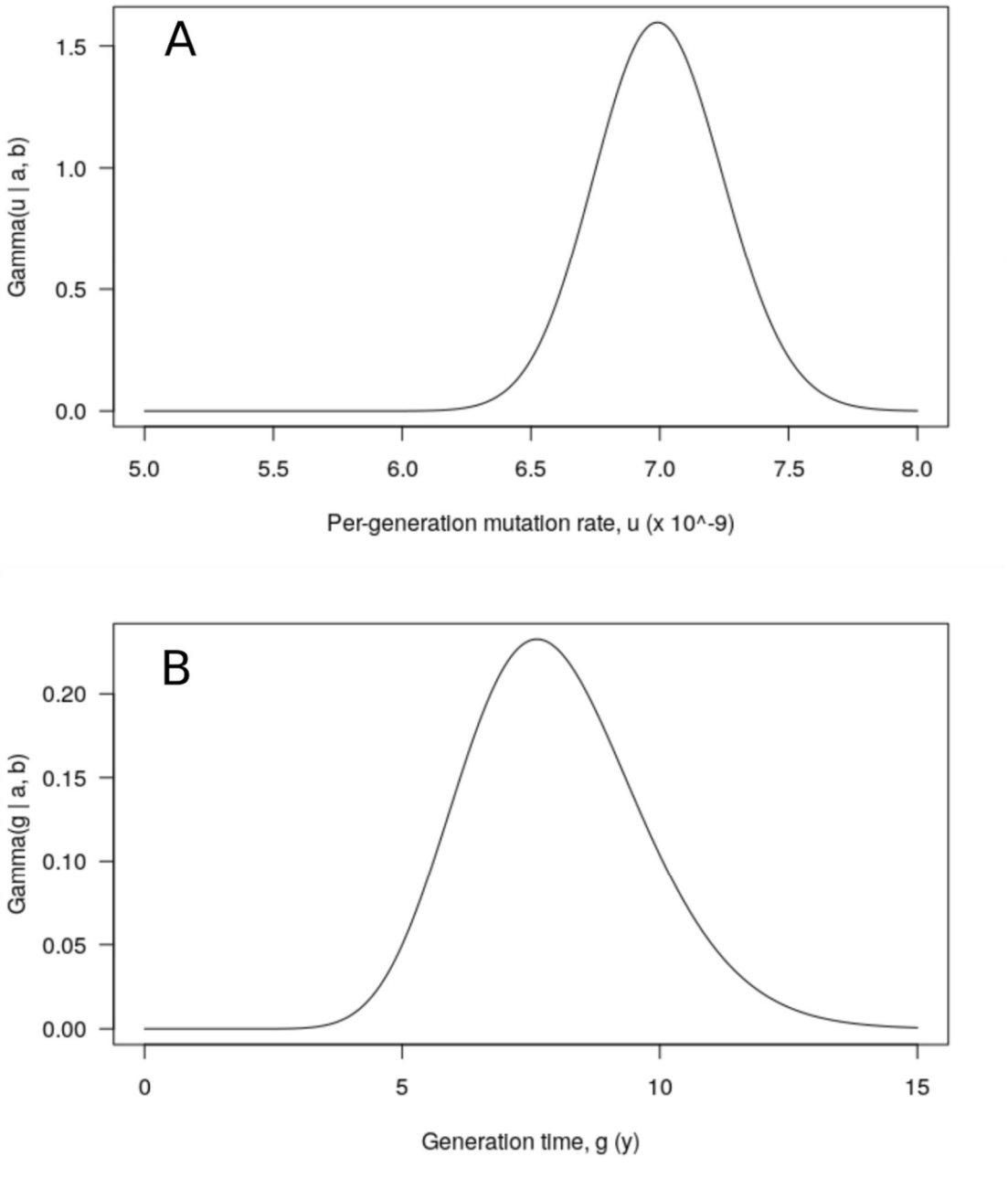


Figure S5. Caballoid horse demographic inference for the past 4M years. Calibrated effective population sizes estimated with G-PhoCS on a four-genome tree topology model.

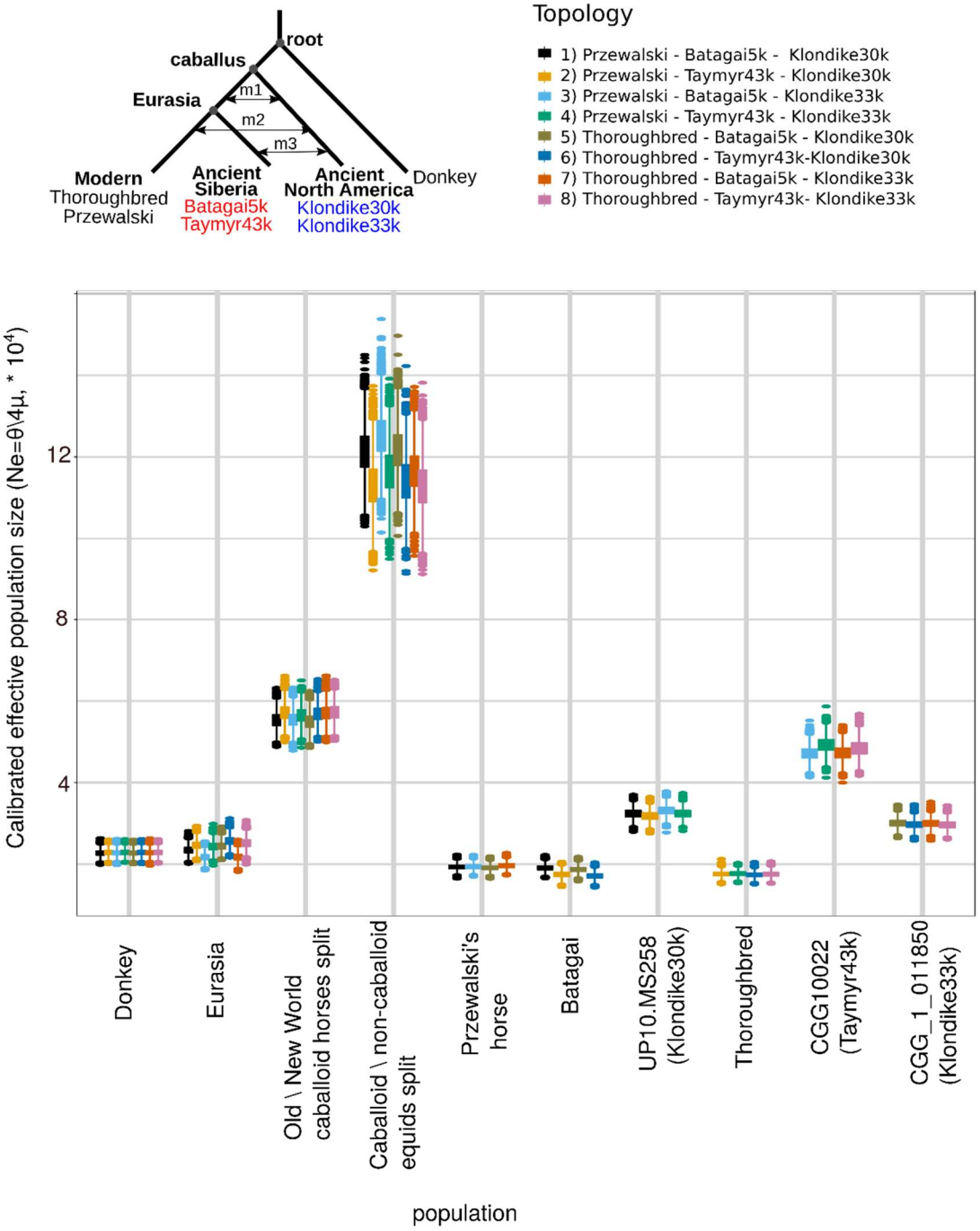


Figure S6. Admixture estimated on autosomes of high coverage caballoid horse genomes.

The D-statistic (ABBA\BABA test) for all possible combinations of two Pleistocene Klondike horses, two Pleistocene Eurasian horses (Taymyr43k and CGG10023), one Holocene Eurasian (Batagai5k) horse, and two modern horses (Thoroughbred and Przewalski's horse). Negative D-statistics suggest gene flow between P1 and P3, while positive - between P2 and P3. The Donkey genome was used as an outgroup.

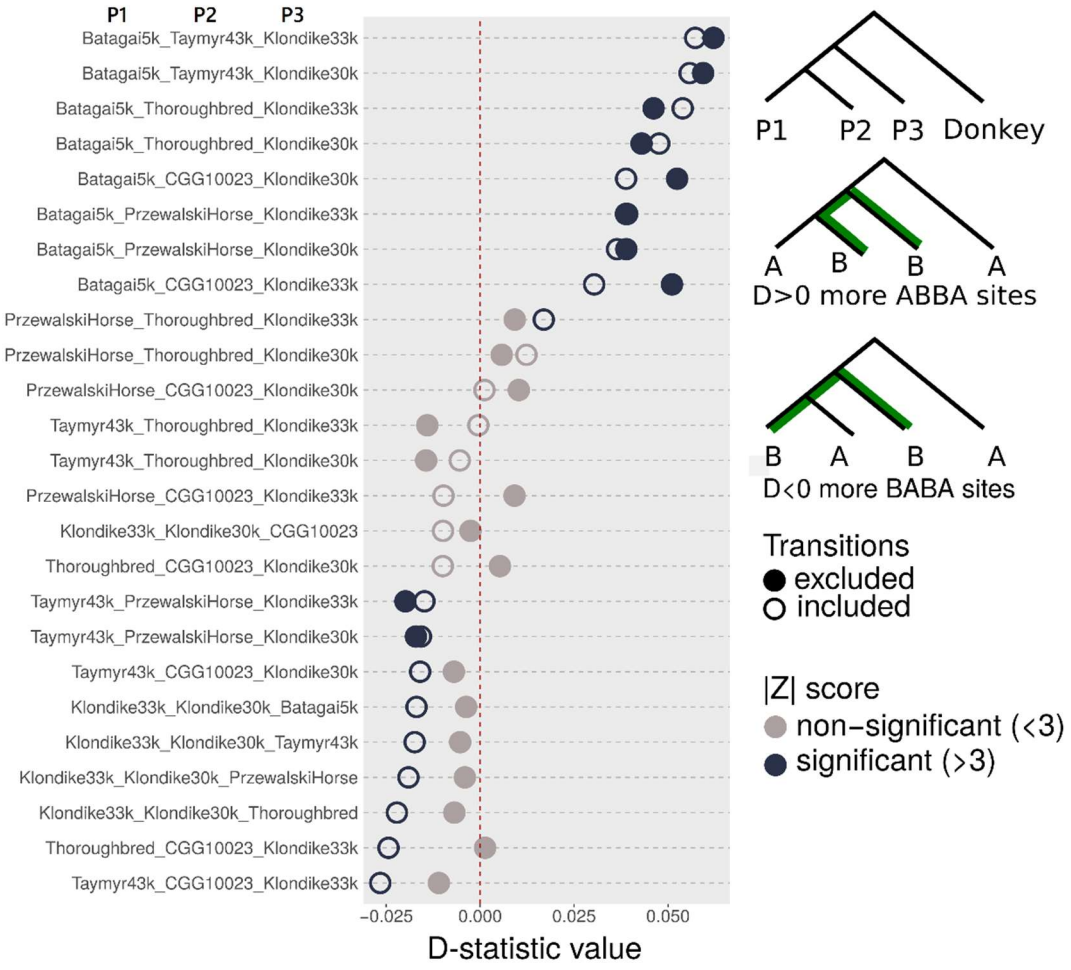


Figure S7. Admixture estimated on X chromosome of high coverage caballoid horse genomes.

The D-statistic (ABBA\BABA test) for all possible combinations of two Pleistocene Klondike horses, two Pleistocene Eurasian horses (Taymyr43k and CGG10023), one Holocene Eurasian (Batagai5k) horse, and two modern horses (Thoroughbred and Przewalski's horse). Negative D-statistics suggest gene flow between P1 and P3, while positive - between P2 and P3. The Donkey genome was used as an outgroup.

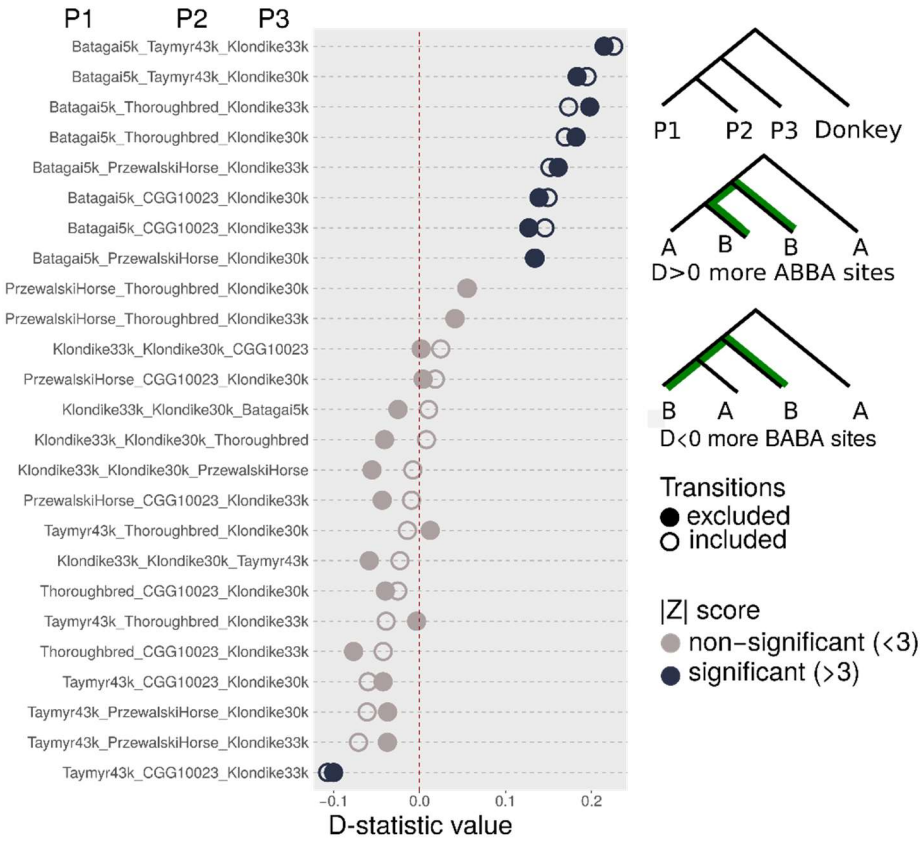
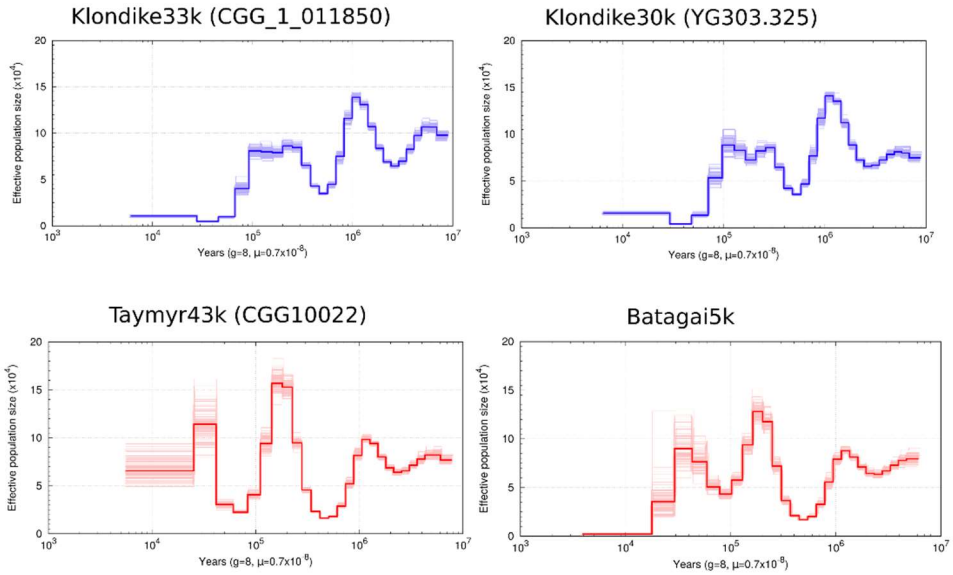


Figure S8. PSMC demographic inference for ancient caballoid horses.

PSMC analysis was performed on quality-filtered reads mapped to EquCab2.0 horse genome. Faint lines show reconstruction for 100 bootstrap replicates.



Chapter 3

A high coverage paleogenome of an Early-Middle Pleistocene Equid from northwest Canada

ABSTRACT

North American equids have been extensively studied since the advent of molecular paleontology, although their high coverage genomes have not yet been recovered and the evolutionary history of major equid groups in the Early Pleistocene remains unclear. Here we fill in this gap by presenting a complete nuclear paleogenome of a new Early-Middle Pleistocene equid species from Paradise Hill (Klondike, Yukon Territory). The specimen find is associated with the Gold Run volcanic ash layer dating it to be older 0.69 - 0.76 million years. We utilized single-stranded DNA library protocol to recover DNA from the petrous bone of the equid's cranium. Due to exceptional DNA preservation we were able to reconstruct the nuclear genome of the Paradise Hill equid up to 12-fold coverage and the mitochondrial genome up to 75-fold coverage, making this dataset the oldest high-coverage paleogenome on record. Tip-dated mitochondrial phylogeny and genotype likelihood analysis of present-day and extinct equid species places the Paradise Hill specimen into a stenonid clade encompassing present-day zebras, Asiatic wild asses, and kiangs. Spatial and temporal overlap of the new to genetics species with *Equus cf. scotti* suggests co-existence of stenonid and caballoid horse groups around 1 million years ago in northwest Canada. Our results demonstrate the power of paleogenomic studies for biodiversity reconstruction and provide insights into the evolution of North American equids.

INTRODUCTION

Equids (family Equidae) originated in North America as archaic ungulates that evolved and diversified for 55 millions of years (Myr) (MacFadden, 2005). The most recent common ancestor of the genus *Equus* is traced back to the Pliocene 4-4.5 Myr ago when two major present-day phylogenetic groups arose: stenonid, comprised of zebras, wild asses, and donkeys (*Equus asinus*), and caballoid, or "true horses", consisting of domestic and Przewalski's horses (*E. caballus,* or *E. ferus*) (Orlando, 2015). The present-day Asiatic wild asses (*E. hemionus*) and kiangs (*E. kiang*) occupy the southern parts of Eurasia. Plains zebra (*E. quagga)*, Mountain zebra (*E. zebra*), endangered Grevyi's zebra (*E. grevyi*) and critically endangered Somali wild ass (*E. asinus somalicus*) inhabit Africa. Caballoid horses were widespread in the Pleistocene, but their populations almost disappeared during the Pleistocene-Holocene transition before their domestication around 5.5 thousand years (Kyr) ago (Fages et al., 2019; Koch & Barnosky, 2006). Present-day Przewalski's horses survive in Mongolia and feral domestic horses inhabit the western United States (Gaunitz et al., 2018). Although the present-day wild equids are limited to Eastern Hemisphere, their ancestors lived in the New World and crossed the Bering Land Bridge at least twice before radiating in the Old World (Eisenmann, 1992). Stenonid horses were the first to disperse around 2 Myr ago, followed by caballoid horses around 1 Myr ago (Orlando, 2019 and Chapter 2 of this dissertation).

Today's diversity and distribution of equids is only a faint snapshot of a much greater collection of species that were abundant across the globe in the past. Thanks to

an ample paleontological record of their adaptive radiation, horse fossil remains have fueled evolutionary studies for a century (MacFadden, 1994, 2005; MacFadden et al., 2012). Despite the incredible richness of forms and good fossil preservation, the taxonomy of the group has been confusing due to morphological variation and debatable species assignments based on fragmentary fossil remains (Azzaroli, 1991; Barrón-Ortiz et al., 2019; Eisenmann, 2006; Eisenmann & Baylac, 2000; Orlando et al., 2009). The taxonomic oversplitting and homoplasies in taxonomically-important characters made reconstructions of the evolutionary history of equids challenging and tracing their recent diversification nearly impossible (Barrón-Ortiz et al., 2019; Eisenmann, 2004; Eisenmann & Baylac, 2000; Forsten, 1991; Heintzman et al., 2017). Due to these limiting factors, genomic data and ancient DNA studies have been essential in the understanding of the temporal and phylogenetic relationships between equids (Orlando, 2019).

Here we sequence and analyze an Early-Middle Pleistocene equid from Paradise Hill mining site, Klondike, in Canada's Yukon Territory. The specimen is associated with the Gold Run volcanic ash layer, dated to be older than 650 Kyr before present (BP) (Westgate et al., 2011). The analysis of the 75-fold coverage mitochondrial genome and 12-fold coverage nuclear genome indicates that the specimen is phylogenetically close to the present-day wild asses and zebras. The new to genetics species likely represents an extinct branch of archaic stenonid ungulates and, due to its temporal and spatial overlap with a caballoid *E. cf. scotti,* suggests the

coexistence of “true horses” and non-caballoid equids in the Early-Middle Pleistocene Yukon.

MATERIALS AND METHODS

The partial equid cranium (museum ID YG412.17, Figure 3.1A,B) was found in association with other incomplete bone fragments at the Paradise Hill placer gold mine in the Klondike region. Stratigraphic assessment associates the finding with the Hollis and Gold Run ash layers (tephra), but closer to the latter, and much below it (Appendix figure S1, Britta Jensen, Duane Froese, and Grant Zazula, personal communication, 3 March 2020).

We extracted the petrous part of the temporal bone from the skull as this bone shows an extremely high DNA preservation profile (Alberti et al., 2018). We then powdered the bone and split the powder into 7 extraction reactions (Appendix table S1). Following the protocol as described in Fulton & Shapiro (2019) we performed all extraction steps and DNA library preparation prior to indexing PCR in a sterile facility, at the University of California, Santa Cruz Paleogenomics Laboratory. We pretreated bone powder with 0.5% solution of sodium hypochlorite followed by three washes with water to decrease DNA contamination (Boessenkool et al., 2017). We extracted DNA following the protocol optimized to increase recovery of short DNA molecules (Dabney & Meyer, 2019).

DNA extracts were prepared into single-stranded Illumina DNA sequencing libraries following the Troll et al. (2019) protocol modified for recovery of degraded ancient DNA (Appendix table S2, Vershinina, Kapp et al. in prep). We calculated the

required concentration of P5 and P7 adapters and SSB proteins depending on the concentration of DNA library input, and quantified the DNA library using qPCR after adapter ligation reaction. Using results of the qPCR with Maxima SYBR Green Master Mix (Thermo Fisher Scientific) and IS7 and IS8 primers, we amplified libraries for 16-18 cycles using Amplitaq Gold 360 Master Mix (Thermo Fisher Scientific) and a unique combination of forward and reverse indexes. We cleaned the resulting libraries with 1.2X Sera-Mag SPRI SpeedBeads (Thermo Fisher Scientific) in 18% PEG-8000. We pooled the resulting libraries in equimolar concentrations and sequenced them on 1.25 lanes of Novaseq 6000 with S4 150bp paired-end read chemistry (Appendix tables S1,S2).

To process the sequencing data, we ran SeqPrep2 (http://github.com/jstjohn/SeqPrep) to cut adapters and merge reads that overlap >15bp. We discarded all reads shorter than 27bp and ran Prinseq to remove low complexity sequences with threshold 7 and "dust" method (Schmieder & Edwards, 2011). We mapped merged and unmerged reads with BWA *aln* and seed disabled, following previously published suggestions on ancient DNA sequence processing (Li & Durbin, 2009; Schubert et al., 2012). To assess DNA damage profile and fragment length distribution we ran MapDamage2 on BWA alignments (Jónsson et al., 2013). We then removed duplicated reads with samtools rmdup, realigned reads around indels with GATK v. 3.7 and assessed quality of BWA alignments with Qualimap2 and bedtools v. 2.25 (Appendix figures S2, S3) (Li et al., 2009; McKenna et al., 2010; Okonechnikov et al., 2016; Quinlan & Hall, 2010). We used MIA

(https://github.com/mpieva/mapping-iterative-assembler) on quality-filtered data as described above to assemble the mitochondrial genome of the Paradise Hill horse following Chapter 1 of this dissertation (Vershinina et al., 2019). To account for unclear taxonomic position of the specimen, we ran MIA several times seeding it with various equid species references (Appendix table S3). For the final consensus sequence, we used a mitochondrial genome assembled with kiang mtDNA as a seed (NCBI ID NC_020433), calling bases with at least 3-fold read coverage and 67% consensus among reads.

To compare Paradise Hill horse nuclear genome with available high-coverage genomes of other equids, we used several previously-published genomic datasets (Table 3.1). First, we processed publicly available SRA data for seven present-day non-caballoid equid species published in Jónsson et al. (2014) and Renaud et al. (2018): European Nucleotide Archive study IDs PRJEB7446 and PRJEB24845. Second, we added to the analysis previously published ancient and present-day genomes of caballoid horses: Batagai, Middle Pleistocene Thistle Creek horse (TC21), CGG10022, CGG10023, and Przewalski's horse from Der Sarkissian, Ermini, et al., 2015; Librado et al., 2015; Orlando et al., 2013; Schubert et al., 2014; Thoroughbred (Twilight) from Wade et al., 2009. Finally, we added the dataset of high coverage genomes of extinct New World caballoid horses, analyzed in Chapter 2 of this dissertation: UP10.MS258 (YG303.325), CGG_1_011850 (YT03-40). In addition, we utilized a 0.08-fold coverage genome of another Klondike horse (CGG_1_011849) that was available from the collaborators (Ludovic Orlando group, University of Toulouse). For all present-day

genomes we used the same pipeline as described above, except for the following changes to account for the non-ancient nature of the specimens: AdapterRemoval v.2.1.7 (Schubert et al., 2016) instead of SeqPrep2, no read merging before BWA read alignment, and read mapping with BWA *mem* algorithm (Table 3.1).

To place Paradise Hill horse into the evolutionary context of other equid specimens, we utilized the strategy outlined in Chapter 2 to ascertain a set of putatively neutral loci: stretches of the genome that lay outside of gene boundaries, CpG islands, tandem and simple repeat regions, and assembly gaps. Briefly, we pseudohaploidized high coverage genomes with ANGSD v.0.923 -doFasta (Korneliussen et al., 2014). Following Chapter 2 (methods section) we collected 23,694 noncoding putatively neutral 1kb loci and concatenated them to reconstruct a maximum likelihood phylogeny. We concatenated these loci into the global alignment of 15 genomes and ran RAxML with the GTR nucleotide substitution model and 1000 bootstrap replicates (Stamatakis, 2014). We visualized the final tree with FigTree v1.4.3 (Table 3.1, Figure 3.1C) (Rambaut, 2014).

To overcome the obstacle of low coverage depth in some nuclear genomes, we used a probabilistic genotype likelihood framework and estimated genotype probabilities taking uncertainty into account (Korneliussen et al., 2014; Meisner & Albrechtsen, 2018). We used ANGSD v.0.923 to call genotype likelihoods of single nucleotide polymorphisms (SNPs) across equid autosomes (Korneliussen et al., 2014). We limited the estimation only to bases that have mapping and sequencing quality higher than 30 Phred, requiring each site to be covered by at least one read in each of

the genome alignments. We omitted A-G and C-T transitions from the analysis due to post-mortem DNA deamination. We then used PCAngsd to infer the allele covariance matrix that we analysed in R with prcomp to compute PCA and factoextra to visualize the results (Table 3.1, Figure 3.2) (Kassambara & Mundt, 2017; Meisner & Albrechtsen, 2018; R Core Team, 2014).

To infer the mitochondrial genome phylogeny, we created a data set of 35 previously published mitochondrial genomes assembled from a world-wide variety of present-day and extinct equid species (Appendix table S4, Figure 3.3) (Der Sarkissian, Vilstrup, et al., 2015; Druzhkova et al., 2017; Heintzman et al., 2017; Plasteeva et al., 2015; Weinstock et al., 2005; Yuan et al., 2019). This analysis made possible to compare the genetic affinity of the Paradise Hill horse to the extinct equid species for which nuclear genomes are not currently available: *Hippidion sp.*, *Haringtonhippus sp.*, and *E. ovodovi*. We used tip-calibrated BEAST with the pipeline outlined below (Drummond & Rambaut, 2007). We aligned the assembled mitochondrial genome of Paradise Hill to this data set using MAFFT v.7 (Nakamura et al., 2018), and manually checked the alignment for discrepancies. Using the annotated *E. caballus* mitochondrial genome (NCBI ID JN398421), we subdivided the mitochondrial alignment into six partitions (first, second, and third codon positions for protein coding genes, concatenated tRNAs, concatenated rRNAs, and the control region). Following Heintzman et al. (2017), we excluded third codon positions of the codon and the control region from downstream analyses to eliminate homoplasious mutations in rapidly evolving positions of the mitochondrial genome. We estimated the appropriate

evolutionary models for each partition using jModeltest2. However, test runs of BEAST with these models failed to converge. To avoid this problem we utilized the less parameter-rich HKY nucleotide substitution model for all partitions which improved the convergence (Darriba et al., 2012). We ran two instances of BEAST 1.8.4 for 100 million iterations per run, sampling model parameters and trees every 10,000 iterations (Drummond & Rambaut, 2007). We assumed the uncorrelated relaxed clock model and calibrated the molecular clock using the radiocarbon dates of ancient samples as priors, with ages of the present-day samples set to 0 (Appendix table S4). For the Paradies Hill specimen date we used normal prior (mean: 700 Kyr, SD: 56.4 Kyr). To calibrate the tree we also specified the divergence of the crown caballine group as 4–4.5 Myr BP (normal prior, mean 4.25 Myr, *SD*: 1.5 Myr, following Orlando et al., 2013). To convert radiocarbon dates of the Late Pleistocene samples into years before present we used IntCal13 curve (Reimer et al., 2013) and OxCal v4.2 (Ramsey, 2009), and assigned the median calibrated age of each sample as prior information. We used the birth-death process with serial samples as a tree prior (Drummond et al., 2006; Stadler, 2010). We estimated the substitution and clock parameters separately for each partition, and estimated a single tree linking all partitions. We discarded the first 25% of MCMC iterations from each run as burnin, and analyzed parameters for convergence in Tracer (Rambaut et al., 2018). All parameters reached an effective sample size > 1000. We then combined trees from the two BEAST runs in logCombiner and calculated the maximum clade credibility (MCC) tree in Tree Annotator (Rambaut &

Drummond, 2010). We visualized reconstructed trees in Figtree v1.4.2 and Densitree (Figure 3.3A,B) (Bouckaert & Heled, 2014; Rambaut, 2014).

RESULTS

We extracted DNA from ~1.2g of the petrous bone, and generated ~4 billion sequencing reads for a total of 21 DNA libraries (Figure 3.1A,B, Appendix tables S1,S2,S3, Appendix figures S1, S2, S3). Removing adapter dimers and merging reads produced ~1 billion filtered sequences that we mapped to three reference genomes: EquCab2, EquCab3, and Donkey (NCBI IDs GCF_000002305.2 (Wade et al., 2009), GCF_002863925.1 (Kalbfleisch et al., 2018), GCA_003033725.1 (Renaud et al., 2018) accordingly, Appendix figure S2). Read mapping showed a high proportion of endogenous equid DNA in comparison to contaminant DNA (20-40%, Appendix table S2), which is exceptionally high for a ~750 Kyr old specimen, but in accordance with potential of the temporal bone for high molecular preservation (Alberti et al., 2018). High ancient DNA concentration allowed sequencing of the nuclear genome of the specimen up to 12-fold coverage and the mitochondrial genome up to 75-fold coverage (Appendix figure S2, Appendix table S4). Interestingly, aligning sequencing reads to the old and improved versions of the domestic horse genome showed a lower average coverage (10.5-fold) than aligning to a donkey genome (11.8-fold, Appendix figure S2). Furthermore, donkey reference mapping produced an alignment where fewer areas of the genome had 0 coverage in comparison to the domestic horse (Appendix figure S2, table S3). All DNA libraries demonstrated characteristics of ancient DNA

degradation: short fragment length (average 48±15bp), abundance of C>T and G>A substitutions at the ends of the molecules (Appendix figure S3).

The RAxML analysis (Table 3.1, Figure 3.1C) divided the panel of pseudo-haplodyzed nuclear genomes into the two major groups: caballoid (with ancient North American Klondike horses) and stenonid (Eurasian and African equids). Surprisingly, the Paradise Hill horse fell into the highly-supported stenonid clade close to present-day wild asses and zebras.

For the dataset of 22,137,398 SNPs called genome-wide from ancient and present-day equids combining low and high coverage genomes, we estimated genotype likelihoods and ran PCA analysis on genotype covariance matrix (Figure 3.2, Appendix figure S4). PC1 accounting for 92.2% of the variance, distinguished between the caballoid and non-caballoid groups, placing the Middle Pleistocene Thistle Creek together with other Klondike horses, except for the Paradise Hill specimen (Figure 3.2, Appendix figure S4). The Paradise Hill horse clustered closer to the Eurasian and African wild asses, rather than to extinct Pleistocene North American caballoid horses, in accordance with the phylogenetic inference (figures 1C and 2). The clustering likely shows the genome-wide differences between the Thistle Creek and the Paradise Hill species, rather than the signal of missing data from the ancient DNA, since we excluded transitions from the analysis to account for sites likely damaged during DNA decay.

We reconstructed the whole mitochondrial genome of the Paradise Hill horse using several references (Appendix table S3) and obtained the highest coverage using the kiang (*E. kiang*) and Asiatic wild ass (*E. hemionus*) mitochondrial genomes

(average coverage: 75.3-fold and 75.1-fold accordingly). The lowest coverage showed assemblies seeded with an extinct South American *Hippidion sp*. (38.4-fold) and the New World stilt-legged *Haringtonhippus francisci* (51.4-fold), followed by the domestic horse (58.8-fold).

A tip-calibrated BEAST analysis (Figure 3.3A,B, Appendix table S4) reconstructed a topology that was consistent with previous results: Old and New World caballoid horses diverged from present-day and extinct stenonids around 3.75-4.34 Myr ago (95% posterior HPD interval, Appendix table S5), preceded by divergence of their sister genus *Haringtonhippus* (4.04-5.03 Myr BP) and South American *Hippidion* (4.74-7.27 Myr BP) further back in time. The Paradise Hill horse fell into the stenonid clade that diversified around 2.48-3.26 Myr BP (Figure 3.3A). While the stenonid clade has high support overall (98%), the resolution of the branches that diverged soon after its split is low. Paradise Hill clustered with the clade of the present-day donkeys, although with low posterior probability support (65%, Figure 3.3A,B).

DISCUSSION

An Early-Middle Pleistocene equid was found in the placer gold mine Paradise Hill, Klondike, Yukon Territory (Figure 3.1A,B). The finding (a cranium and bone fragments) was located below the Gold Run tephra which dated the equid to be around 0.69-0.76 Myr BP, associating it with the Irvingtonian mammal fauna (Buryak et al., 2019; Westgate et al., 2009, 2011). Exceptionally high DNA preservation (Appendix table S2) of its petrous bone allowed us to sequence the genome of the sample up to 12-fold coverage making it more than ten times higher in coverage than the genome of

the Thistle Creek horse dated to around the same time period (Dabney et al., 2013; Meyer et al., 2016; Orlando et al., 2013). We took advantage of the good quality genomic data to place the Paradise Hill horse into the evolutionary context of Pleistocene and present-day equids (Figure 3.1C, Figure 3.2, Figure 3.3A,B, Appendix figure S4). The phylogenetic analysis of nuclear and mitochondrial genomes suggests that the Paradise Hill horse belongs to an archaic stenonid equid lineage, close to the present-day Asiatic wild asses and zebras, and likely represents a species previously unknown to genetics. Phylogenetic affinity of the specimen to the non-caballoid clade is further supported by the genome-wide analysis of the nuclear polymorphisms and improved mapping of reads to non-caballoid reference genomes (Figure 3.2, 3, Appendix figure S2, table S3).

The oldest equid genome is sequenced from the caballoid *E. cf. scotii* specimen found on the Thistle Creek (Orlando et al., 2013). The Thistle Creek specimen, as the Paradise Hill horse, is associated with Gold Run tephra bed. Thistle Creek is located ~100km to the south of the Paradise Hill mine, which together with the same age of the specimens suggests temporal overlap between the stenonid and caballoid horse species. The best-preserved element of the Paradise Hill skeletal assemblage is a cranium, while the Thistle Creek sample is represented by an incomplete metapodial bone, which makes it difficult to compare the specimens morphologically. The genetic signal, however, unequivocally suggests that they belong to two distinct groups that diverged at least 3 Myr before their coexistence in Klondike region (Figure 3.3A). It is not unusual for several equid species to be sympatric. For instance, Stilt Legged

Haringtonhippus diverged from the *Equus* stem before the latter diversified into the stenonid and caballoid groups, and its lineage survived as late as 15 Kyr BP, coexisting across their entire range with caballoid horses (Heintzman et al., 2017). Furthermore, it has been shown previously that non-caballoid equids diverged from caballoid horses in the presence of gene flow that gradually ceased after their dispersal into the Old World (Jónsson et al., 2014). Future research should utilize the archaic Paradise Hill genome to test whether the gene flow between various equids continued until the Middle Pleistocene.

Despite extremely well preserved Late Pleistocene megafauna, and abundant ancient DNA studies of equids inhabiting North America, there have been no Late Pleistocene finds of equids that would belong to the stenonid group. Genetically distant from the Paradise Hill species, *E. cf. scotti* is an extinct lineage of New World caballoid horses, closely-related to the present-day domestic horses (Chapter 2 of this dissertation; Heintzman et al., 2017; Orlando, 2019; Orlando et al., 2013; Vilstrup et al., 2013). Specimens morphologically identified as hemionids, which could have been a source of candidate species for Paradise Hill, genetically belong to *Haringtonhippus* (Harington, 1980, 2011; Heintzman et al., 2017). Lack of known taxa genetically close to Paradise Hill horse suggests that the group it belongs to either went extinct before the Late Pleistocene or was so sparsely distributed across North America that its fossils are extremely rare or non-existent. Studies of bone assemblages found that large-bodied herbivores were the most abundant mammals in the North American Late Pleistocene paleoenvironment, with the local environment supporting their rich biomass (Lorenzen

et al., 2011; Mann et al., 2013, 2015). Thus, the lack of equid groups related to the Paradise Hill in paleontological record of the Late Pleistocene raises questions about the range distribution and population size of stenonid horses during this time. The increasingly unstable Arctic environment of the Middle to Late Pleistocene North America may explain the lack of specimens close to the Paradise Hill genetic lineage (Mann et al., 2018). During rapid cold to warm transitions of glacial-interglacial cycles more morphologically "derived" caballoid and stilt-legged horses had advantage over species with more archaic morphology (such as present-day zebras and extinct South American *Hippidion*), which were able to survive only in lower latitudes (Forsten, 1991; Forsten, 1996; Heintzman et al., 2017). Testing whether the "high altitude turnover" was the case for the Paradise Hill equid lineage requires further morphological analysis of its fossil remains.

What species does Paradise Hill equid belong to? Since none of its relaties have been identified genetically so far, we can only rely on the existing literature on the matter. One species candidate is *Plesippus (Equus) simplicidens* (originally *E. shoshonensis*), or Hagerman's zebra - the most archaic representative of the *Equus* genus that was first found in Pliocene Idaho and whose possible occurrence dates range from 3.3 to 1.7 Myr BP (Forsten & Eisenmann, 1995; Gidley, 1930). *E. simplicidens* is thought to be the most recent archaic form of *Equus* (Azzaroli, 1991; Azzaroli & Voorhies, 1993). Comparative morphological analysis of various Pliocene and Pleistocene equids from Eurasia, Africa and North America trace the origin of the Old World stenonid groups to *E. simplicidens*, which would explain occurrence of horses

phylogenetically close to it in Eastern Beringia (Bernor et al., 2019; Eisenmann & Deng, 2005). DNA survival beyond 1 Myr is highly unlikely (Kistler et al., 2017), however a test of the dental enamel proteome would make possible comparison between the Paradise Hill sample and *E. simplicidens* (Demarchi et al., 2016). Other species evolutionarily close to the Hagerman's zebra, but occurring later in paleontological record, such as *E. idahoensis* and *E. enormis*, are also good candidates for further investigation (Buckley, 2019; Scott et al., 2006; Scott, 2004).

The second species candidate for the Paradise Hill specimen is the Early Pleistocene *E. verae* - a large non-caballoid horse described initially from Chukotka, but believed to be distributed across the north of Holarctic, including the Klondike, Old Crow river (~400km to the North from the Paradise Hill location) and Sixtymile area (~70km to the West) (Boulbes & van Asperen, 2019; Harington, 1990; Harington, 2011). Fossil remains of *E. verae* are dated to 1-0.2 Myr BP, however its close relative, Ovodov's horse *E. ovodovi*, inhabited southern regions of Eurasia and survived until the end of Late Pleistocene in Eastern Asia (Druzhkova et al., 2017; Eisenmann, 2010; Eisenmann & Sergej, 2011; Plasteeva et al., 2015; Yuan et al., 2019). BEAST analysis of the Paradise Hill mitochondrial genome places it together with *E. ovodovi* into the unresolved polytomy with wild asses (Figure 3.3A,B), which likely appears due to the incomplete lineage sorting and\or hybridization between the species (Jónsson et al., 2014; Rosenbom et al., 2015; Steiner et al., 2012; Vilstrup et al., 2013). Nonetheless, from the BEAST analysis we conclude that Paradise Hill horse mitochondrial genome belongs to the haplogroup of stenonid equids. Thus to resolve the outlined hypothesis

a nuclear genome of *E. ovodovi* and ancient DNA data from specimens identified as *E. verae* are necessary.

CONCLUSION

The assignment of the Paradise Hill specimen as a certain species is speculative without additional genomic, proteomic and morphometric data. Nonetheless, the finding is important for paleoecology of the Yukon region. Both Thistle Creek and Paradise Hill horses belong to equid species that together with steppe bison, wooly mammoth, muskox, and other large-bodied herbivores were part of the developing cold-adapted steppe megafauna (Harington, 2011; Mann et al., 2013). Cold-adapted species dominated the Holarctic in the Late Pleistocene, however very little is known about how this community formed during earlier glacial-interglacial cycles (Mann et al., 2015, 2018; Zimov et al., 2012). Ancient DNA studies similar to the one presented here allow going further back in time beyond the relatively well-studied Late Pleistocene (Parducci, 2019). Here we demonstrate that genome-wide studies of Early and Middle Pleistocene fossil remains are possible and paleogenomics should be utilized to track large-scale biodiversity and ecosystem change.

BIBLIOGRAPHY

Alberti, F., Gonzalez, J., Paijmans, J. L. A., Basler, N., Preick, M., Henneberger, K., Trinks, A., Rabeder, G., Conard, N. J., Münzel, S. C., Joger, U., Fritsch, G., Hildebrandt, T., Hofreiter, M., & Barlow, A. (2018). Optimized DNA sampling of ancient bones using Computed Tomography scans. *Molecular Ecology Resources*, *18*(6), 1196–1208.

Azzaroli, A. (1991). Ascent and decline of monodactyl equids: a case for prehistoric overkill. *Annales Zoologici Fennici*, *28*(3/4), 151–163.

Azzaroli, A., & Voorhies, M. R. (1993). The genus Equus in North America. The Blancan species. *Palaeontographia Italica*, *80*, 175–198.

Barrón-Ortiz, C. I., Avilla, L. S., Jass, C. N., Bravo-Cuevas, V. M., Machado, H., & Mothé, D. (2019). What Is Equus? Reconciling Taxonomy and Phylogenetic Analyses. *Frontiers in Ecology and Evolution*, *7*, 343.

Bernor, R. L., Cirilli, O., Jukar, A. M., Potts, R., Buskianidze, M., & Rook, L. (2019). Evolution of Early Equus in Italy, Georgia, the Indian Subcontinent, East Africa, and the Origins of African Zebras. *Frontiers in Ecology and Evolution*, *7*, 166.

Boessenkool, S., Hanghøj, K., Nistelberger, H. M., Der Sarkissian, C., Gondek, A. T., Orlando, L., Barrett, J. H., & Star, B. (2017). Combining bleach and mild predigestion improves ancient DNA recovery from bones. *Molecular Ecology Resources*, *17*(4), 742–751.

Bouckaert, R. R., & Heled, J. (2014). DensiTree 2: Seeing Trees Through the Forest. In *bioRxiv* (p. 012401). https://doi.org/10.1101/012401

Boulbes, N., & van Asperen, E. N. (2019). Biostratigraphy and Palaeoecology of European Equus. *Frontiers in Ecology and Evolution*, *7*, 301.

Buckley, M. (2019). Paleoproteomics: An Introduction to the Analysis of Ancient Proteins by Soft Ionisation Mass Spectrometry. In C. Lindqvist & O. P. Rajora (Eds.), *Paleogenomics: Genome-Scale Analysis of Ancient DNA* (pp. 31–52). Springer International Publishing.

Buryak, S., Reyes, A., Froese, D., & Jensen, B. (2019). Zircon U-Pb dates for Gold Run tephra confirm the antiquity of ancient permafrost and the world's oldest recovered genome. *20th Congress of the International Union for Quaternary Research (INQUA)*. INQUA, Dublin, Ireland. https://app.oxfordabstracts.com/events/574/program-app/submission/94741

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., & Meyer, M. (2013).

Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(39), 15758–15763.

Dabney, J., & Meyer, M. (2019). Extraction of Highly Degraded DNA from Ancient Bones and Teeth. In B. Shapiro, A. Barlow, P. D. Heintzman, M. Hofreiter, J. L. A. Paijmans, & A. E. R. Soares (Eds.), *Ancient DNA: Methods and Protocols* (pp. 25–29). Springer New York.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772.

Demarchi, B., Hall, S., Roncal-Herrero, T., Freeman, C. L., Woolley, J., Crisp, M. K., Wilson, J., Fotakis, A., Fischer, R., Kessler, B. M., Rakownikow Jersie-Christensen, R., Olsen, J. V., Haile, J., Thomas, J., Marean, C. W., Parkington, J., Presslee, S., Lee-Thorp, J., Ditchfield, P., … Collins, M. J. (2016). Protein sequences bound to mineral surfaces persist into deep time. *eLife*, *5*. https://doi.org/10.7554/eLife.17092

Der Sarkissian, C., Ermini, L., Schubert, M., Yang, M. A., Librado, P., Fumagalli, M., Jónsson, H., Bar-Gal, G. K., Albrechtsen, A., Vieira, F. G., Petersen, B., Ginolhac, A., Seguin-Orlando, A., Magnussen, K., Fages, A., Gamba, C., Lorente-Galdos, B., Polani, S., Steiner, C., … Orlando, L. (2015). Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Current Biology: CB*, *25*(19), 2577–2583.

Der Sarkissian, C., Vilstrup, J. T., Schubert, M., Seguin-Orlando, A., Eme, D., Weinstock, J., Alberdi, M. T., Martin, F., Lopez, P. M., Prado, J. L., Prieto, A., Douady, C. J., Stafford, T. W., Willerslev, E., & Orlando, L. (2015). Mitochondrial genomes reveal the extinct Hippidion as an outgroup to all living equids. *Biology Letters*, *11*(3). https://doi.org/10.1098/rsbl.2014.1058

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, *4*(5), e88.

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, *7*, 214.

Druzhkova, A. S., Makunin, A. I., Vorobieva, N. V., Vasiliev, S. K., Ovodov, N. D., Shunkov, M. V., Trifonov, V. A., & Graphodatsky, A. S. (2017). Complete mitochondrial genome of an extinct Equus (Sussemionus) ovodovi specimen from Denisova cave (Altai, Russia). *Mitochondrial DNA Part B*, *2*(1), 79–81.

Eisenmann, V. (1992). Origins, dispersals, and migrations of Equus (Mammalia, Perissodactyla). *Courier Forschungsinstitut Senckenberg*, *153*, 161–170.

Eisenmann, V. (2004). Equus: an evolution without lineages. *Terra Nostra*. https://www.academia.edu/download/50967163/Equus_an_evolution_without_lineages__20161219-24452-2uqqh4.pdf

Eisenmann, V. (2006). Pliocene and Pleistocene Equids: Paleontology versus Molecular Biology. *CFS Courier Forschungsinstitut Senckenberg*, *256*, 71–89.

Eisenmann, V. (2010). Sussemionus, a new subgenus of Equus (Perissodactyla, Mammalia). *Comptes Rendus Biologies*, *333*(3), 235–240.

Eisenmann, V., & Baylac, M. (2000). Extant and fossil Equus (Mammalia, Perissodactyla) skulls: A morphometric definition of the subgenus Equus. *Zoologica Scripta*, *29*(2). https://www.researchgate.net/publication/40663737_Extant_and_fossil_Equus_Mammalia_Perissodactyla_skulls_A_morphometric_definition_of_the_subgenus_Equus

Eisenmann, V., & Deng, T. (2005). Equus qingyangensis (Equidae, Perissodactyla) of the Upper Pliocene of Bajiazui, China: evidence for the North American origin of an Old World lineage distinct from E. stenonis. *Quaternaire: International Journal of the French Quaternary Association*, *2*, 113–122.

Eisenmann, V., & Sergej, V. (2011). Unexpected finding of a new Equus species (Mammalia, Perissodactyla) belonging to a supposedly extinct subgenus in late Pleistocene deposits of Khakassia (Southwestern Siberia). *Geodiversitas*, *33*(3), 519–530.

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K. A. S., Albizuri, S., Alfarhan, A. H., Allentoft, M., Alquraishi, S., Anthony, D., Baimukhanov, N., Barrett, J. H., Bayarsaikhan, J., Benecke, N., Bernáldez-Sánchez, E., … Orlando, L. (2019). Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell*, *177*(6), 1419–1435.e31.

Forsten, A. (1991). Mitochondrial-DNA time-table and the evolution of Equus: comparison of molecular and paleontological evidence. *Annales Zoologici Fennici*. https://www.jstor.org/stable/23735454

Forsten, A. (1996). Climate and the evolution of Equus (Perissodactyla, Equidae) in the Plio-Pleistocene of Eurasia. *Acta Zoologica Cracoviensia*, *39*(1), 161–166.

Forsten, A., & Eisenmann, V. (1995). Equus (Plesippus) Simplicidens (Cope), not Dolichohippus. *Mammalia*, *59*(1), 85–90.

Fulton, T. L., & Shapiro, B. (2019). Setting Up an Ancient DNA Laboratory. *Methods in Molecular Biology* , *1963*, 1–13.

Gaunitz, C., Fages, A., Hanghøj, K., Albrechtsen, A., Khan, N., Schubert, M., Seguin-Orlando, A., Owens, I. J., Felkel, S., Bignon-Lau, O., de Barros Damgaard, P., Mittnik, A., Mohaseb, A. F., Davoudi, H., Alquraishi, S., Alfarhan, A. H., Al-Rasheid, K. A. S., Crubézy, E., Benecke, N., … Orlando, L. (2018). Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science*, *360*(6384), 111–114.

Gidley, J. W. (1930). A New Pliocene Horse from Idaho. *Journal of Mammalogy*, *11*(3), 300–303.

Harington, C. (1990). Vertebrates of the last interglaciation in Canada: A review, with new data. *Géographie Physique et Quaternaire*, *44*(3), 375–387.

Harington, C. R. (1980). Pleistocene mammals from Lost Chicken Creek, Alaska. *Canadian Journal of Earth Sciences*, *17*(2), 168–198.

Harington, C. R. (2011). Pleistocene vertebrates of the Yukon Territory. *Quaternary Science Reviews*, *30*(17), 2341–2354.

Heintzman, P. D., Zazula, G. D., MacPhee, R. D. E., Scott, E., Cahill, J. A., McHorse, B. K., Kapp, J. D., Stiller, M., Wooller, M. J., Orlando, L., Southon, J., Froese, D. G., & Shapiro, B. (2017). A new genus of horse from Pleistocene North America. *eLife*, *6*. https://doi.org/10.7554/eLife.29944

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* , *29*(13), 1682–1684.

Jónsson, H., Schubert, M., Seguin-Orlando, A., Ginolhac, A., Petersen, L., Fumagalli, M., Albrechtsen, A., Petersen, B., Korneliussen, T. S., Vilstrup, J. T., Lear, T., Myka, J. L., Lundquist, J., Miller, D. C., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A. S., Stagegaard, J., Strauss, G., … Orlando, L. (2014). Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), 18655–18660.

Kalbfleisch, T. S., Rice, E., DePriest, M. S., Walenz, B. P., Hestand, M. S., Vermeesch, J. R., O'Connell, B. L., Fiddes, I. T., Vershinina, A. O., Petersen, J. L., Finno, C. J., Bellone, R. R., McCue, M. E., Brooks, S. A., Bailey, E., Orlando, L., Green, R. E., Miller, D. C., Antczak, D. F., & MacLeod, J. N. (2018). EquCab3, an Updated Reference Genome for the Domestic Horse. In *bioRxiv* (p. 306928). https://doi.org/10.1101/306928

Kassambara, A., & Mundt, F. (2017). Factoextra: extract and visualize the results of multivariate data analyses. *R Package Version*, *1*(4), 2017.

Kistler, L., Ware, R., Smith, O., Collins, M., & Allaby, R. G. (2017). A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Research*, *45*(11), 6310–6320.

Koch, P. L., & Barnosky, A. D. (2006). Late Quaternary Extinctions: State of the Debate. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 215–250.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*, 356.

Librado, P., Der Sarkissian, C., Ermini, L., Schubert, M., Jónsson, H., Albrechtsen, A., Fumagalli, M., Yang, M. A., Gamba, C., Seguin-Orlando, A., Mortensen, C. D., Petersen, B., Hoover, C. A., Lorente-Galdos, B., Nedoluzhko, A., Boulygina, E., Tsygankova, S., Neuditschko, M., Jagannathan, V., … Orlando, L. (2015). Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(50), E6889–E6897.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079.

Lorenzen, E. D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., Ugan, A., Borregaard, M. K., Gilbert, M. T. P., Nielsen, R., Ho, S. Y. W., Goebel, T., Graf, K. E., Byers, D., Stenderup, J. T., Rasmussen, M., Campos, P. F., Leonard, J. A., Koepfli, K.-P., … Willerslev, E. (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, *479*(7373), 359–364.

MacFadden, B. J. (1994). *Fossil Horses: Systematics, Paleobiology, and Evolution of the Family Equidae*. Cambridge University Press.

MacFadden, B. J. (2005). Fossil horses - evidence for evolution. *Science*, *307*(5716), 1728–1730.

MacFadden, B. J., Oviedo, L. H., Seymour, G. M., & Ellis, S. (2012). Fossil Horses, Orthogenesis, and Communicating Evolution in Museums. *Evolution: Education and Outreach*, *5*(1), 29–37.

Mann, D. H., Groves, P., Gaglioti, B. V., & Shapiro, B. A. (2018). Climate-driven ecological stability as a globally shared cause of Late Quaternary megafaunal

extinctions: the Plaids and Stripes Hypothesis. *Biological Reviews of the Cambridge Philosophical Society*. https://doi.org/10.1111/brv.12456

Mann, D. H., Groves, P., Kunz, M. L., Reanier, R. E., & Gaglioti, B. V. (2013). Ice-age megafauna in Arctic Alaska: extinction, invasion, survival. *Quaternary Science Reviews*, *70*, 91–108.

Mann, D. H., Groves, P., Reanier, R. E., Gaglioti, B. V., Kunz, M. L., & Shapiro, B. (2015). Life and extinction of megafauna in the ice-age Arctic. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(46), 14301–14306.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

Meisner, J., & Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics*, *210*(2), 719–731.

Meyer, M., Arsuaga, J.-L., de Filippo, C., Nagel, S., Aximu-Petri, A., Nickel, B., Martínez, I., Gracia, A., Bermúdez de Castro, J. M., Carbonell, E., Viola, B., Kelso, J., Prüfer, K., & Pääbo, S. (2016). Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*, *531*(7595), 504–507.

Nakamura, T., Yamada, K. D., Tomii, K., & Katoh, K. (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* , *34*(14), 2490–2492.

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* , *32*(2), 292–294.

Orlando, L. (2015). Equids. *Current Biology: CB*, *25*(20), R973–R978.

Orlando, L. (2019). An Ancient DNA Perspective on Horse Evolution. In C. Lindqvist & O. P. Rajora (Eds.), *Paleogenomics: Genome-Scale Analysis of Ancient DNA* (pp. 325–351). Springer International Publishing.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P. L. F., Fumagalli, M., Vilstrup, J. T., Raghavan, M., Korneliussen, T., Malaspinas, A.-S., Vogt, J., Szklarczyk, D., Kelstrup, C. D., … Willerslev, E. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, *499*(7456), 74–78.

Orlando, L., Metcalf, J. L., Alberdi, M. T., Telles-Antunes, M., Bonjean, D., Otte, M., Martin, F., Eisenmann, V., Mashkour, M., Morello, F., Prado, J. L., Salas-Gismondi, R., Shockey, B. J., Wrinn, P. J., Vasil'ev, S. K., Ovodov, N. D., Cherry, M. I., Hopwood, B., Male, D., … Cooper, A. (2009). Revising the recent evolutionary history of equids using ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(51), 21754–21759.

Parducci, L. (2019). Quaternary DNA: A Multidisciplinary Research Field. *Quaternary*, *2*(4), 37.

Plasteeva, N. A., Vasiliev, S. K., & Kosintsev, P. A. (2015). Equus (Sussemionus) ovodovi Eisenmann et Vasiliev, 2011 from the Late Pleistocene of Western Siberia. *Russian Journal of Theriology*, *14*(2), 187–200.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* , *26*(6), 841–842.

Rambaut, A. (2014). FigTree 1.4. 2 software. *Institute of Evolutionary Biology, Univ. Edinburgh*. https://scholar.google.ca/scholar?cluster=5129645832770813008&hl=en&as_sdt=0,5&sciodt=0,5

Rambaut, A., & Drummond, A. J. (2010). TreeAnnotator version 1.6. 1. *University of Edinburgh, Edinburgh, UK*. https://scholar.google.ca/scholar?cluster=16034997356612446945&hl=en&as_sdt=0,5&sciodt=0,5

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904.

Ramsey, C. B. (2009). Bayesian Analysis of Radiocarbon Dates. *Radiocarbon*, *51*(1), 337–360.

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Reimer, P. J., Bard, E., Bayliss, A., Warren Beck, J., Blackwell, P. G., Ramsey, C. B., Buck, C. E., Cheng, H., Lawrence Edwards, R., Friedrich, M., Grootes, P. M., Guilderson, T. P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T. J., Hoffmann, D. L., Hogg, A. G., Hughen, K. A., … van der Plicht, J. (2013). IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon*, *55*(4), 1869–1887.

Renaud, G., Petersen, B., Seguin-Orlando, A., Bertelsen, M. F., Waller, A., Newton, R., Paillot, R., Bryant, N., Vaudin, M., Librado, P., & Orlando, L. (2018).

Improved de novo genomic assembly for the domestic donkey. *Science Advances*, *4*(4), eaaq0392.

Rosenbom, S., Costa, V., Chen, S., Khalatbari, L., Yusefi, G. H., Abdukadir, A., Yangzom, C., Kebede, F., Teclai, R., Yohannes, H., Hagos, F., Moehlman, P. D., & Beja-Pereira, A. (2015). Reassessing the evolutionary history of ass-like equids: insights from patterns of genetic variation in contemporary extant populations. *Molecular Phylogenetics and Evolution*, *85*, 88–96.

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* , *27*(6), 863–864.

Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J. F., Al-Rasheid, K. A. S., Willerslev, E., Krogh, A., & Orlando, L. (2012). Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics*, *13*, 178.

Schubert, M., Jónsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., Albrechtsen, A., Dupanloup, I., Foucal, A., Petersen, B., Fumagalli, M., Raghavan, M., Seguin-Orlando, A., Korneliussen, T. S., Velazquez, A. M. V., Stenderup, J., Hoover, C. A., Rubin, C.-J., Alfarhan, A. H., … Orlando, L. (2014). Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), E5661–E5669.

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes*, *9*, 88.

Scott, E. (2004). Pliocene and Pleistocene Horses from Porcupine Cave. In *Biodiversity Response to Climate Change in the Middle Pleistocene*.

Scott, E., Jefferson, G. T., & Lindsey, L. (2006). Extinct horses and their relatives. *Fossil Treasures of the Anza-Borrego Desert (GT Jefferson and L. Lindsay, Editors). Sunbelt Publications, Inc. , San Diego, California*, 253–271.

Shorthouse, D. P. (2010). *SimpleMappr*. https://www.simplemappr.net

Stadler, T. (2010). Sampling-through-time in birth-death trees. *Journal of Theoretical Biology*, *267*(3), 396–404.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* , *30*(9), 1312–1313.

Steiner, C. C., Mitelberg, A., Tursi, R., & Ryder, O. A. (2012). Molecular phylogeny of extant equids and effects of ancestral polymorphism in resolving species-level phylogenies. *Molecular Phylogenetics and Evolution*, *65*(2), 573–581.

Troll, C. J., Kapp, J., Rao, V., Harkins, K. M., Cole, C., Naughton, C., Morgan, J. M., Shapiro, B., & Green, R. E. (2019). A ligation-based single-stranded library preparation method to analyze cell-free DNA and synthetic oligos. *BMC Genomics*, *20*(1), 1023.

Vershinina, A. O., Kapp, J. D., Baryshnikov, G. F., & Shapiro, B. (2019). The case of an arctic wild ass highlights the utility of ancient DNA for validating problematic identifications in museum collections. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13130

Vilstrup, J. T., Seguin-Orlando, A., Stiller, M., Ginolhac, A., Raghavan, M., Nielsen, S. C. A., Weinstock, J., Froese, D., Vasiliev, S. K., Ovodov, N. D., Clary, J., Helgen, K. M., Fleischer, R. C., Cooper, A., Shapiro, B., & Orlando, L. (2013). Mitochondrial phylogenomics of modern and ancient equids. *PloS One*, *8*(2), e55950.

Wade, C. M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T. L., Adelson, D. L., Bailey, E., Bellone, R. R., Blöcker, H., Distl, O., Edgar, R. C., Garber, M., Leeb, T., Mauceli, E., MacLeod, J. N., Penedo, M. C. T., Raison, J. M., … Lindblad-Toh, K. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science*, *326*(5954), 865–867.

Weinstock, J., Willerslev, E., Sher, A., Tong, W., Ho, S. Y. W., Rubenstein, D., Storer, J., Burns, J., Martin, L., Bravi, C., Prieto, A., Froese, D., Scott, E., Xulong, L., & Cooper, A. (2005). Evolution, systematics, and phylogeography of pleistocene horses in the new world: a molecular perspective. *PLoS Biology*, *3*(8), e241.

Westgate, J. A., Preece, S. J., Froese, D. G., Telka, A. M., Storer, J. E., Pearce, N. J. G., Enkin, R. J., Jackson, L. E., LeBarge, W., & Perkins, W. T. (2009). Gold Run tephra: a Middle Pleistocene stratigraphic and paleoenvironmental marker across west-central Yukon Territory, Canada. *Canadian Journal of Earth Sciences*, *46*(6), 465–478.

Westgate, J. A., Preece, S. J., & Jackson, L. E. (2011). Revision of the tephrostratigraphy of the lower Sixtymile River area, Yukon Territory, Canada. *Canadian Journal of Earth Sciences*, *48*(3), 695–701.

Yuan, J.-X., Hou, X.-D., Barlow, A., Preick, M., Taron, U. H., Alberti, F., Basler, N., Deng, T., Lai, X.-L., Hofreiter, M., & Sheng, G.-L. (2019). Molecular identification of late and terminal Pleistocene Equus ovodovi from northeastern China. *PloS One*, *14*(5), e0216883.

Zimov, S. A., Zimov, N. S., Tikhonov, A. N., & Chapin, F. S. (2012). Mammoth steppe: a high-productivity phenomenon. *Quaternary Science Reviews*, *57*, 26–45.

Table 3.1. Equid specimens and their genome sequencing data, listed by taxonomic groups.
Unless specified with a star, age estimates are reported as radiocarbon dates before present, calibrated with IntCal13 as implemented in OxCal 4.3 (Ramsey, 2009; Reimer et al., 2013). Median and sigma values for 99.7 calibration probability range are reported. Ages marked with a star are based on stratigraphic association with the Gold Run volcanic ash layer. Mean genome coverage depth is reported based on mapping to the horse EquCab2.0 reference genome.

Specimen ID	Species	Common name	Age estimate	Age error	genome coverage (fold)	Reference
Caballoid horses	**domestic**					
Arabian	*Equus ferus caballus*	Domestic horse	Present-day		11.0	Orlando et al. 2013
Icelandic	*Equus ferus caballus*	Domestic horse	Present-day		8.4	Orlando et al. 2013
Norwegian Fjord	*Equus ferus caballus*	Domestic horse	Present-day		7.9	Orlando et al. 2013
Standardbred	*Equus ferus caballus*	Domestic horse	Present-day		12.2	Orlando et al. 2013
Thoroughbred (Twilight)	*Equus ferus caballus*	Domestic horse	Present-day		21.0	Orlando et al. 2013
	wild					
PRZ	*E. ferus przewalskii*	Przewalski's horse	Present-day		9.1	Orlando et al. 2013
	ancient					
Batagai	*E. ferus*	Ancient Taymyr horse	5,104	109	18.3	Librado et al. 2015
CGG10022	*E. ferus*	Ancient Taymyr horse	42,692	891	24.3	Orlando et al. 2013; Schubert et al., 2014

CGG10023	*E. ferus*	Ancient Taymyr horse	16,112	96	7.4	Schubert et al., 2014
UP10.MS258 (YG303.325)	*E. ferus*	Ancient Klondike horse	30,292	245	24.8	Chapter 2
CGG_1_011850 (YT03-40)	*E. ferus*	Ancient Klondike horse	27,947	130	25.1	Chapter 2
CGG_1_011849 (YT03-185)	*E. ferus*	Ancient Klondike horse	35,572	237	0.08	this study
TC21 (YG148.2)	*E. cf. scotti*	Thistle Creek horse	* 560–780 kyr BP		0.74	Orlando et al. 2013
Stenonid horses						
YG412.17	*Equus sp.*	Paradise Hill horse	*690–760 kyr BP		10.5	this study
BOE	*E. quagga boehmi*	Plains zebra	Present-day		27.7	Jónsson et al. 2014
GRE	*E. grevyi*	Grevy's zebra	Present-day		22.6	Jónsson et al. 2014
HAR	*E. zebra hartmannae*	Hartmann's mountain zebra	Present-day		22.9	Jónsson et al. 2014
KIA	*E. kiang*	Tibetan wild ass	Present-day		13.1	Jónsson et al. 2014
ONA	*E. hemionus onager*	Asiatic wild ass (onager)	Present-day		27.3	Jónsson et al. 2014
SOM	*E. asinus somalicus*	Somali wild ass	Present-day		33.8	Jónsson et al. 2014
DON (Willy)	*E. asinus asinus*	Domestic donkey	Present-day		24.8	Renaud et al. 2018

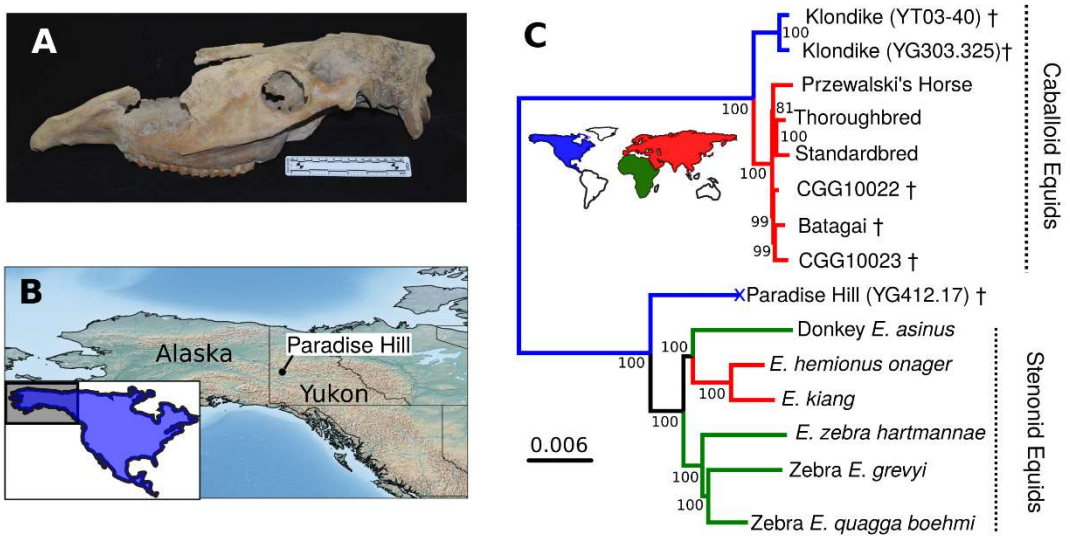


Figure 3.1. The Early-Middle Pleistocene Paradise Hill horse. **A.** A cranium of the equid specimen YG412.17. **B.** Geographical locality of the finding - Klondike Paradise Hill site, Canada. Drawn using SimpleMappr (Shorthouse, 2010) **C.** Phylogenetic relationships between present-day and ancient caballoid and stenonid equids. Maximum likelihood analysis of 23,694 noncoding putatively-neutral 1kb loci. Genomes recovered from paleontological samples are marked with daggers.

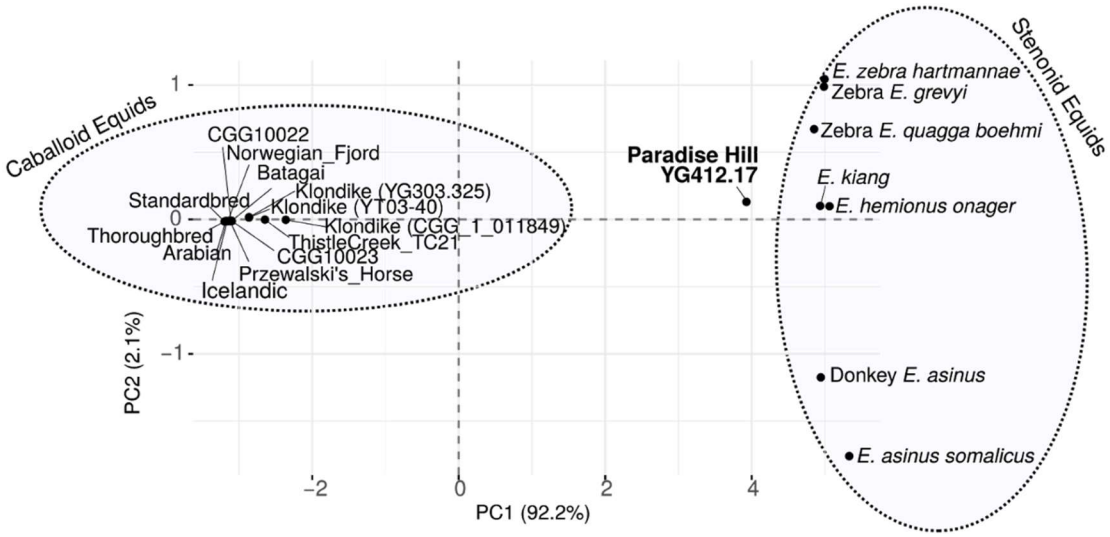


Figure 3.2. PCA on genotype likelihoods estimated on low and high coverage genomes of present-day and extinct equids. Genotype likelihoods were called only on transversions. The proportion of variance explained by principal components is indicated on the X- and Y-axes.

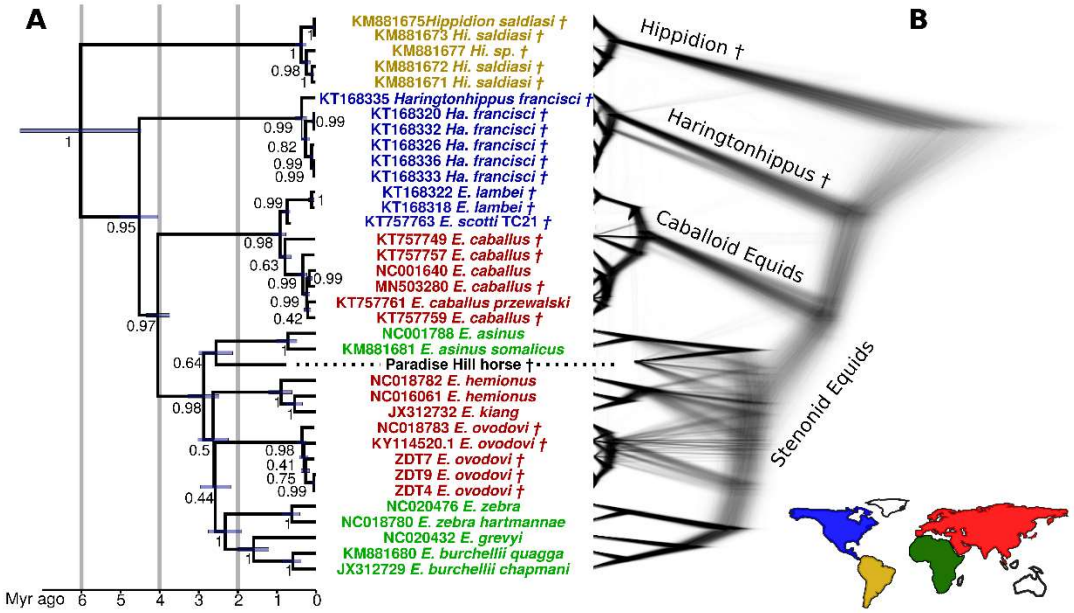


Figure 3.3. Mitochondrial phylogeny of extant and Early-Middle and Late Pleistocene equids inferred with BEAST. Genomes recovered from paleontological samples are marked with daggers. **A.** Maximum clade credibility tree. Purple bars are 95% highest posterior density intervals of node heights are indicated with numbers. **B.** Phylogenetic uncertainty visualized with Densitree. Each thin line is a topology reconstructed during BEAST runs.

APPENDICES

Table S1. Summary sequencing statistics for the Paradise Hill horse DNA libraries. Reads passing filters are paired-end sequencing reads that have a proper adapter on both ends and are longer than 27bp after adapter-trimming.

Extract ID	Library ID	# sequencing clusters	Mean sequencing quality score (Phred)	# reads passing filters
JKSC77	JKSC77	225,478,873	34.87	182,921,151
JKSC78	JKSC78	186,094,891	35.49	148,929,751
JKSC79	JKSC79	172,008,301	34.49	141,807,007
JKSC80	JKSC80	203,707,471	35.72	164,675,706
AV214	AV214-1	283,536,328	34.9	203,185,788
AV215	AV215-1	247,176,057	35.88	201,383,338
AV216	AV216-1	261,086,859	35.07	200,014,485
AV217	AV217-1	350,182,303	35.8	252,160,804
AV218	AV218-1	276,107,539	35.5	238,448,403
AV219	AV219-1	144,224,178	35.77	119,240,304
AV220	AV220-1	218,643,266	35.68	182,291,419
AV214	AV214-2	148,340,456	35.64	103,910,624
AV215	AV215-2	170,370,236	35.44	116,896,155

AV216	AV216-2	135,679,135	36.06	108,597,390
AV217	AV217-2	149,990,467	35.88	126,247,363
AV218	AV218-2	117,746,542	35.4	95,842,009
AV219	AV219-2	135,348,626	35.62	90,399,881
AV220	AV220-2	110,587,391	35.85	76,868,884
Pool 1	AVcombo1-1	93,990,564	35.84	65,715,044
Pool 2	AVcombo2-1	122,373,094	35.14	91,760,405
Pool 3	AVcombo3-1	158,368,971	35.24	115,939,695

Table S2. Summary mapping statistics for the Paradise Hill horse DNA libraries.

Library ID	EquCab2 reference			EquCab3 reference			Donkey reference		
	reads mapped total	reads mapped, no duplicates	% unique mapped reads	reads mapped total	reads mapped, no duplicates	% unique mapped reads	reads mapped total	reads mapped, no duplicates	% unique mapped reads
JKSC77	80,121,211	62,447,088	77.9	80,350,339	62,529,699	77.8	85,303,202	64,591,030	75.7
JKSC78	66,685,731	52,980,095	79.4	66,851,729	53,036,168	79.3	70,916,666	54,793,389	77.3
JKSC79	58,860,705	46,350,191	78.7	59,008,469	46,401,866	78.6	62,648,019	47,941,741	76.5
JKSC80	72,715,542	56,552,504	77.8	72,875,491	56,590,711	77.7	77,340,877	58,452,471	75.6
AV214-1	67,192,771	26,588,688	39.6	67,748,252	26,752,575	39.5	72,365,251	27,863,169	38.5
AV215-1	57,552,421	23,318,185	40.5	58,061,118	23,466,143	40.4	62,034,758	24,483,622	39.5
AV216-1	68,236,869	33,847,396	49.6	68,742,670	34,018,948	49.5	73,508,505	35,521,539	48.3
AV217-1	60,731,727	29,681,015	48.9	61,209,533	29,848,388	48.8	65,504,371	31,157,365	47.6
AV218-1	60,681,011	25,306,970	41.7	61,100,747	25,429,387	41.6	64,976,519	26,479,658	40.8
AV219-1	39,153,962	8,438,715	21.6	39,452,498	8,502,115	21.6	42,007,730	8,843,803	21.1
AV220-1	53,293,576	16,848,192	31.6	53,733,925	16,961,700	31.6	57,330,785	17,690,038	30.9
AV214-2	36,721,178	25,326,309	69.0	36,955,050	25,442,746	68.8	39,478,990	26,601,560	67.4
AV215-2	44,258,746	27,759,833	62.7	44,565,083	27,899,047	62.6	47,562,959	29,112,929	61.2
AV216-2	39,827,575	27,437,187	68.9	40,043,905	27,539,290	68.8	42,723,626	28,820,751	67.5
AV217-2	33,577,009	21,891,628	65.2	33,744,005	21,969,518	65.1	35,947,357	22,989,258	64.0

AV218-2	30,232,436	13,695,348	45.3	30,382,222	13,747,841	45.2	32,261,204	14,330,768	44.4
AV219-2	33,753,307	20,472,867	60.7	33,989,923	20,576,658	60.5	36,253,145	21,439,411	59.1
AV220-2	27,676,030	14,782,956	53.4	27,870,864	14,864,885	53.3	29,735,034	15,512,263	52.2
AVcombo1-1	23,483,986	16,289,163	69.4	23,665,120	16,387,936	69.2	25,305,696	17,121,278	67.7
AVcombo2-1	30,025,542	18,258,288	60.8	30,211,603	18,343,149	60.7	32,216,019	19,175,162	59.5
AVcombo3-1	40,285,494	27,617,535	68.6	40,518,095	27,726,147	68.4	43,171,136	28,924,869	67.0

Table S3. Mapping Paradise Hill horse sequencing reads on various mitochondrial genome references.
Sequencing reads were mapped to mitochondrial genomes downloaded from GenBank NCBI. The number of unique reads mapped is estimated with BWA alignment and MIA assembly program. Coverage is reported for the Paradise Hill mitogenome assembled with MIA.

Reference species	**NCBI ID**	**# mapped (BWA)**	**# mapped (MIA)**	**assembly coverage (x-fold)**
Hippidion saldiasi	KM881673	13986	13546	38.4
Equus kiang	NC_020433	23826	23511	75.3
Haringtonhippus francisci	KT168336	18101	17512	51.4
Equus zebra	NC_020476	22703	22317	69.8
Equus caballus	NC_001640	20179	19719	58.8
Equus hemionus onager	JX312730.1	23849	23494	75.2
Equus asinus	KX669267.1	22133	21799	67.1

Table S4. Mitochondrial genomes used for phylogenetic analysis.
NCBI numbers (where available), taxonomic position, uncalibrated radiocarbon age (if published), and calendar dates used for BEAST analysis. The star (*) marks specimen ages estimated by stratigraphic dating. All other non-present day specimens are radiocarbon dated.

Specimen ID	Species	Age (uncalibrated)	Calendar date BP (IntCal13 median)
Stenonid equids			
Hemiones			
NC016061	*Equus hemionus*	Present day	0
NC018782	*E. hemionus*	Present day	0
JX312732	*E. kiang*	Present day	0
Donkeys			
KM881681	*E. asinus somalicus*	Present day	0
NC001788	*E. asinus*	Present day	0
Zebras			
NC018780	*E. zebra hartmannae*	Present day	0
NC020476	*E. zebra*	Present day	0
NC020432	*E. grevyi*	Present day	0
JX312729	*E. burchellii chapmani*	Present day	0
KM881680	*E. burchellii quagga*	Present day	0
Extinct stenonid equids			
YG412.17	*Equus sp.*	*690,000–760,000	*700,000
NC018783	*E. ovodovi*	*45,000	*45,000
KY114520.1	*E. ovodovi*	Not available	0
ZDT4	*E. ovodovi*	Not available	29,206
ZDT7	*E. ovodovi*	Not available	39,525
ZDT9	*E. ovodovi*	Not available	12,683
Caballoid equids			
MN503280	*E. caballus*	19,470±70	23,457
KT757749	*E. caballus*	28,800±1100	32,918

KT757757	*E. caballus*	34,460±240	38,961
KT757759	*E. caballus*	13,940±55	16,908
KT757761	*E. caballus przewalski*	Present day	0
NC001640	*E. caballus*	Present day	0
KT757763	*E. scotti*	*560,000-780,000	*650,000
KT168318	*E. lambei*	33,760±400	38,138
KT168322	*E. lambei*	21,420±80	25,749
New World Stilt Legged horses			
KT168320	*Haringtonhippus francisci*	28,740±570	32,767
KT168326	*H. francisci*	46,500±1900	47,770
KT168332	*H. francisci*	33,400±430	37,664
KT168333	*H. francisci*	33,560±440	37,858
KT168335	*H. francisci*	14,450±90	17,598
KT168336	*H. francisci*	28,390±240	32,311
Extinct South American equids			
KM881671	*Hippidion saldiasi*	13,990±150	16,980
KM881672	*H. saldiasi*	11,900±60	13,709
KM881673	*H. saldiasi*	13,890±60	16,830
KM881675	*H. saldiasi*	10,680±40	12,653
KM881677	*H. sp*	13,275±30	15,961

Table S5. Divergence times estimated with BEAST analysis of mitochondrial genomes.

Summary Statistic	Caballoid\Stenonid split	Hippidion\(Equus+Haringtonhippus) split	Haringtonhippus\Equus split	Paradise Hill \present-day stenonid equids split
mean	4.05E+06	6.04E+06	4.53E+06	2.87E+06
Standard Error of mean	1298.1751	9767.4691	2262.2679	2168.8905
SD	1.51E+05	6.43E+05	2.52E+05	2.01E+05
variance	2.28E+10	4.13E+11	6.37E+10	4.04E+10
median	4.05E+06	6.04E+06	4.52E+06	2.87E+06
geometric mean	4.04E+06	6.01E+06	4.52E+06	2.87E+06
95% HPD Interval	[3.7457E6, 4.342E6]	[4.7379E6, 7.2754E6]	[4.0424E6, 5.0342E6]	[2.4796E6, 3.2597E6]
auto-correlation time (ACT)	11112.1037	34662.1711	12045.1565	17485.5863
effective sample size (ESS)	13500.594	4328.0613	12454.7988	8579.638

Figure S1. Stratigraphic description of the Paradise Hill site and location of the horse cranium.

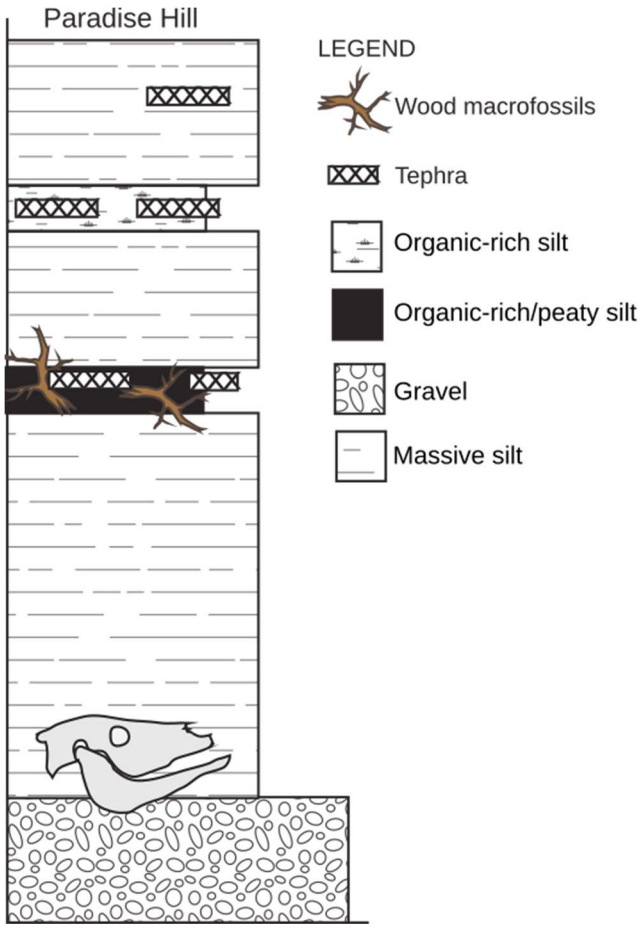


Figure S2. Depth of coverage distribution for Paradise Hill horse sequencing reads. Paradise Hill horse sequencing reads mapped to a donkey and two domestic horse reference genomes.

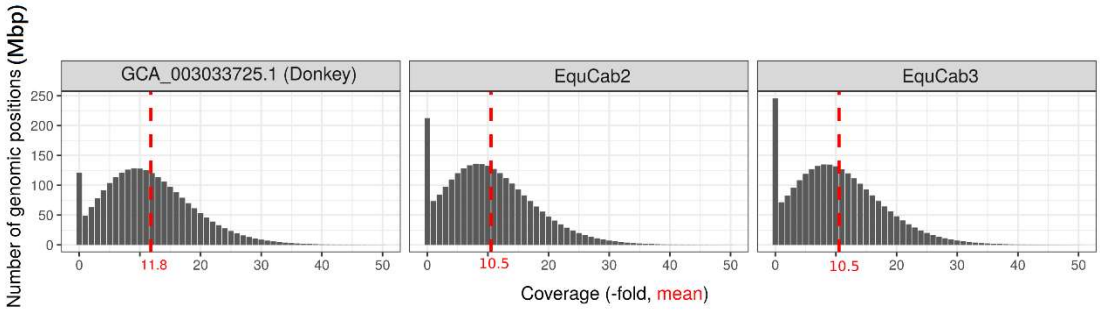


Figure S3. DNA degradation in Paradise Hill genome.
Deamination profile and fragment length distribution in the exemplary DNA library (AV214-1). **A.** Histogram of fragment length distribution. **B,C,D.** Frequency of deaminated sites at the ends of the molecules. **E.** Deamination profile along the sequencing reads.

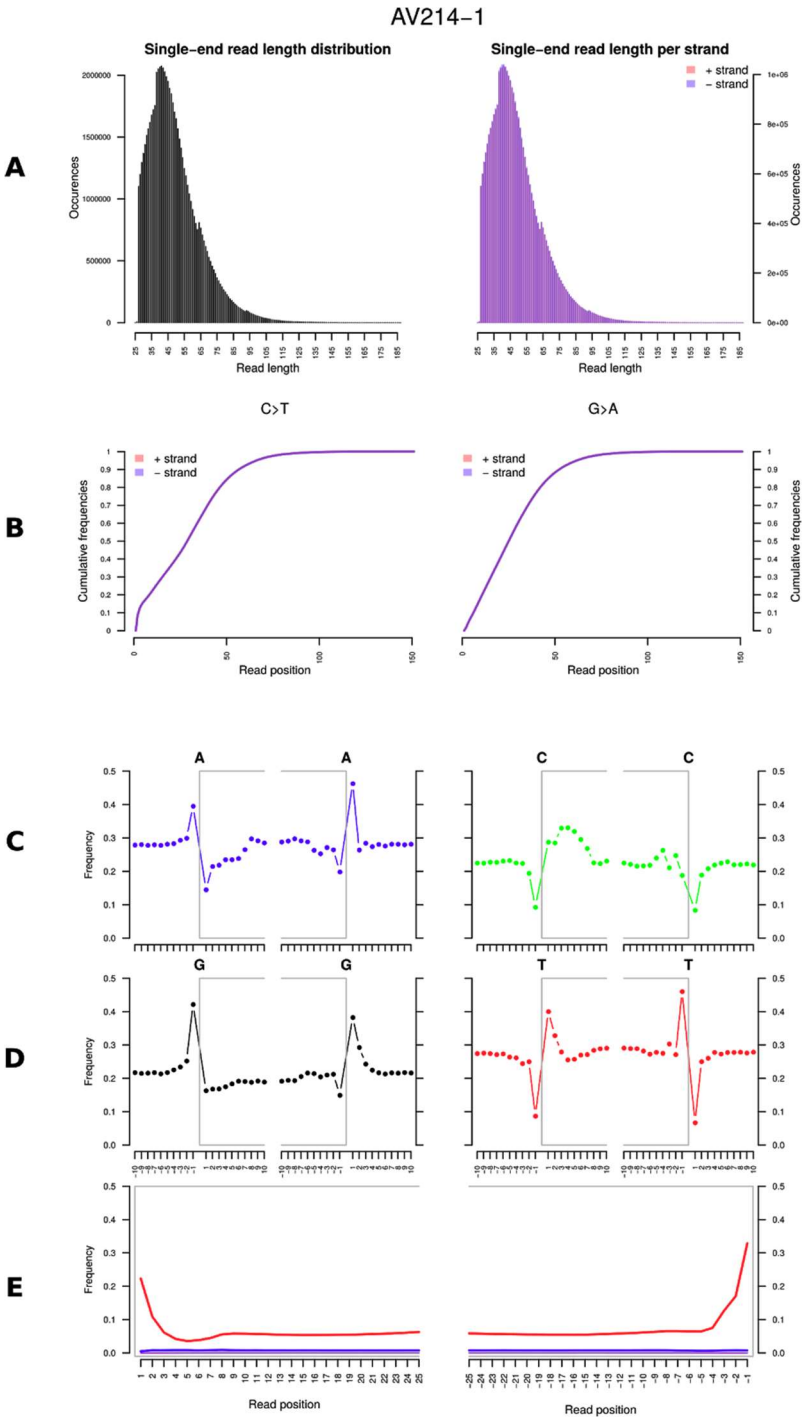
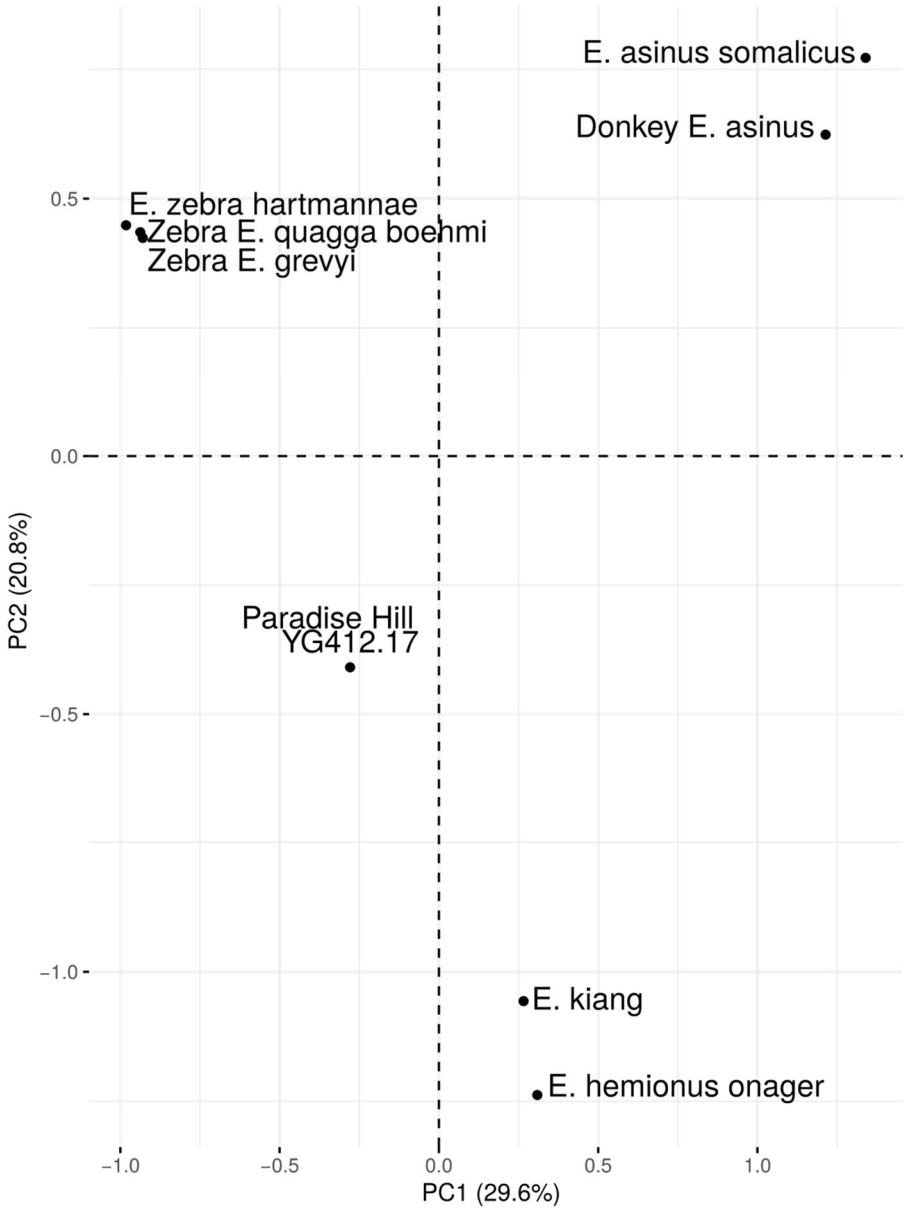


Figure S4. PCA on genotype likelihoods estimated on genomes of stenonid equids. Genotype likelihoods were called only on transversions. The proportion of variance explained by principal components is indicated on the X- and Y-axes.



Chapter 4

Ancient DNA analysis of a Holocene bison from the Rauchua River, North Western Chukotka, and the existence of a deeply divergent mitochondrial clade

Alisa Vershinina[1], Joshua D. Kapp[1], André Elias Rodrigues Soares[2], Peter D. Heintzman[3], Christian Lowson[1], Molly Cassatt-Johnstone[1], Fedor K. Shidlovskiy[4], Irina V. Kirillova[4], Beth Shapiro[1,5]

[1] Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064, USA

[2] Laboratório Nacional de Computação Científica, Rua Getúlio Vargas 333, Quitandinha, Petrópolis, RJ 25651-070, Brazil

[3] Tromsø University Museum, UiT - The Arctic University of Norway, Tromsø NO-9037, Norway

[4] "National Alliance of Shidlovskiy "Ice Age", All-Russian Exhibition Centre, Bld. 71, Moscow 129223, Russia

[5] Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz CA 95064, USA

Zoologicheskiĭ zhurnal (2019), 98(10): 1091-1099

ABSTRACT

The remains of a Holocene extinct steppe bison, *Bison priscus*, that died 9.5 thousand years ago were discovered on the Rauchua River (Chukotka, Russia) in 2012. The sample F-3246 yielded ancient DNA and, when compared to other extant and extinct *Bison* lineages, clustered outside of known bison genetic diversity, suggesting that this bison represented a divergent and heretofore unknown lineage of extinct bison. While this conclusion is supported by morphological analysis of sample F-3246, additional examples of this divergent bison phenotype and genotype are required in order to understand its position in bison evolutionary history. Here, we assemble complete mitochondrial genomes from 26 additional ancient *Bison* samples from Northern Siberia, aiming to improve knowledge of the evolutionary history and geographic range of the Rauchua bison lineage. Surprisingly, we did not identify the Rauchua haplotype in any of the newly tested bison, including those discovered from the Rauchua River locality. Nonetheless, additional mitochondrial genomic data from Siberian bison, and a new high-coverage mitogenome recovered from the original Rauchua bison (F-3246), confirm the existence of the deeply divergent clade and shed new light on the evolutionary history of bison during the Pleistocene to Holocene transition in Siberia.

INTRODUCTION

The remains of a Late Pleistocene steppe bison, *Bison priscus* were discovered by the Rauchua River, North-Western Chukotka, Russia, in 2012 (Kirillova et al., 2015). The sample F-3246 was well preserved, yielded an uncalibrated radiocarbon age

of 9497 ± 92 years before present (BP) (AA101271), and contained sufficient endogenous ancient DNA (aDNA) to assemble a complete mitochondrial genome with 163x coverage. When compared to data generated from a large dataset of extant and extinct *Bison* lineages from across the Northern Hemisphere, the Rauchua bison fell outside of, but sister to, known steppe bison genetic diversity.

Morphological data from specimen F-3246 also support its belonging to a divergent bison clade. The specimen includes an almost complete carcass with remains of soft tissues, and differs morphologically from other known ancient bison (Kirillova et al., 2015). Although the specimen is an adult bison, it is distinctively more gracile than any other bison known from Siberia and North America. While the length of the long bones are similar to other Pleistocene bison, its metapodials are longer.

The phylogenetic position of the Rauchua lineage, augmented by additional genetic evidence, led us to conclude that specimen F-3246 represents an extinct lineage that diverged from all other bison prior to the expansion of bison into eastern Beringia around 160,000 years ago (Froese et al., 2017; Kirillova et al., 2015). Based on the Holocene age of F-3246, its population may have been one of the last to survive past the Pleistocene in the Northern Siberian refuge.

The goal of our current study was twofold. First, because the Rauchua bison haplotype was so diverged from other bison haplotypes, we aimed to replicate and therefore confirm the results generated previously from the Rauchua bison using additional bone elements that were shipped and processed separately. Second, we aimed to build on the previous results from the Rauchua bison study to provide new

insights into bison taxonomy and evolution by generating and analyzing DNA from additional ancient bison remains. Specifically, we aimed to isolate mitochondrial DNA from other bison samples from the region of the Rauchua River, as well as from newly collected samples from Chukotka and other locations in the Russian Far East. Using data recovered from these specimens, we infer a mitochondrial phylogeny and explore the phylogenetic diversity of bison in Western Beringia during the last Ice Age and the beginning of Holocene warming, with a particular focus on the divergent Rauchua lineage.

MATERIALS AND METHODS

We extracted ancient DNA from 29 bison samples (collection of the "National Alliance of Shidlovskiy "Ice Age") from five locations (Figure 4.1, Table 4.1) following the protocol described by (Dabney et al., 2013), with in-house modifications to reduce contamination (Korlević et al., 2015). We then prepared double-stranded DNA sequencing libraries from these extracts following (Meyer & Kircher, 2010). We cleaned the resulting libraries with SPRI beads in 18% PEG-8000 solution and sequenced each on an Illumina MiSeq machine using the v3 chemistry and a paired-end 2x75 cycle strategy. Most libraries were barcoded using only one index, although some (F-1267 - F-1297) were dual-indexed to avoid index swaps and incorrect read identification (Costello et al., 2018). Libraries were pooled so as to recover at least 500,000 reads from each library. To avoid contamination, all protocols preceding indexing PCR were executed in the PCR-free sterile ancient DNA laboratory facility at the Paleogenomics lab, University of California Santa Cruz. After shotgun sequencing,

we then assessed the quality and authenticity of each DNA extract to learn (1) endogenous content (proportion of DNA assigned to *Bison*); (2) library complexity (diversity of molecules recovered during library preparation); (3) average DNA fragment length (expected to be less than 100bp due to DNA fragmentation in samples that are older than 5,000 years); and (4) deamination profile (cytosine damage reflecting DNA degradation).

Using SeqPrep (https://github.com/jstjohn/SeqPrep), we removed adapters and then merged reads that overlapped by at least 15bp and discarded reads shorter than 28bp to allow confident mapping. We then mapped all non-discarded reads to the bison nuclear genome UMD 1.0 (GenBank ID GCF_000754665.1) using BWA 'aln' with the seeding option disabled and a minimum mapping Phred quality score of Q>20 (Li & Durbin, 2009). We then checked the authenticity of the aDNA reads (exploring cytosine deamination patterns) and the fragment size distribution using MapDamage2 (Jónsson et al., 2013).

After determining whether each extract was sufficiently well preserved to proceed, we enriched each library for bison mitochondrial DNA using an in-solution hybridization capture protocol with bison mitogenome RNA baits (Arbor Biosciences, Ann Arbor, USA, previously MYcroarray), following the MYbaits recommended protocol versions 2 and 3, except that the hybridisation step was extended to 36 hours. The enriched libraries were cleaned and pooled following the strategy described above, and sequenced on the Illumina MiSeq.

To process sequenced bait-captured libraries we followed the same protocol as described above, but using *B. bonasus* NC_014044 mitogenome as a reference for mapping (Zeyland et al., 2012). Finally, we assembled complete mitogenomes using the Mapping Iterative Assembler MIA (Briggs et al., 2009), as described in (Heintzman et al., 2016).

We aligned the 16 newly assembled *Bison* mitochondrial genomes to a set of 29 previously published extinct and extant *Bison* species and other bovids using Muscle (Edgar, 2004). For phylogenetic reconstruction, we split the alignments into four partitions and identified the most appropriate models of nucleotide substitution for each partition using jModeltest2 (Darriba et al., 2012). The following setup was used according to results of the test: HKY+G for protein coding sequences, HKY+I+G for D-loop, GTR+G for rRNAs, and HKY+G for tRNAs.

Finally, we performed phylogenetic inference on the partitioned dataset using RAxML (Stamatakis, 2014) and MrBayes (Ronquist et al., 2012) as described in (Maschenko et al., 2017), but with the nucleotide models listed above.

RESULTS

We extracted aDNA and prepared next-generation sequencing libraries from four new specimens collected at the mouth of the Rauchua River (North-Western Chukotka), eight specimens from the Middle Indigirka River (North Yakutia), and 14 specimens from the Bilibinsky region, North-Western Chukotka: Maly Anuy River; Ostrovnovskaya tundra - to the West of the mouth of the Rauchua River; the East Siberian Sea coast - 200 km to the West of the Rauchua River mouth (Figure 4.1, Table

4.1). In addition, we also re-extracted DNA from three new F-3246 sub-samples (the original Rauchua lineage individual). Most of the bison samples listed in Table 4.1 were bones, with the exclusion of two samples of horn sheath that are discussed below.

As expected, aDNA preservation dramatically varied between samples (Figure 4.2, Table 4.1). DNA preservation is often reported as a fraction of sequenced reads that map to a conspecific or closely related genome (endogenous proportion of DNA) in relation to the fraction of all other (exogenous, i.e. contaminant) reads that do *not* map to the host genome. Sources of exogenous DNA include co-extracted co-preserved DNA from plants, animals, and microbes that invade the sample after death. One of the subsamples of the the original Rauchua River sample, subsample F-3246/12, was the best preserved among the three tested, with 55% of recovered DNA mapping to the *Bison* nuclear genome. Bones sampled from the Middle Indigirka River were the least well preserved. From this set of samples, F-1291 and F-1294 (Figure 4.3, Table 4.1) yielded about 1% endogenous DNA. Interestingly, both of these are horn sheath tissue, rather than bone. Samples from Maly Anuy River, Ostrovnovskaya tundra, and East Siberian Sea coast show very good DNA preservation (Table 4.1) based on endogenous bison DNA.

The Rauchua bison lineage was an unexpected finding. We decided that, as part of our undertaking to identify additional samples belonging to this lineage, we would return to the original sample and replicate these results, using samples that were sent and processed entirely independently of the original sample. We extracted new DNA from this specimen using three different bones (F-3246/9, F-3246/12, F-3246/13): right

humerus, right fore autopodia, and left ulna. All three DNA extracts were successfully enriched, producing high coverage mitogenomes (Table 4.1). None of these mitogenomes showed discrepancies with the original published Rauchua haplotype, suggesting that the Rauchua bison lineage as reported is valid and is not an artifact of mitochondrial genome assembly, contamination, or other sources of error.

Despite poor preservation of bison DNA in half of the newly processed bison samples, we were able to enrich most DNA libraries for sufficient quantities of mitochondrial DNA to assemble mitogenomes to >3X coverage for 16 samples, which we then used for phylogenetic analysis. Both Bayesian and Maximum Likelihood approaches produced consistent tree topologies (Figure 4.3). All newly assembled bison mitochondrial lineages fall within the known diversity of steppe bison mitochondrial genomes. None of the new samples clustered with the original Rauchua lineage, including a new sample from the Rauchua River mouth (F-4171), the same locality where the original Rauchua bison was found.

Two samples, F-1297 (from the Middle Indigirka River) and F-4172 (from the Rauchua River mouth), have very little DNA preserved, and resulted in an average mitochondrial coverage of around 1x. While we were unable to include them in our phylogenetic reconstructions, we were nonetheless able to genotype these individuals from the portions of the mitochondrial genome that were recovered. Using the Rauchua mitochondrial haplotype and a panel of extinct *B.priscus* as a reference, we identified nucleotide substitutions in the mitochondrial genome that were distinct from other bison and unique to the Rauchua lineage. Most of these substitutions are located in the

cytochrome oxidase unit 1, the gene that has been used as a barcode for confident identification of various taxa (Luo et al., 2011). We manually inspected those fragments of the poorly preserved F-1297 and F-4172 mitochondrial genomes that were covered by a minimum of three reads (to ensure that any differences were supported by independent reads and not therefore due to DNA damage or sequencing error) looking for informative substitutions. In both samples we identified 35 informative nucleotide sites that were both covered by at least three reads and span nucleotide positions that would have Rauchua-specific substitutions. Neither F-1297 nor F-4172 had the Rauchua allele at any of these positions, and instead both had that of the main lineage of Siberian bison. These results indicate that both F-1297 and F-4172 are part of the main lineage of Siberian *B. priscus*, and not the Rauchua lineage. We would like to mention, however, that only F-1297 was dual indexed to avoid incorrect read assignment. Base calling of the single-indexed F-4172 library should be interpreted with caution due to a possibility of index hopping (Costello et al., 2018).

DISCUSSION

The results of this study demonstrate that the Rauchua lineage is even more enigmatic than previously thought. Despite extensive sampling in the northern Russian Far East, covering most of the territory where Eurasian steppe bison were present during the Pleistocene - Holocene transition and the Rauchua locality itself, we did not find any new specimens belonging to the Rauchua bison lineage (Figure 4.3). Instead, our results show that the majority of western Beringian *Bison priscus* cluster together

into a clade within the diversity of Holarctic *Bison priscus* mitochondrial genomes, as described previously (Froese et al., 2017; Heintzman et al., 2016; Shapiro et al., 2004).

We currently do not know why the population to which the Rauchua bison belonged is so rare in paleontological record. It is notable that the Rauchua bison itself has a relatively recent radiocarbon date of only 9497 ± 92 years BP. The end of the Pleistocene was marked by megafaunal extinctions across the globe, with most mainland extinctions in Siberia occuring by 14-10 thousand years ago (Koch & Barnosky, 2006; Stuart et al., 2004). The Rauchua bison therefore postdates these extinctions, and is one of the youngest bison thus far dated from the Russian Far East (Markova et al., 2015; Van Geel et al., 2014). It is possible that the Rauchua bison lineage persisted in more southerly locations during the Pleistocene, where bones are either not preserved or have yet to be recovered. The Rauchua lineage may have moved northward as the climate improved after the transition into the Holocene, and established populations, albeit briefly, in the region that was previously occupied by the main lineage of Siberian steppe bison.

Bison is among the most well represented taxon in Late Pleistocene fossil deposits and among the taxa best represented by ancient mitochondrial DNA data (Froese et al., 2017; Heintzman et al., 2016; Palacio et al., 2017; Shapiro et al., 2004; Soubrier et al., 2016; Zazula et al., 2017). This research has revealed several surprises, including the presence of other cryptic bison lineages. For example, (Soubrier et al., 2016) reported a genetically distinct but morphologically indistinguishable bison lineage that may have coexisted with *Bison priscus* in the Southern Ural region during

the Late Pleistocene. This cryptic lineage, which has been referred to as Bb1 or *Bison schoetensacki*, has mitochondrial data that is more similar to European bison, *Bison bonasus,* than it is to the steppe bison *Bison priscus* (Palacio et al., 2017; Soubrier et al., 2016). The discovery and genetic characterization of both Clade X and the Rauchua bison lineage highlight the extensive genetic diversity of this geographically widespread taxon, which further paleontological and genetic research will be necessary to explore. In particular, additional bison samples from southern regions of Eurasia are necessary to shed light on evolution of the *Bison* genus, and on the divergent Rauchua lineage specifically.

Despite an extremely rich paleontological record, hypotheses about *Bison* evolution and taxonomy have been conflicting for a long time, because of the extensive morphological diversity within the genus (Grange et al., 2018). While mitochondrial DNA data can provide new insights into the evolution of this particular group of taxa, mitochondria represent only a single, maternally inherited locus. The evolutionary history revealed by analysis of mitochondrial DNA tells, therefore, only a part of the evolutionary story of the species in question. To fully resolve the evolutionary history of bison, and to fully understand the relationship between the Rauchua lineage and other Siberian steppe bison, analysis of nuclear DNA will be necessary. Fortunately, some of the samples presented in this study are sufficiently well preserved to attempt this work.

The results of our study confirm the existence of the deeply divergent Rauchua clade and shed new light on the evolutionary history of bison during the Pleistocene to

Holocene transition in Siberia. Future analysis of samples processed for this study will reveal when the Rauchua bison lineage diverged from that of the other bison and whether and to what extent the Rauchua bison population admixed with other Late Pleistocene bovids.

ACKNOWLEDGEMENTS

This paper is dedicated to Andrei Sher. The authors are tremendously grateful to Dr. Sher for his commitment to paleontology, Arctic ecology, and to applying cutting-edge tools like genomics to learn more about the ecosystem and animals he loved. BS in particular is thankful for Dr Sher's willingness to mentor and guide her during the early stages of her career, and for his many fascinating and often surprising stories about Pleistocene megafauna, Arctic paleontology, and Joseph and the Amazing Technicolor Dreamcoat. We are also thankful to Beth Nelson, Heather Milne, and Christopher Garrison for technical support and assistance in the laboratory. Alisa Vershinina and Beth Shapiro were supported in part by National Science Foundation grant NSF ANS 1417036. Christian Lowson is supported by the UCSC Kenneth S. Norris Center for Natural History.

BIBLIOGRAPHY

Briggs, A. W., Good, J. M., Green, R. E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Schmitz, R., Doronichev, V. B., Golovanova, L. V., de la Rasilla, M., Fortea, J., Rosas, A., & Pääbo, S. (2009). Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science*, *325*(5938), 318–321.

Costello, M., Fleharty, M., Abreu, J., Farjoun, Y., Ferriera, S., Holmes, L., Granger, B., Green, L., Howd, T., Mason, T., Vicente, G., Dasilva, M., Brodeur, W., DeSmet, T., Dodge, S., Lennon, N. J., & Gabriel, S. (2018). Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics*, *19*(1), 332.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., & Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(39), 15758–15763.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Froese, D., Stiller, M., Heintzman, P. D., Reyes, A. V., Zazula, G. D., Soares, A. E. R., Meyer, M., Hall, E., Jensen, B. J. L., Arnold, L. J., MacPhee, R. D. E., & Shapiro, B. (2017). Fossil and genomic evidence constrains the timing of bison arrival in North America. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(13), 3457–3462.

Grange, T., Brugal, J.-P., Flori, L., Gautier, M., Uzunidis, A., & Geigl, E.-M. (2018). The Evolution and Population Diversity of Bison in Pleistocene and Holocene Eurasia: Sex Matters. *Diversity*, *10*(3), 65.

Heintzman, P. D., Froese, D., Ives, J. W., Soares, A. E. R., Zazula, G. D., Letts, B., Andrews, T. D., Driver, J. C., Hall, E., Hare, P. G., Jass, C. N., MacKay, G., Southon, J. R., Stiller, M., Woywitka, R., Suchard, M. A., & Shapiro, B. (2016). Bison phylogeography constrains dispersal and viability of the Ice Free Corridor in western Canada. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 8057–8063.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* , *29*(13), 1682–1684.

Kirillova, I. V., Zanina, O. G., Chernova, O. F., Lapteva, E. G., Trofimova, S. S., Lebedev, V. S., Tiunov, A. V., Soares, A. E. R., Shidlovskiy, F. K., & Shapiro, B. (2015). An ancient bison from the mouth of the Rauchua River (Chukotka, Russia). *Quaternary Research*, *84*(2), 232–245.

Koch, P. L., & Barnosky, A. D. (2006). Late Quaternary Extinctions: State of the Debate. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 215–250.

Korlević, P., Gerber, T., Gansauge, M.-T., Hajdinjak, M., Nagel, S., Aximu-Petri, A., & Meyer, M. (2015). Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *BioTechniques*, *59*(2), 87–93.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760.

Luo, A., Zhang, A., Ho, S. Y., Xu, W., Zhang, Y., Shi, W., Cameron, S. L., & Zhu, C. (2011). Potential efficacy of mitochondrial genes for animal DNA barcoding: a case study using eutherian mammals. *BMC Genomics*, *12*, 84.

Markova, A. K., Puzachenko, A. Y., van Kolfschoten, T., Kosintsev, P. A., Kuznetsova, T. V., Tikhonov, A. N., Bachura, O. P., Ponomarev, D. V., van der Plicht, J., & Kuitems, M. (2015). Changes in the Eurasian distribution of the musk ox (Ovibos moschatus) and the extinct bison (Bison priscus) during the last 50 ka BP. *Quaternary International: The Journal of the International Union for Quaternary Research*, *378*, 99–110.

Maschenko, E. N., Potapova, O. R., Vershinina, A., Shapiro, B., Streletskaya, I. D., Vasiliev, A. A., Oblogov, G. E., Kharlamova, A. S., Potapov, E., van der Plicht, J., Tikhonov, A. N., Serdyuk, N. V., & Tarasenko, K. K. (2017). The Zhenya Mammoth (Mammuthus primigenius (Blum.)): Taphonomy, geology, age, morphology and ancient DNA of a 48,000 year old frozen mummy from western Taimyr, Russia. *Quaternary International: The Journal of the International Union for Quaternary Research*, *445*, 104–134.

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, *2010*(6), db.prot5448.

Palacio, P., Berthonaud, V., Guérin, C., Lambourdière, J., Maksud, F., Philippe, M., Plaire, D., Stafford, T., Marsolier-Kergoat, M.-C., & Elalouf, J.-M. (2017). Genome data on the extinct Bison schoetensacki establish it as a sister

species of the extant European bison (Bison bonasus). *BMC Evolutionary Biology*, *17*(1), 48.

Reimer, P. J., Bard, E., Bayliss, A., Warren Beck, J., Blackwell, P. G., Ramsey, C. B., Buck, C. E., Cheng, H., Lawrence Edwards, R., Friedrich, M., Grootes, P. M., Guilderson, T. P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T. J., Hoffmann, D. L., Hogg, A. G., Hughen, K. A., … van der Plicht, J. (2013). IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon*, *55*(4), 1869–1887.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539–542.

Shapiro, B., Drummond, A. J., Rambaut, A., Wilson, M. C., Matheus, P. E., Sher, A. V., Pybus, O. G., Gilbert, M. T. P., Barnes, I., Binladen, J., Willerslev, E., Hansen, A. J., Baryshnikov, G. F., Burns, J. A., Davydov, S., Driver, J. C., Froese, D. G., Harington, C. R., Keddie, G., … Cooper, A. (2004). Rise and fall of the Beringian steppe bison. *Science*, *306*(5701), 1561–1565.

Soubrier, J., Gower, G., Chen, K., Richards, S. M., Llamas, B., Mitchell, K. J., Ho, S. Y. W., Kosintsev, P., Lee, M. S. Y., Baryshnikov, G., Bollongino, R., Bover, P., Burger, J., Chivall, D., Crégut-Bonnoure, E., Decker, J. E., Doronichev, V. B., Douka, K., Fordham, D. A., … Cooper, A. (2016). Early cave art and ancient DNA record the origin of European bison. *Nature Communications*, *7*, 13158.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* , *30*(9), 1312–1313.

Stuart, A. J., Kosintsev, P. A., Higham, T. F. G., & Lister, A. M. (2004). Pleistocene to Holocene extinction dynamics in giant deer and woolly mammoth. *Nature*, *431*(7009), 684–689.

Van Geel, B., Protopopov, A., Bull, I., Duijm, E., Gill, F., Lammers, Y., Nieman, A., Rudaya, N., Trofimova, S., Tikhonov, A. N., Vos, R., Zhilich, S., & Gravendeel, B. (2014). Multiproxy diet analysis of the last meal of an early Holocene Yakutian bison: DIET ANALYSIS OF AN EARLY HOLOCENE YAKUTIAN BISON. *Journal of Quaternary Science*, *29*(3), 261–268.

Zazula, G. D., Hall, E., Hare, P. G., Thomas, C., Mathewes, R., La Farge, C., Martel, A. L., Heintzman, P. D., & Shapiro, B. (2017). A middle Holocene steppe bison and paleoenvironments from the Versleuce Meadows, Whitehorse, Yukon, Canada. *Canadian Journal of Earth Sciences*, *54*(11), 1138–1152.

Zeyland, J., Wolko, L., Lipiński, D., Woźniak, A., Nowak, A., Szalata, M., Bocianowski, J., & Słomski, R. (2012). Tracking of wisent-bison-yak mitochondrial evolution. *Journal of Applied Genetics*, *53*(3), 317–322.

Table 4.1. Specimens used in the study. Abbreviations: NAS - National Alliance of Shidlovskiy "Ice Age", UCSC - University of California Santa Cruz, endogenous DNA - proportion of non-contaminant *Bison* DNA, mtDNA - mitochondrial genome. For location details please see the Results section and Figure 4.1.

NAS ID	UCSC ID	Locality	Endogenous DNA (%)	mtDNA coverage (x)
F-1267	SC18.CL011	Middle Indigirka River (North-Eastern Yakutia)	0.46	0
F-1286	SC18.CL010	Middle Indigirka River (North-Eastern Yakutia)	0.2	0
F-1287	SC18.CL009	Middle Indigirka River (North-Eastern Yakutia)	0.24	0
F-1288	SC18.CL007	Middle Indigirka River (North-Eastern Yakutia)	0.21	0
F-1291	SC18.CL001	Middle Indigirka River (North-Eastern Yakutia)	1.08	3.3
F-1294	SC18.CL003	Middle Indigirka River (North-Eastern Yakutia)	1.6	0
F-1295	SC18.CL002	Middle Indigirka River (North-Eastern Yakutia)	0.12	0

F-1297	SC18.CL004	Middle Indigirka River (North-Eastern Yakutia)	0.07	1.2
F-2519	SC14.AE008	Maly Anuy River (North-Western Chukotka)	21.01	223
F-3006	SC14.AE009	Maly Anuy River (North-Western Chukotka)	56.11	837
F-3246/9	SC14.PH043	Rauchua River mouth (North-Western Chukotka)	26.4	4
F-3246/12	SC14.PH043	Rauchua River mouth (North-Western Chukotka)	54.99	128
F-3246/13	SC14.PH043	Rauchua River mouth (North-Western Chukotka)	4.74	12
F-3267	SC14.AE010	Maly Anuy River (North-Western Chukotka)	12.45	22
F-3521	SC14.AE014	Ostrovnovskaya tundra (North-Western Chukotka)	6.9	20
F-3524	SC14.AE015	Ostrovnovskaya tundra (North-Western Chukotka)	22.63	133

F-3531	SC14.AE016	Ostrovnovskaya tundra (North-Western Chukotka)	5.42	13
F-3807	SC14.AE012	Maly Anuy River (North-Western Chukotka)	27.12	139
F-3941	SC14.AE013	Maly Anuy River (North-Western Chukotka)	46.13	480
F-3952	SC14.AE011	Maly Anuy River (North-Western Chukotka)	2.34	7
F-4089	SC14.AE023	East Siberian Sea coast (North-Western Chukotka)	15.3	39
F-4090	SC14.AE022	East Siberian Sea coast (North-Western Chukotka)	11.81	72
F-4091	SC14.AE019	East Siberian Sea coast (North-Western Chukotka)	1.73	17
F-4092	SC14.AE020	East Siberian Sea coast (North-Western Chukotka)	1.05	47
F-4095	SC14.AE018	East Siberian Sea coast (North-Western Chukotka)	9.3	60

F-4171	SC19.CL001	Rauchua River mouth (North-Western Chukotka)	38.28	6
F-4172	SC19.CL002	Rauchua River mouth (North-Western Chukotka)	1.5	1.8
F-4173	SC19.CL003	Rauchua River mouth (North-Western Chukotka)	1.37	0.15
F-4176	SC19.CL004	Rauchua River mouth (North-Western Chukotka)	0.59	0.55

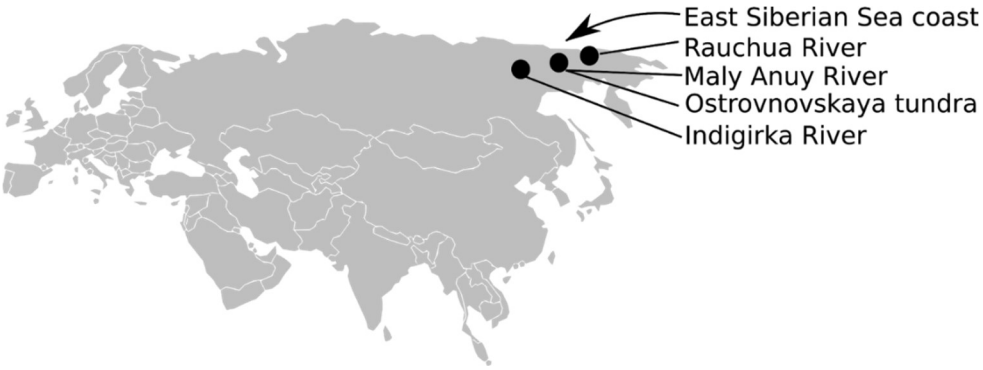


Figure 4.1. The map of Eurasia with sample locations marked with a circle: Middle Indigirka River (North-Eastern Yakutia), Bilibinsky region (North-Western Chukotka), Rauchua River (North-Western Chukotka).

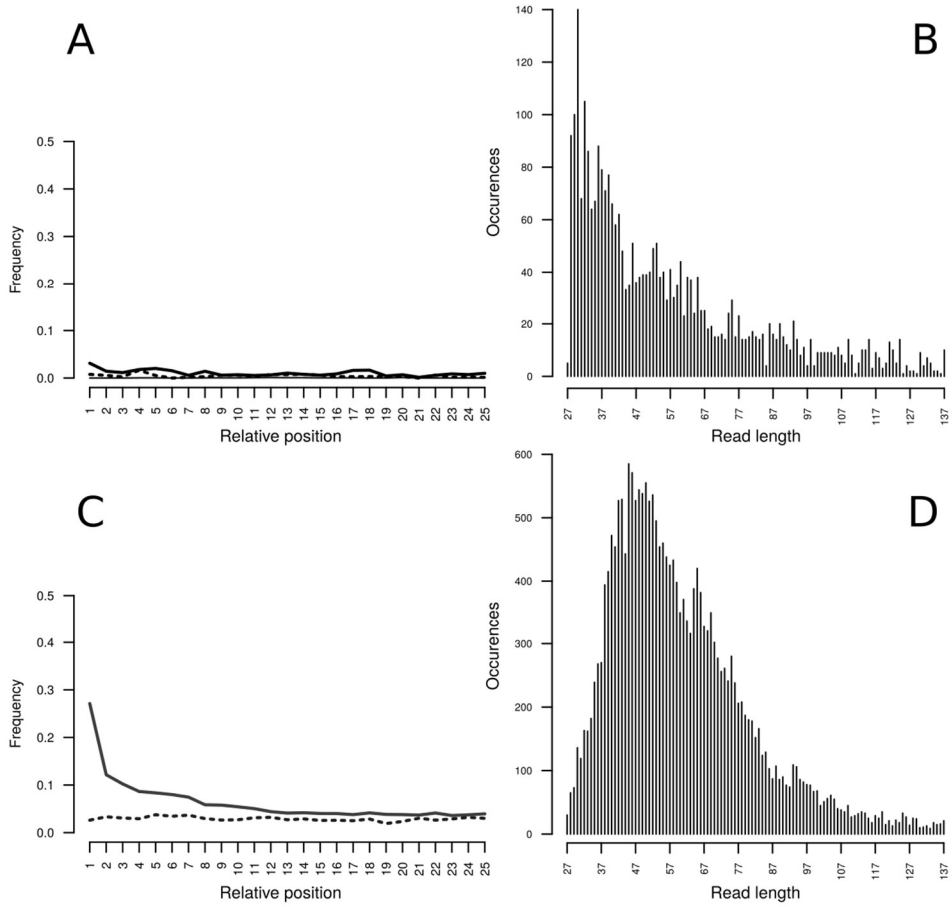


Figure 4.2. Ancient DNA degradation profile of F-1291 from Middle Indigirka (panels A,B) and F-3246/9, the new subsample of the original Rauchua bison (panel C, D). Deamination profile is shown in the panels A and C. Solid black line - frequency of C to T substitutions at 5` ends of DNA molecules. Numbers on the x-axis are relative nucleotide positions on each DNA fragment mapped to the Bison reference nuclear genome. Dotted line - frequency of G to A substitutions. Fragment length distribution is shown on panels B and D.

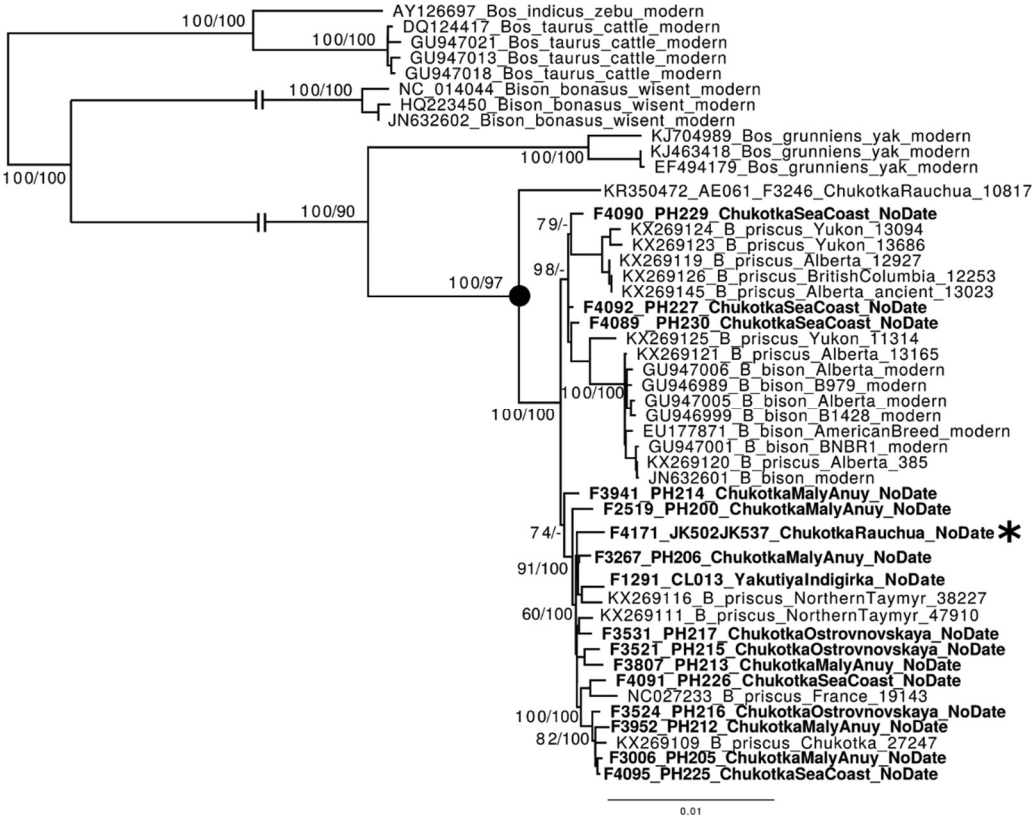


Figure 4.3. Maximum likelihood phylogenetic tree reconstructed on complete mitochondrial genomes of various bovids, including extinct and extant *Bison* species. Support for major divergences is marked below or above the branches: bootstrap calculated with RAxML, and posterior probability (in percent, calculated with MrBayes). Both numbers are shown in cases where corresponding node was supported by both phylogeny reconstruction methods. Divergence of Rauchua lineage is shown with a circle. The new specimen from Rauchua locality is marked with a star. Mitochondrial genomes assembled in this study are in bold. All previously published sequences are shown with corresponding GenBank ID numbers. Numbers at the end of sequence names are median IntCal13-calibrated radiocarbon ages (Reimer et al., 2013).

SYNTHESIS

Since the publishing of the DNA extracted from quagga in 1987, paleogenetics, a science of recovery of DNA from archival, museum, paleontological, or otherwise degraded tissue samples (ancient DNA, aDNA) has dramatically changed (Higuchi et al. 1984). Moving into the era of high throughput next-generation sequencing opened the door into relatively inexpensive and fast specimen processing, and aDNA field blooms on these recent technological advantages (Brunson and Reich 2019). In my dissertation, I utilized aDNA techniques to examine the interplay between the evolutionary history of megafauna species and timing and pattern of major environmental shifts of the Northern Hemisphere during the last five million years.

All results of my dissertation are based on the analysis of DNA extracted from specimens stored in natural history museum collections. In Chapter 1 I demonstrated the benefit of aDNA methods in species identification and provided a pipeline together with a set of guidelines on how paleogenomics can utilize museum-stored specimens for biogeographical research and eco-evolutionary analyses. I also demonstrated that paleogenetic studies are most powerful when used in combination with other methods, including morphological analysis, radiocarbon dating, stable isotope analysis, and other techniques. The following chapters 2, 3, and 4 further support this conclusion. Additionally, based on my experience while conducting the research for this dissertation, I acknowledged that some of the aDNA approaches are inherently destructive with respect to the sampled material. Thus in Chapter 1, I took an opportunity to suggest how researchers together with museum curators should consider

the long-term impact of aDNA studies on the museum collections when deciding what techniques are best for any particular sample. Future developments in the growing diversity of DNA processing techniques should be focused both on advancing our understanding of evolutionary history, paleoecology, and conservation, as well as on minimizing invasive and destructive sampling of museum specimens.

In chapters 1, 2 and 4 I studied the Late Pleistocene, the period of the last 50,000 years that was marked by extremely high population sizes of equids and other megafauna species, such as the steppe bison *Bison priscus*. Their ubiquitous presence in museum collections and the applicability of the radiocarbon dating to this period makes extensive evolutionary inferences possible. It is harder, however, to track the diversification and phylogeography of more archaic populations that are often not preserved in paleontological record or are represented by specimens with irretrievable DNA. I overcame this obstacle in Chapter 3 by targeting parts of the bone fossil remains with a high DNA preservation potential for DNA extraction as well as utilizing a new single-stranded DNA library protocol (Troll et al. 2019; Pinhasi et al. 2015). The outlined approach allowed the recovery of the ~700,000 year-old equid nuclear genome with a 12-fold coverage, making it the oldest multi-fold coverage paleogenome known to date. My results suggest that genome-wide studies of Early and Middle Pleistocene fossil remains are possible and paleogenomics should be utilized for tracking large-scale biodiversity changes.

The common theme across all my dissertation chapters is utilizing present-day advantages of the paleogenomic field to look into the past epochs for insights on a

correlation between extinction patterns and environmental change. I was particularly interested in understanding the role of population fragmentation in the reduction of gene flow between populations and decreasing genetic diversity. Reduced gene flow is generally thought to promote speciation via genetic differentiation. On the other hand, gene flow can have antagonistic effects on adaptive potential through the influx of deleterious alleles and the formation of new genetic combinations (Slatkin 1987). Furthermore, complete absence of admixture between populations and shrinking population size increase genetic drift, which in turn puts long-lived species in peril due to their limited adaptability in a face of rapidly changing climate (Stuart et al. 2004; Rogers and Slatkin 2017). Many questions about how population fragmentation will affect biodiversity remain open. Will habitat change increase rates of local extinction and endanger biodiversity? How important is connectivity between populations in maintaining genetic diversity and mitigating extinction risk? What is the role of geographic barriers and their permeability in population connectivity? Present-day habitat fragmentation and environmental change are not exceptional and in chapters 2, 3 and 4 I used environmental changes of the past to shed light on the outlined questions.

Building on the methodology described in Chapter 1, in chapters 2 and 3 I made an in-depth exploration of the biogeography of equids, the genus *Equus*, a group that served as a model for evolutionary studies for decades (MacFadden 2005). First, I examined the role of Beringia in cross-continental dispersals and migrations of caballoid equids, including populations close to the present-day *E. ferus*, during the Pleistocene. Ice Age Beringia was a large landmass that facilitated various biotic

transitions of the Late Pleistocene by linking Eurasia and North America (Elias and Crocker 2008; Zimov et al. 1995; Dale Guthrie 2001; Guthrie 1982). When opened, the Bering Land Bridge played a key role as a Quaternary dispersal corridor for the steppe bison, mammoth, and humans, whose populations migrated from Eastern Siberia into the New World and for equids migrating in the opposite Westward direction (Debruyne et al. 2008; Shapiro et al. 2004; Meiri et al. 2014; Weinstock et al. 2005). I found evidence of gene flow as well as a signal of genetic diversity in this area which demonstrates that Beringia acted as a critical contact zone for various caballoid horse populations. The productive ecosystem of Late Pleistocene Beringia, the Mammoth Steppe biome, likely allowed caballoid horses to expand into the Old World (Zimov et al. 2012). Furthermore, the unique biome of the Mammoth Steppe located in the dispersal corridor area served as a conduit to gene flow between the continents. Disappearance of the Bering Land Bridge together with the transformation of the Mammoth Steppe into the less productive shrub-tundra during the Pleistocene-Holocene transition likely exacerbated the decline of caballoid horses and other megafauna species. The future research should explore to what degree will global climatic change and human impacts further transform Arctic ecosystems.

The extent of admixture between caballoid horse populations, as well as between other megafauna species inhabiting Beringia remains unclear. Due to the limited number of nuclear genomes analyzed in Chapter 2, I was unable to test whether hybridization between local caballoid horse populations stretched beyond the area of the Bering Land Bridge. Thus, it is still unknown whether a North American admixture

signal exists in European Late Pleistocene horses as well as whether Eurasian admixture signal is present in ancient caballoid horses inhabiting south of the North American Cordilleran-Laurentide ice sheets. One hypothesis explaining patterns of genetic relatedness discovered in Chapter 2 is the coexistence of a Late Pleistocene meta-population of caballoid horses that stretched from present-day Europe to Chukotka and further to Yukon and central North America. An alternative hypothesis is a genetic isolation between two major (North American and Eurasian) caballoid horse lineages which coexisted in sympatry through niche partitioning. The outlined hypotheses and the circumpolar-wide signal of admixture could be resolved with single nucleotide polymorphism (SNP) data recovered from fossil caballoid horse specimens sampled across the Holarctic. The scope of the Bering Land Bridge project was limited to the analysis of a few high coverage nuclear genomes supplemented by a wide-scale mitochondrial phylogeny, thus I did not recover population-scale SNP data for this chapter. Future analysis should be focused on exploring the extent of gene flow between caballoid horse lineages using population genetic methods on SNP markers. The panel of > 100 specimens processed for aDNA in Chapter 2 makes such research possible. Furthermore, studies of critical contact zones and dispersal corridors similar to one presented in Chapter 2, should be utilized to improve conservation strategies.

Paleontological studies, including aDNA ones, have shown that past extinctions appear to result from multiple causes. Two of the most important factors are loss of genetic diversity associated with habitat destruction and decreased population continuity (Lorenzen et al. 2011; Koch and Barnosky 2006; MacPhee and Greenwood

2013; C. N. Johnson 2002). Population fragmentations and resulting from it increased genetic drift are the possible mechanisms underlying local extirpations of high latitude populations (Rogers and Slatkin 2017; Shapiro et al. 2004). In Chapter 2 I found that critical dispersal corridors, such as the Bering Land Bridge, maintained genetic diversity among equids concurrently with their demographic decline in the Pleistocene. Present-day Arctic mammals, such as caribou and arctic fox, are currently experiencing a decrease of available habitat and increase in its patchiness (Weir and Schluter 2007; Mann et al. 2015; Berteaux et al. 2004; Dalén et al. 2007; C. J. Johnson, Ehlers, and Seip 2015). In light of the current sixth mass extinction, studies revealing the role of population fragmentation in increasing extinction risk are bringing critical insights that should be used to aid future conservation strategies.

In Chapter 3 I investigated another prominent common theme of equid evolutionary research - the convoluted taxonomy and diversification of the genus *Equus* (Barrón-Ortiz et al. 2019). Despite the incredible richness of forms and good fossil preservation, the evolutionary history of the genus has been confusing due to morphological variation, debatable species assignments, and taxonomic oversplitting (Azzaroli and Voorhies 1993; Eisenmann 2004; Heintzman et al. 2017). Due to these factors, genomic data and aDNA studies have been essential for understanding the temporal and phylogenetic relationships between equids (Orlando 2019). I contributed to this understanding in Chapter 3, where I sequenced and analyzed an Early-Middle Pleistocene equid from Paradise Hill mining site, Klondike, Canada's Yukon Territory. The specimen finding is associated with the Gold Run volcanic ash layer dating the

specimen to be older than ~700,000 years (Westgate, Preece, and Jackson 2011). My genetic analysis has revealed that the Paradise Hill equid belongs to stenonid horses - a clade encompassing present-day zebras, Asiatic wild asses, and kiangs. Furthermore, when the population of the Paradise Hill individual was present in the Early-Middle Pleistocene, it coexisted with the more divergent caballoid horse group in Klondike suggesting high megafauna species diversity in the Early Middle Pleistocene Yukon. Interestingly, subsequently to the Middle Pleistocene, there have been no more findings of equids in Klondike that would belong to the stenonid group. The latter suggests a species turnover, during which stenonid horse populations disappeared from Beringian high latitude locations, and those became dominated by caballoid horses up until 11 thousand years ago. Based on these findings, I suggest to focus the future research of ancient Yukon paleoenvironments in two directions: (1) identifying the local ecological factors that underlie megafauna species turnover and (2) revealing how these factors impacted formation of the present-day Arctic fauna.

Lastly, one common misconception about equids, which turns out to be detrimental for feral horse management in the Western United States, is the number of species in the genus *Equus*. Numerous species of caballoid horses were described using subtle differences in bone morphology without any evidence of the putative "species" to be a real biological entity (Kefena et al. 2012). As a result, domestic horses, including their feral populations, and extinct North American caballoid hoses are seen as two different species (Fazio 1997). It is on this basis that mustangs are considered to be invasive in North America and aggressive management strategies are implemented to

sustain their populations in the US (The Bureau of Land Management 2018). However, as I described above as well as in chapters 1, 2, and 3, fossil remains are often problematic for morphological interpretations, and taxonomical oversplitting is common in equid paleontology. Furthermore, caballoid horses as well as their Pleistocene ancestors have been a component of North American fauna for several million years, until their (recent in evolutionary timescale) extinction on the continent 11 thousand years ago (MacFadden 1994; Guthrie 2003). The results of my dissertation supported this notion and indicated that caballoid equids, regardless of their species identification, are a native component of the North American ecosystem. Thus, the status of feral horses roaming in the Western US should be changed from an invasive animal to a reintroduced megafauna species. I suggest modifying current wild horse management strategies into a comprehensive approach that takes the evolutionary and ecological data on caballoid horses into account.

BIBLIOGRAPHY

Azzaroli, A., and M. R. Voorhies. 1993. "The Genus Equus in North America. The Blancan Species." *Palaeontographia Italica* 80: 175–98.

Barrón-Ortiz, Christina I., Leonardo S. Avilla, Christopher N. Jass, Víctor M. Bravo-Cuevas, Helena Machado, and Dimila Mothé. 2019. "What Is Equus? Reconciling Taxonomy and Phylogenetic Analyses." *Frontiers in Ecology and Evolution* 7: 343.

Berteaux, Dominique, Denis Réale, Andrew G. McAdam, and Stan Boutin. 2004. "Keeping Pace with Fast Climate Change: Can Arctic Life Count on Evolution?" *Integrative and Comparative Biology* 44 (2): 140–51.

Brunson, Katherine, and David Reich. 2019. "The Promise of Paleogenomics Beyond Our Own Species." *Trends in Genetics: TIG* 35 (5): 319–29.

Dale Guthrie, R. 2001. "Origin and Causes of the Mammoth Steppe: A Story of Cloud Cover, Woolly Mammal Tooth Pits, Buckles, and inside-out Beringia." *Quaternary Science Reviews* 20 (1): 549–74.

Dalén, Love, Veronica Nyström, Cristina Valdiosera, Mietje Germonpré, Mikhail Sablin, Elaine Turner, Anders Angerbjörn, Juan Luis Arsuaga, and Anders Götherström. 2007. "Ancient DNA Reveals Lack of Postglacial Habitat Tracking in the Arctic Fox." *Proceedings of the National Academy of Sciences of the United States of America* 104 (16): 6726–29.

Debruyne, Regis, Genevieve Chu, Christine E. King, Kirsti Bos, Melanie Kuch, Carsten Schwarz, Paul Szpak, et al. 2008. "Out of America: Ancient DNA Evidence for a New World Origin of Late Quaternary Woolly Mammoths." *Current Biology: CB* 18 (17): 1320–26.

Eisenmann, V. 2004. "Equus: An Evolution without Lineages." *Terra Nostra*. https://www.academia.edu/download/50967163/Equus_an_evolution_without_lineages__20161219-24452-2uqqh4.pdf.

Elias, Scott A., and Barnaby Crocker. 2008. "The Bering Land Bridge: A Moisture Barrier to the Dispersal of Steppe–tundra Biota?" *Quaternary Science Reviews* 27 (27): 2473–83.

Fazio, Patricia Mabee. 1997. "The Fight to Save a Memory: Creation of the Pryor Mountain Wild Horse Range." *Annals of Wyoming* 69: 28.

Guthrie, R. Dale. 1982. "Mammals of the Mammoth Steppe as Paleoenvironmental Indicators." In *Paleoecology of Beringia*, edited by David M. Hopkins, John V. Matthews, Charles E. Schweger, and Steven B. Young, 307–26. Academic Press.

———. 2003. “Rapid Body Size Decline in Alaskan Pleistocene Horses before Extinction.” *Nature* 426 (6963): 169–71.

Heintzman, Peter D., Grant D. Zazula, Ross D. E. MacPhee, Eric Scott, James A. Cahill, Brianna K. McHorse, Joshua D. Kapp, et al. 2017. “A New Genus of Horse from Pleistocene North America.” *eLife* 6 (November). https://doi.org/10.7554/eLife.29944.

Higuchi, R., B. Bowman, M. Freiberger, O. A. Ryder, and A. C. Wilson. 1984. “DNA Sequences from the Quagga, an Extinct Member of the Horse Family.” *Nature* 312 (5991): 282–84.

Johnson, Chris J., Libby P. W. Ehlers, and Dale R. Seip. 2015. “Witnessing Extinction--Cumulative Impacts across Landscapes and the Future Loss of an Evolutionarily Significant Unit of Woodland Caribou in Canada.” *Biological Conservation* 186: 176–86.

Johnson, Chrisopher N. 2002. “Determinants of Loss of Mammal Species during the Late Quaternary ‘megafauna’ extinctions: Life History and Ecology, but Not Body Size.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269 (1506): 2221–27.

Kefena, E., Y. Mekasha, J. L. Han, S. Rosenbom, A. Haile, T. Dessie, and A. Beja-Pereira. 2012. “Discordances between Morphological Systematics and Molecular Taxonomy in the Stem Line of Equids: A Review of the Case of Taxonomy of Genus Equus.” *Livestock Science* 143 (2): 105–15.

Koch, Paul L., and Anthony D. Barnosky. 2006. “Late Quaternary Extinctions: State of the Debate.” *Annual Review of Ecology, Evolution, and Systematics* 37 (1): 215–50.

Lorenzen, Eline D., David Nogués-Bravo, Ludovic Orlando, Jaco Weinstock, Jonas Binladen, Katharine A. Marske, Andrew Ugan, et al. 2011. “Species-Specific Responses of Late Quaternary Megafauna to Climate and Humans.” *Nature* 479 (7373): 359–64.

MacFadden, Bruce J. 1994. *Fossil Horses: Systematics, Paleobiology, and Evolution of the Family Equidae*. Cambridge University Press.

———. 2005. “Fossil Horses - Evidence for Evolution.” *Science* 307 (5716): 1728–30.

MacPhee, Ross D. E., and Alex D. Greenwood. 2013. “Infectious Disease, Endangerment, and Extinction.” *International Journal of Evolutionary Biology* 2013 (January): 571939.

Mann, Daniel H., Pamela Groves, Richard E. Reanier, Benjamin V. Gaglioti, Michael L. Kunz, and Beth Shapiro. 2015. "Life and Extinction of Megafauna in the Ice-Age Arctic." *Proceedings of the National Academy of Sciences of the United States of America* 112 (46): 14301–6.

Meiri, Meirav, Adrian M. Lister, Matthew J. Collins, Noreen Tuross, Ted Goebel, Simon Blockley, Grant D. Zazula, et al. 2014. "Faunal Record Identifies Bering Isthmus Conditions as Constraint to End-Pleistocene Migration to the New World." *Proceedings. Biological Sciences / The Royal Society* 281 (1776): 20132167.

Orlando, Ludovic. 2019. "An Ancient DNA Perspective on Horse Evolution." In *Paleogenomics: Genome-Scale Analysis of Ancient DNA*, edited by Charlotte Lindqvist and Om P. Rajora, 325–51. Cham: Springer International Publishing.

Pinhasi, Ron, Daniel Fernandes, Kendra Sirak, Mario Novak, Sarah Connell, Songül Alpaslan-Roodenberg, Fokke Gerritsen, et al. 2015. "Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone." *PloS One* 10 (6): e0129102.

Rogers, Rebekah L., and Montgomery Slatkin. 2017. "Excess of Genomic Defects in a Woolly Mammoth on Wrangel Island." *PLoS Genetics* 13 (3): e1006601.

Shapiro, Beth, Alexei J. Drummond, Andrew Rambaut, Michael C. Wilson, Paul E. Matheus, Andrei V. Sher, Oliver G. Pybus, et al. 2004. "Rise and Fall of the Beringian Steppe Bison." *Science* 306 (5701): 1561–65.

Slatkin, M. 1987. "Gene Flow and the Geographic Structure of Natural Populations." *Science* 236 (4803): 787–92.

Stuart, A. J., P. A. Kosintsev, T. F. G. Higham, and A. M. Lister. 2004. "Pleistocene to Holocene Extinction Dynamics in Giant Deer and Woolly Mammoth." *Nature* 431 (7009): 684–89.

The Bureau of Land Management. 2018. "Report to Congress: Management Options for a Sustainable Wild Horse and Burro Program." https://www.blm.gov/sites/blm.gov/files/wildhorse_2018ReporttoCongress.pdf.

Troll, Christopher J., Joshua Kapp, Varsha Rao, Kelly M. Harkins, Charles Cole, Colin Naughton, Jessica M. Morgan, Beth Shapiro, and Richard E. Green. 2019. "A Ligation-Based Single-Stranded Library Preparation Method to Analyze Cell-Free DNA and Synthetic Oligos." *BMC Genomics* 20 (1): 1023.

Weinstock, Jaco, Eske Willerslev, Andrei Sher, Wenfei Tong, Simon Y. W. Ho, Dan Rubenstein, John Storer, et al. 2005. "Evolution, Systematics, and Phylogeography of Pleistocene Horses in the New World: A Molecular Perspective." *PLoS Biology* 3 (8): e241.

Weir, Jason T., and Dolph Schluter. 2007. “The Latitudinal Gradient in Recent Speciation and Extinction Rates of Birds and Mammals.” *Science* 315 (5818): 1574–76.

Westgate, J. A., S. J. Preece, and L. E. Jackson. 2011. “Revision of the Tephrostratigraphy of the Lower Sixtymile River Area, Yukon Territory, Canada.” *Canadian Journal of Earth Sciences* 48 (3): 695–701.

Zimov, S. A., V. I. Chuprynin, A. P. Oreshko, F. S. Chapin, J. F. Reynolds, and M. C. Chapin. 1995. “Steppe-Tundra Transition: A Herbivore-Driven Biome Shift at the End of the Pleistocene.” *The American Naturalist* 146 (5): 765–94.

Zimov, S. A., N. S. Zimov, A. N. Tikhonov, and F. S. Chapin. 2012. “Mammoth Steppe: A High-Productivity Phenomenon.” *Quaternary Science Reviews* 57 (December): 26–45.