MAMMOTHS, MASTODONS, AND CHRONOSPATIAL WARMING

MAMMOTHS, MASTODONS, AND CHRONOSPATIAL WARMING: EVOLUTIONARY ANALYSES OF PLEISTOCENE PROBOSCIDEANS FROM TEMPERATE AND TROPICAL LOCALES

By EMIL KARPINSKI, H.B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy

DOCTOR OF PHILOSOPHY (2021) McMaster University
(Biology) Hamilton, Ontario, Canada

TITLE: Mammoths, mastodons, and chronospatial warming: evolutionary analyses of Pleistocene proboscideans from temperate and tropical locales

AUTHOR: Emil Karpinski, Hons. B.Sc. (McMaster University)

SUPERVISOR: Professor Hendrik Nicholas Poinar

NUMBER OF PAGES: xi, 103

Lay Abstract

Pleistocene North America was a time period of immense climatic turbidity, with temperature swings greater than 15°C in response to the expansion and contraction of continental ice-sheets. Despite these massive swings in temperature, many species managed to thrive on the continent and adapt to glacial-associated ecosystem restructuring. Ancient DNA from Pleistocene megafauna can serve as a very useful tool to answer many questions about the distribution of megafaunal species, and how they may have responded to these climatic events. However, most studies have largely focused on species adapted to cold environments and from the last fifty thousand years. In this thesis I extend our knowledge of the genetic landscape of Pleistocene proboscideans, characterizing the mammoth inhabitants of Bechan Cave, Utah, and producing the first look at American mastodon diversity through space and time. This work increases our representation of warm-adapted specimens and characterizes the effects of glacial cycles on megafauna populations.

Abstract

The Quaternary (the approximately the last 2.6 million years) of North America is a tremendously exciting time period to study with respect to ecology. It saw periods of immense climatic turbidity - the expansion and retreat of continental ice-sheets and large swings in temperature, resulting in the wide scale restructuring of terrestrial ecosystem. It also saw widespread migrations of many species and out of Eurasia, mostly notably of modern humans. Ancient DNA offers powerful tools to examine the relationships and responses of megafunal species to these events, but has largely focused on cold-adapted species, and within radiocarbon-time (*i.e.* the last 50 thousand years). In this thesis I work to expand our understanding of the genetic landscape of Pleistocene megafuna in three ways. First, I describe the analysis of coprolites from Bechan Cave, Utah and characterize the mammoth inhabitants in the broader context of North American mammoths. Second, I characterize the diversity of American mastodons across the continent and through time, showing that their range likely repeatedly expanded and contracted in response to Pleistocene glaciations. Lastly, I begin to fill in some of the gaps in the American mastodon dataset from chapter 3, and begin to address some of the taxonomic and biogeographic questions about American and Pacific mastodons in Idaho. Understanding how North American megafauna responded to these climatic and anthropogenic stresses may help to explain why so many species went extinct at the end of the last glaciation, and how species may respond to present day warming. However, it is important to include taxa from warmer locales and environments to ensure our models and hypotheses are comprehensive.

Acknowledgements

To start I would like to thank my supervisor, Hendrik Poinar, who took in a very eager undergrad student in 2014. Thank you so much for this opportunity to learn from you and everyone else in the lab over the last seven years. You guys have become very close friends. I know you joke about my complaining, but know it is only because I care – and I think we can both agree has been worth it most of the time. ☺

Brian Golding and Ben Evans, thank you both so much for your continuous support and pushing me through this PhD. The end result would have been much worse without your help, and I am truly luckily to have had you as members of my committee and to get to know you over the last few years.

I also wanted to extend a special thank you to all the collaborators I have worked with over the last few years. I consider myself truly fortunate to have worked along such great people, and that you entrusted me with the specimens in your care. These projects could not have happened without your trust.

All members of the Ancient DNA Centre are wonderful people, and I'm sure will go very far in life, but I also wanted to take this opportunity to mention a few people explicitly. Firstly, Jacob and Alison Enk, thank you both for sparking my interest in the field and showing me the ropes.

Thank you Melanie Kuch for your help and guidance over these last few years. Some of the most fun parts of this PhD have been the times sitting in your office solving problems and talking through borderline ridiculous hypotheticals! ☺

Ana Duggan, thank you so much for all you taught me and your help on my projects. And thank you for pushing back – I know the arguments sometimes got a little heated but I think they were a vital part of the decision making process and some of the most challenging and thought provoking debates I've been a part of! ☺

Katherine Eaton and Jessica Hider, you guys are the some of the nicest and most empathetic people I have had the pleasure of ever meeting! Always there to listen to any frustration and just chat during a long day.

Matthew Emery and Tyler Murchie, you guys are the source of much personal growth and change over the last few years. You are both wonderful people and I consider myself very fortunate to have made such great friends!

Dirk Hackenberger, buddy, you are easily the best student I have ever had the pleasure of teaching and your breadth of knowledge in biochemistry is a little scary! I am proud to call a friend, and know some day you will develop some protocol that will make you rich and famous. ☺

Thank you also to all members past and present of the Golding-Evans lab group and later of BEAP, for helping me think through things and getting really good at presentations. You guys are all insanely clever and I know you'll go far in life.

I also wanted to extend a special thank you to the staff of the Biology and Anthropology departments, but in particular to Alison Cowie, Mihaela Georgescu, Barb Reuter, and Bonnie Kahlon for putting up with me and all their help and advice these last few years. They were also available to bounce some teaching or project ideas off, help me navigate the ridiculously convoluted nature in which administrative documents had to be submitted, or just to chat.

Thank you also to the BGSS Games Night and Anthro Friday Night Delights groups. During this, seemingly at times never-ending covid lockdown, you guys have been something to look forward to at the end of the week, and have no doubt kept me from going more insane.

Thank you also to the Anime Night crew. You're all fantastic and I'm truly lucky to have known you guys for so long! Not being able to see you guys regularly has been one of the hardest parts of the pandemic, but I'm happy we've at least got the chance to play together and stay in touch online. ☺

Lastly, I wanted to thank my family. Monika, Janusz, and Basia Karpinski, thank you for your support and love. To my grandparents, Barbara and Cezary Kowalski, and Bronisława Porzezińska, thank you all for the roles you played in raising me as well as your never ending support and encouragement. I love you all!

Table of Contents

List of Figures and Tables

List of Abbreviations and Symbols

^{14}C	Carbon-14 / radiocarbon
AMS	Accelerator mass spectrometry
aDNA	Ancient DNA
bp	base pairs
yr	years
BP	Before present
rpm	Rotations per minute
kDa	Kilodalton
vX.X.X	Version
VNTR	Variable number tandem repeat region of the mitochondrial genome
AICc	Corrected Akaike information criterion
FLD	Fragment length distribution
ky/kya	Thousand years / thousand years ago
my/mya	Million years / million years ago
Ma	Mega-annum (a million years)
MIS	Marine isotope stage
95% HPD	Highest posterior density interval containing 95% of the samples
π	Nucleotide diversity
SD	Standard deviation
UDG	Uracil-DNA glycosylase
NCBI	National center for Biotechnology Information
SRA	Sequence Read Archive
^{18}O	Oxygen-18
BLAST	Basic local alignment search tool
σ	Standard deviation
PS / SS	Path sampling / stepping-stone sampling
GSS	Generalized stepping-stone sampling
MLE	Log marginal likelihood
AFR	American Falls Reservoir

Declaration of Academic Achievement

This thesis contains five chapters: an introduction to the thesis (Chapter 1), three chapters describing published works or works in preparation for publication (Chapters 2-4), and a conclusion summarizing the main findings of the thesis and possible future avenues of exploration (Chapter 5). I have written Chapters 1 and 5 in their entirety. Chapters 2, 3, and 4 were produced in collaboration with many other researchers, with myself as lead author. Mine and the contributions of each co-author are outlined in the acknowledgments section within each of those chapters.

Science is made up of so many things that appear obvious after they are explained.

– Frank Herbert, *Dune*

Chapter 1: Introduction

The Quaternary (approximately the last 2.6 million years) was a period of immense climatic turbidity. This is especially true as the world transitioned from the Early to Middle Pleistocene (~770 kya; though the transition spanned from 1.4 – 0.4 mya)[1,2], when glaciations decrease in frequency (from approximately a 41 ky cycle to a 100 ky cycle), but significantly increase in magnitude[3,4]. For continental North America, this mid-Pleistocene change in climatic periodicity manifests in the cyclic growth and retreat of two ice sheets: the Laurentide Ice Sheet, centered over present-day Hudson Bay; and the Cordilleran Ice Sheet, which spread out from the Rocky Mountains[5].

The expansion and retreat of these ice sheets was coupled with large-scale temperature fluctuations (in excess of -15 °C from the last glacial maximum to today in parts of North America)[6,7], as well as changes in sea-levels and climate variability[7,8]. Accordingly, as these abiotic factors changed between glacial and interglacial periods, the ecosystems they supported varied, resulting in the expansion and contraction of many plant and animal species[9,10]. While the ebb and flow of species and environments are still an area of active research (see Chapter 3 which addresses this question in American mastodons), these environmental changes have been implicated in the intra and intercontinental spread of megafauna that would become key constituents of modern landscapes. This includes the migration of horses into Eurasia[11], bison into North America[12], and perhaps most famously, of anatomically modern humans out of Africa and into the rest of the world[13]. These migrations were not only restricted to extant species, but included many now extinct species such as mammoths[14,15].

The capability of many of these species to adapt to changing conditions, and sometimes even cross into new continents, raises questions about why such a large percentage of them went extinct at the Pleistocene-Holocene transition (~11.7 kya)[16,17]. These extinctions happened relatively synchronously (within the last 50-60 ky) and were believed to have primarily affected megafauna (animals with mean body size >44 kg). Additionally, while this extinction event claimed megafauna from across the globe, genera from Australia and the Americas were particularly hard hit[16,17].

While not the focus of research addressed in this thesis (and it likely being too early to answer these questions with any certainty), it is worth quickly summarizing the main proposed hypotheses often put forth to explain the rapid disappearance of megafauna as, regardless of the root cause, both almost certainly played a role in stressing many late Pleistocene populations. Currently most research attempts to quantify the relative roles of two main causes: anthropogenic pressure from humans directly as a new top-level predator or indirectly as ecosystem modulators; or a particularly devastating episode of climatic change[16,17]. A variety of other less favoured causes have been put forth including a mass pandemic[18] and meteor impacts[19,20], the latter of which has been

gaining some support recently with the discovery of craters which possibly date to around the time of extinction[21,22].

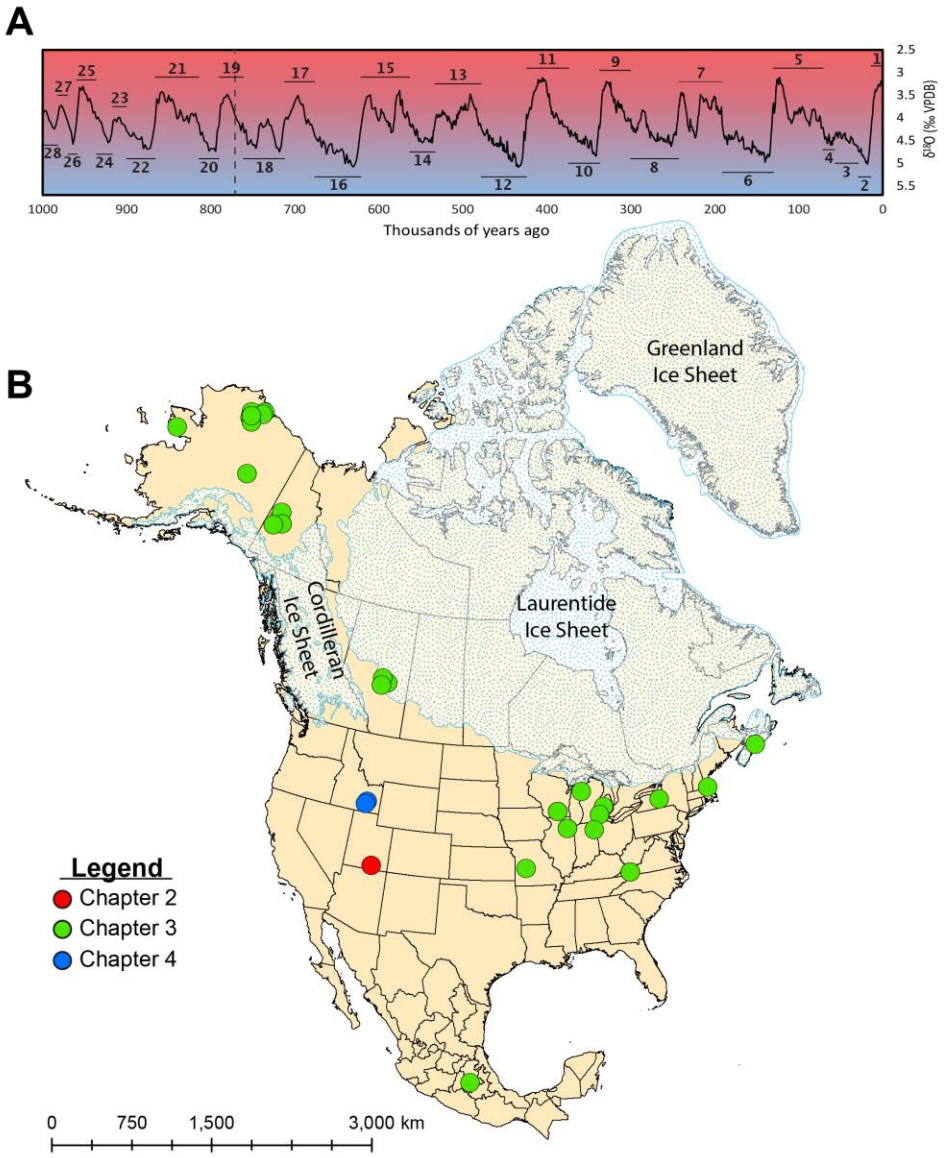


Figure 1: Summary of glacial cycles and samples analyzed in this study. a, Global stack of benthic foraminifera $\delta^{18}O$ for the last 1 million years, which tracks changes in deep-water temperature and global ice volume. The y-axis has been inverted so that periods of low ice buildup (and higher temperatures – red) are at the top of the graph, and periods of greater ice buildup (and lower temperatures – blue) are at the bottom. Marine Isotope Stage (MIS) extents are indicated with black bars above (interglacials) or below (glacials) the $\delta^{18}O$ record. $\delta^{18}O$ values and MIS terminations can be found in Lisiecki & Raymo[4]. The dashed grey line denotes the approximate age of the Matuyama – Bruhes paleomagnetic polarity boundary which official denotes the start of the Chibanian (formerly Middle) stage of the Pleistocene[2]. **b,** A map of North America showing the two main continental ice sheets, as well as the Greenland ice sheet, at 12 ^{14}C ky BP[5]. The location of successful specimens analyzed in Chapters 2 – 4 are also shown.

Researchers who favour a human-driven cause for late Pleistocene extinctions primarily attribute the loss of megafauna to human hunting[23], though also occasionally to anthropogenic practices of environmental change such as fire-clearing[17]. Perhaps the most influential paper in the human-mediated extinction debate is Paul Martin's Overkill/Blitzkrieg model, which posits that the extinction of megafauna was quickly perpetuated by humans against "naive" megafauna[24]. More recent models usually argue for less rapid or less direct approaches, but still center around the idea that the arrival of humans as a new top-level predator introduced too much stress on megafauna populations and drove them to extinction[25,26]. Proponents of this hypothesis often point to the correlation duration of hominin presence and the extent of megafauna extinction, with Africa and Eurasia both losing less genera then Australia and the Americas, where humans only arrive in the late Pleistocene[23]. Critics of human-driven extinction models usually point to the lack of species associated with kill sites[25,27,28]; in North America for example only about 5 taxa can be confidently linked to human hunting[28]. Additionally, as the overlap in time between human colonization and megafauna extinctions in Australia and the Americas increases, critics question why megafauna appeared to live alongside humans for so long before going extinct[28].

Researchers who attribute extinctions to environmental causes largely argue that the seesawing of climate during the transition out of the last glacial was particularly severe and rapid, such that many species could not easily adapt[16,25]. In the Americas, the effects of the Antarctic Cold Reversal and the Younger-Dryas events, which sharply reversed post-LGM warming trends, are usually implicated as a contributing factor to environmental extinctions[29]. Critics of climate extinction scenarios often question how extinct species managed to survive numerous other glacial/interglacial transitions during the mid-to-late Pleistocene (*i.e.* what makes the final glacial-interglacial transition unique?)[26,28]. Additionally, it's unlikely that many species with continental or intercontinental distributions would have been equally impacted across the entirety of their range[28].

Among the 38 megafauna genera lost in North America during the late-Pleistocene extinctions, were two proboscidean genera – *Mammut* (mastodons), with two recognized species (*M. americanum, and M. pacificus*), and *Mammuthus* (mammoths), with at least three recognized species (*M. primigenius, M. columbi, and M. exilis*)[28]. Despite casual similarity in appearance, these two groups represented highly diverged lineages, sharing a common ancestor ~24-28 million years ago in Africa[30]. Mastodons are considered first radiation proboscideans, being the descendants of the first proboscideans that leave Africa ~20-23 mya[31], while mammoths along with most other elephants represent some of the latest groups to diverge and eventually leave Africa[31,32].

One of the key changes observed between earlier and later radiating lineages of proboscideans, is the shift from zygodonty to lophodonty, and an increase in the number

of tooth lamallae and crown height[33,34]. Consequently these species were adapted to unique ecological niches, with mastodons typically considered browsers with diets primarily of trees and shrubs, while mammoths are considered grazers with diets centered around grasses and forbs[33,35]. Recent tooth-wear, calculus, and isotope investigations have supported these claims, but also suggest both taxa did exhibit some dietary variability[35–37].

In line with their dietary preferences both taxa are believed to have responded very differently to the climatic stresses of the Pleistocene. Previous genetic work on mammoths has suggested they were restricted to small refugia during interglacials and expanded outwards during glaciations[38]. Palaeontological analyses in the American midcontinent found a similar trend, with mammoths dominant in the region during the last glacial maximum[39]. This same study also noted that American mastodons were absent from this region during the glaciation, likely due to the lack of suitable environments[39]. Similar hypotheses have also been proposed for the extirpation of mastodons in East Beringia during the last glacial period[40].

Both mammoths and mastodons likely acted as quintessential keystone species that exhibited immense ecological impact within their respective environments[33,41]. Modern African and Asian elephant studies highlight their important roles as ecosystem engineers, whose environmental modifications increase species diversity and are important for maintaining savannah landscapes[42–44]. Ancient sedimentary DNA studies have also implicated Pleistocene megafauna in maintaining the diversity of steppe-tundra habitats[41]. Combined with the relative ease with which these large taxa can be identified, these animals represent prime candidates to better understand the roles and effects of Pleistocene stressors.

Since the first ancient DNA sequences reported in the late 20th century, palaeogenomics has rapidly become an important tool in understanding the evolutionary relationships of extinct taxa[45]. The ability to reconstruct mitochondrial and nuclear genomes from ancient organisms and environmental samples allows for a more robust and complete evolutionary picture, and opens the fossil record to population genetics, demographics, and other genetic tools previously only applicable to extant taxa[46]. Broadly speaking, ancient DNA studies interested in the source organism (*i.e.* the animal represented by the palaeontological material) can be split into four categories based on the material, amount of data retrieved, and number of individuals targeted: systematic/taxonomy studies; those working with non-identifiable remains; large-scale phylogenetics; and nuclear genomics. While these four categories are by no means mutually exclusive, I would argue they offer a good way to categorize palaeogenomic studies and highlight their utility in understanding the Pleistocene and its megafauna.

Addressing questions of taxonomy or systematics is one of the earliest applications of ancient DNA[45,47]. These studies are usually interested in the relationships

between extinct taxa and modern relatives[30,48] or interrogating the taxonomic divisions proposed by morphological analysis[49]. They typically feature few, well-identified specimens, and target sub-nuclear genetic loci (*i.e.* partial/complete mitochondrial genomes or nuclear fragments) which are also available for extent species[30,48].

Studies on non-identifiable remains share many of the same characteristics as taxonomy/systematic projects (*i.e.* a low number of specimens and sub-nuclear loci), but as the name would suggest, deal with remains which cannot confidently be assigned to a particular taxa. These non-identifiable remains can range from associated material (*e.g.* palaeofeces)[50–52] to mixed assemblages of broken or otherwise damaged skeletal material[29]. These studies are usually focused on identifying the source organism responsible for the material.

Large-scale phylogenetic analyses seek to shift their focus from the identification or relationships between species, to questions about a single (or a small handful of related) species. By incorporating many individuals, these studies can provide palaeontologically-inaccessible insights into the fossil record, such as identifying migration events or population replacements[38,53]. As their name would suggest, these studies require data from many members of a species, with the temporal and geographic scope of the dataset heavily influencing the type of questions that can be addressed. Given the large amount of data associated with these projects, these studies also typically use sub-nuclear loci.

Nuclear or whole genome sequencing studies, are the most recent application of ancient DNA and enable unprecedented insight into the evolution and extinction of Pleistocene megafauna. These studies have become feasible in the last five to ten years as a result of better methodologies, and the decreasing costs of next-generation sequencing[54]. Despite these advancements, the amount of sequencing necessary to recover complete genomes is still cost-prohibitive for many specimens, and particularly for those samples with poorer preservation. As such, most of these studies have been restricted primarily to low specimen numbers, and target remains from colder and drier environments (*e.g.* Siberia). However, the recovery of full nuclear genomes enables a much more detailed look at the population histories of a species, allowing for the identification of introgression events[55], and the examination of functional adaptations at the genic level[56].

Regardless of the design or questions, palaeogenetic studies are ultimately constrained by their ability to recover authentic ancient DNA. Consequently, most genetic work on Pleistocene megafauna has dealt with specimens from environments that are favourable to DNA preservation (*i.e.* relatively cool, stable temperatures, and low moisture) and primarily from the very late Pleistocene (the last ~50 ky) (see for example[12,14,53,57]). In this thesis I extend our understanding of the genetic landscape of extinct proboscideans, across three chapters, coinciding with the first three categories of

ancient DNA studies outlined above. In chapter 2, I examine work on palaeofecal remains from Bechan Cave, Utah, confirming previous palaoentological hypotheses about the occupants. In chapter 3, I look at the effects of glacial-interglacial cycling on American mastodons through the mid-Pleistocene, and the parallels we may be observing in extant taxa. Finally in chapter 4, I take a closer look at the genetics of mastodons from the American Falls Reservoir in Idaho, and the implications for *Mammut* taxonomy more generally. Together, this work extends the representation of extinct North American taxa from temperate regions, provides the first comprehensive analysis of Mastodon genetics within the continent, and illustrates the immense biogeographic effects that Pleistocene glaciations had on browsing megafauna.

Chapter 1 Bibliography

1. Head, M. J. & Gibbard, P. L. Early-Middle Pleistocene transitions: Linking terrestrial and marine realms. *Quat. Int.* **389**, 7–46 (2015).
2. Suganuma, Y. *et al.* Formal ratification of the Global Boundary Stratotype Section and Point (GSSP) for the Chibanian Stage and Middle Pleistocene Subseries of the Quaternary System: the Chiba Section, Japan. *Episodes*, 1–31 (2021).
3. Augustin, L. *et al.* Eight glacial cycles from an Antarctic ice core. *Nature* **429**, 623–628 (2004).
4. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ 18 O records. *Paleoceanography* **20**, PA1003 (2005).
5. Dyke, A. S. An outline of the deglaciation of North America with emphasis on central and northern Canada. *Quat. Glaciat. Chronol. Part II North Am.* **2b**, 373-424 (2004).
6. Tierney, J. E. *et al.* Glacial cooling and climate sensitivity revisited. *Nature* **584**, 569–573 (2020).
7. Osman, M. B. *et al. Globally resolved surface temperatures since the Last Glacial Maximum*. (2021). doi:https://doi.org/10.31223/X5S31Z
8. Gowan, E. J. *et al.* A new global ice sheet reconstruction for the past 80 000 years. *Nat. Commun.* **12**, 1199 (2021).
9. Dyke, A. S. Late Quaternary Vegetation History of Northern North America Based on Pollen, Macrofossil, and Faunal Remains. *Géographie Phys. Quat.* **59**, 211–262 (2005).
10. Binney, H. *et al.* Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. *Quat. Sci. Rev.* **157**, 80–97 (2017).
11. Warmuth, V. *et al.* Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proc. Natl. Acad. Sci.* **109**, 8202–8206 (2012).
12. Froese, D. *et al.* Fossil and genomic evidence constrains the timing of bison arrival in North America. *Proc. Natl. Acad. Sci.* **114**, 3457–3462 (2017).
13. Timmermann, A. & Friedrich, T. Late Pleistocene climate drivers of early human migration. *Nature* **538**, 92–95 (2016).
14. Chang, D. *et al.* The evolutionary and phylogeographic history of woolly mammoths: a comprehensive mitogenomic analysis. *Sci. Rep.* **7**, 44585 (2017).

15. van der Valk, T. *et al.* Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* **591**, 265–269 (2021).

16. Stuart, A. J. Late Quaternary megafaunal extinctions on the continents: a short review. *Geol. J*. **50**, 338–363 (2015).

17. Koch, P. L. & Barnosky, A. D. Late Quaternary Extinctions: State of the Debate. *Annu. Rev. Ecol. Evol. Syst*. **37**, 215–250 (2006).

18. Kathleen Lyons, S., Smith, F. A., Wagner, P. J., White, E. P. & Brown, J. H. Was a 'hyperdisease' responsible for the late Pleistocene megafaunal extinction? *Ecol. Lett*. **7**, 859–868 (2004).

19. Firestone, R. B. *et al.* Evidence for an extraterrestrial impact 12,900 years ago that contributed to the megafaunal extinctions and the Younger Dryas cooling. Proc. *Natl. Acad. Sci*. **104**, 16016–16021 (2007).

20. Jorgeson, I. A., Breslawski, R. P. & Fisher, A. E. Radiocarbon simulation fails to support the temporal synchroneity requirement of the Younger Dryas impact hypothesis. *Quat. Res. (United States)* **96**, 123–139 (2020).

21. Moore, A. M. T. *et al.* Evidence of Cosmic Impact at Abu Hureyra, Syria at the Younger Dryas Onset (~12.8 ka): High-temperature melting at >2200 °C. *Sci. Rep.* **10**, 4185 (2020).

22. Kjær, K. H. *et al.* A large impact crater beneath Hiawatha Glacier in northwest Greenland. *Sci. Adv*. **4**, 1–11 (2018).

23. Lyons, S. K., Smith, F. A & Brown, J. H. Of mice, mastodons and men: human-mediated extinctions on four continents. *Evol. Ecol. Res*. **6**, 339–358 (2004).

24. Martin, P. S. The Discovery of America: The first Americans may have swept the Western Hemisphere and decimated its fauna within 1000 years. *Science* **179**, 969–974 (1973).

25. Meltzer, D. J. Pleistocene Overkill and North American Mammalian Extinctions. *Annu. Rev. Anthropol*. **44**, 33–53 (2015).

26. Haynes, G. The catastrophic extinction of North American mammoths and mastodonts. *World Archaeol*. **33**, 391–416 (2002).

27. Nagaoka, L., Rick, T. & Wolverton, S. The overkill model and its impact on environmental research. *Ecol. Evol*. **8,** 9683–9696 (2018).

28. Meltzer, D. J. Overkill, glacial history, and the extinction of North America's Ice Age megafauna. *Proc. Natl. Acad. Sci*. **117**, 28555–28563 (2020).

29. Seersholm, F. V. *et al.* Rapid range shifts and megafaunal extinctions associated with late Pleistocene climate change. *Nat. Commun.* **11**, 2770 (2020).

30. Rohland, N. *et al.* Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biol.* **5**, 1663–1671 (2007).

31. van der Made, J. The evolution of elephants and their relatives in the context of a changing climate and geography. in *Elefantentreich - Eine Fossilwelt in Europa* (eds. Höhne, D. & Schwarz, W.) 340–360 (Landesamt für Denkmalpflege und Archälogie Sachsen-Anhalt and Landesmuseum für Vorgeschichte, 2010).

32. Fisher, D. C. Paleobiology of Pleistocene Proboscideans. *Annu. Rev. Earth Planet. Sci.* **46,** 229–260 (2018).

33. Shoshani, J. Understanding proboscidean evolution: a formidable task. *Trends Ecol. Evol.* **13**, 480–487 (1998).

34. Lister, A. M. The role of behaviour in adaptive morphological evolution of African proboscideans. *Nature* **500**, 331–334 (2013).

35. Green, J. L., DeSantis, L. R. G. & Smith, G. J. Regional variation in the browsing diet of Pleistocene *Mammut americanum* (Mammalia, Proboscidea) as recorded by dental microwear textures. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **487**, 59–70 (2017).

36. Cammidge, T. S., Kooyman, B. & Theodor, J. M. Diet reconstructions for end-Pleistocene *Mammut americanum* and *Mammuthus* based on comparative analysis of mesowear, microwear, and dental calculus in modern Loxodonta africana. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **538**, 109403 (2020).

37. Widga, C. *et al.* Life histories and niche dynamics in late Quaternary proboscideans from midwestern North America. Quat. Res. **100,** 224–239 (2020).

38. Palkopoulou, E. *et al.* Holarctic genetic structure and range dynamics in the woolly mammoth. *Proc. R. Soc. B Biol. Sci.* **280**, 20131910 (2013).

39. Widga, C. *et al.* Late Pleistocene proboscidean population dynamics in the North American Midcontinent. *Boreas* **46**, 772–782 (2017).

40. Zazula, G. D. *et al.* American mastodon extirpation in the Arctic and Subarctic predates human colonization and terminal Pleistocene climate change. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18460–18465 (2014).

41. Willerslev, E. *et al.* Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* **506**, 47–51 (2014).

42. Barnosky, A. D. *et al.* Variable impact of late-Quaternary megafaunal extinction in causing ecological state shifts in North and South America. *Proc. Natl. Acad. Sci*. **113**, 1–6 (2015).

43. Haynes, G. Elephants (and extinct relatives) as earth-movers and ecosystem engineers. *Geomorphology* **157–158**, 99–107 (2012).

44. Remmers, W., Gameiro, J., Schaberl, I. & Clausnitzer, V. Elephant (Loxodonta africana) footprints as habitat for aquatic macroinvertebrate communities in Kibale National Park, south-west Uganda. *Afr. J. Ecol*. **55,** 342–351 (2016).

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

46. Swift, J. A. *et al.* Micro Methods for Megafauna: Novel Approaches to Late Quaternary Extinctions and Their Contributions to Faunal Conservation in the Anthropocene. *Bioscience* **69**, 877–887 (2019).

47. Hagelberg, E. *et al.* DNA from ancient mammoth bones. *Nature* **370**, 333–334 (1994).

48. Rohland, N. *et al.* Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. *PLoS Biol.* **8**, 1–10 (2010).

49. Delsuc, F. *et al.* Ancient Mitogenomes Revisit the Evolutionary History and Biogeography of Sloths the evolutionary history and biogeography of sloths. *Curr. Biol*. **29,** 2031–2042 (2019). doi:10.1016/j.cub.2019.05.043

50. Poinar, H. N. *et al.* Molecular coproscopy: dung and diet of the extinct ground sloth Nothrotheriops shastensis. *Science* **281**, 402–406 (1998).

51. Campos, P. F., Willerslev, E., Mead, J. I., Hofreiter, M. & Gilbert, M. T. P. Molecular identification of the extinct mountain goat, *Oreamnos harringtoni* (Bovidae). *Boreas* **39**, 18–23 (2010).

52. Bon, C. *et al.* Coprolites as a source of information on the genome and diet of the cave hyena. *Proc Biol Sci* **279**, 2825–2830 (2012).

53. Heintzman, P. D. *et al.* Bison phylogeography constrains dispersal and viability of the Ice Free Corridor in western Canada. *Proc. Natl. Acad. Sci*. **113**, 8057–8063 (2016).

54. (NHGRI), N. H. G. R. I. DNA Sequencing Costs: Data. (2020). Available at: https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data.

55. Palkopoulou, E. *et al.* A comprehensive genomic history of extinct and living elephants. *Proc. Natl. Acad. Sci. U. S. A.* **115**, (2018).

56. Lynch, V. J. *et al.* Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. *Cell Rep*. **12**, 217–228 (2015).

57. Campos, P. F. *et al.* Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5675–5680 (2010).

CHAPTER 2

Molecular identification of paleofeces from Bechan Cave, southeastern Utah, USA

Emil Karpinski[a,b], Jim I. Mead[c], Hendrik N. Poinar[a,b,d]

[a] *Department of Biology, McMaster University, 1280 Main St. West, Hamilton, Ontario, Canada, L8S 4K1*
[b] *McMaster Ancient DNA Centre, Department of Anthropology, McMaster University, 1280 Main St. West, Hamilton, Ontario, Canada, L8S 4L9*
[c] *The Mammoth Site, Hot Springs, SD 57747, USA*
[d] *Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, 1280 Main St. West, Hamilton, Ontario, Canada, L8N 3Z5*

A version of this work has been published in Quaternary International (Volume 443, Part A). The main text of the study is presented herein. Please refer to the published work for access to Supplementary Materials.

Karpinski E, Mead JI, and Poinar HN. 2017. Molecular identification of paleofeces from Bechan Cave, southeastern Utah, USA. Quaternary International 443: 140-146. doi: https://doi.org/10.1016/j.quaint.2017.03.068

Author Contributions

JIM and HP conceived the study with feedback from EK. EK conducted all wet lab and bioinformatic analyses. EK wrote the first draft of the manuscript and EK, JIM, and HP revised and wrote subsequent drafts.

Abstract

Recent advances in ancient DNA methodologies have enabled the retrieval of highly degraded DNA from contexts with poor preservation conditions. While paleofeces have previously been shown to contain endogenous DNA of the defecator, the preserved DNA is composed of a mixture of diverse microbial, floral and fungal constituents, with limited DNA from the host. However, in situations where skeletal remains are unavailable, paleofeces can serve as an important alternative genetic source, allowing for the molecular identification of the target species and diet. Here, we describe the extraction of ancient DNA from a paleofecal sample found within Bechan Cave (southeastern Utah, USA). Previous work in the cave has suggested that these remains likely stem from *Mammuthus*. We used a comprehensive proboscidean bait set which was used to enrich a nearly complete mitochondrial genome (81.6%) at an average coverage depth of 8.1x. Phylogenetic analysis of the derived consensus sequence revealed that the Bechan Cave bolus does indeed derive from *Mammuthus*, and its sequence falls within Clade 1 (haplogroups F or C), most similar to specimens identified as *Mammuthus columbi*.

Introduction

Desiccated paleofecal remains from extinct and extant mammals are well known from the dry caves and rock shelters of the arid American Southwest (see overview in Mead and Swift[1]). Some of the earliest discoveries and analyses come from the dung of the extinct herbivore *Nothrotheriops shastensis* (Shasta ground sloth)[2–5]. The predominant means of identification of these dried paleofecal remains (*Nothrotheriops*, *Oreamnos* [mountain goat]) is using external morphology and/or the size fraction of the contents[6–9]. More recently, molecular analyses have enhanced our ability to identify unknown dung producers, as well as elucidate aspects of their diets (*e.g.* shrub ox, *Euceratherium*, Harrington's mountain goat, *Oreamnos harringtoni*; *Nothrotheriops shastensis*)[10–14].

Large accumulations of paleofeces have been found in a rock shelter known as Bechan Cave, along the Colorado Plateau, Utah[15,16] (Fig. 1). The overall morphology of the dung boluses, their size (230 x 170 x 85 mm) (Fig. 2) and plant constituents, suggest that the producer was likely to be *Mammuthus* (mammoth)[8], however, due to the large size of the boluses, there are at minimum two other possible producers in that vicinity and at that time: mastodon (*Mammut*) and two species of ground sloths (*Mylodon* and *Megalonyx)*. Here we provide a molecular analysis of a single bolus from Bechan Cave to further augment the putative identification of the paleofeces.

Regional Setting

Bechan (the Navajo word for 'big feces') Cave is a large sandstone rock shelter on the southern Colorado Plateau of southeastern Utah. Excavation of the sand floor deposit exposed a rich organic layer upwards of 40 cm thick with a total volume in the cavern estimated to be 300 m^3 [8,17]. A minimum of eight different morphologies of dung were recovered, with the dominant form (trampled and entire) identified as belonging to mammoth (see Mead *et al.*[8], for details about the morphology and identification). No skeletal remains of mammoths have been recovered from the cave.

Chronology

Paleofeces attributed to *Mammuthus* based on morphology have been recovered from a number of dry alcoves on the Colorado Plateau including Grobot Grotto, Mammoth Alcove, Oak Haven, Shrubox Alcove, Wither's Wallow, Bechan Cave, and possibly in Cowboy Cave[17,18]. *Mammuthus* dung from Bechan Cave produced radiocarbon dates on six isolate boluses ranging in age from 11,670 ± 300 to 13,505 ± 580 uncorrected radiocarbon years before present. The contents of an additional *Mammuthus* bolus were separated (by O.K. Davis and P.S. Martin) into different

taxonomic groups (*e.g.* grass culms, *Atriplex*, sedge achenes) and AMS dated, producing 14 ages between 11,630 ± 150 to 13,040 ± 280 yr BP[18]. The paleofecal specimen used here for the molecular analysis is from the same dung bed unit that yielded the radiocarbon dates listed above.

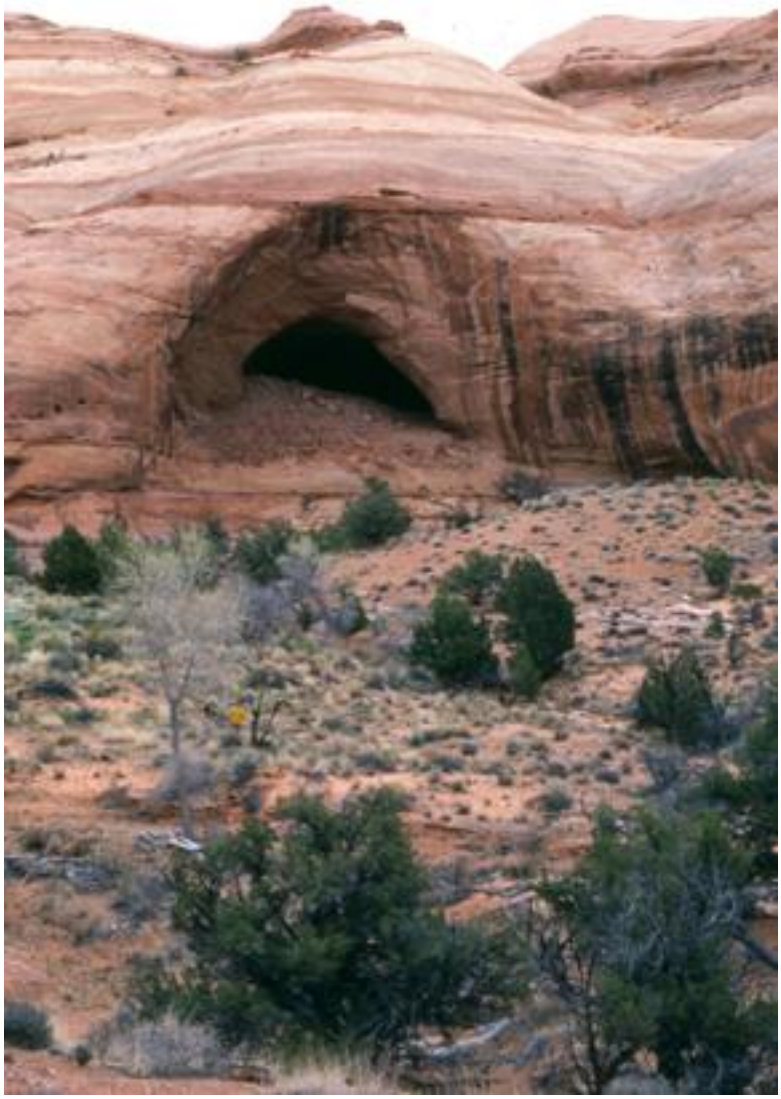


Figure 1: The entrance to Bechan Cave, southeastern Utah. The entrance is approximately 10 m high.

Materials and Methods

A sample of the presumed mammoth bolus (based on characters outlined in Mead *et al.*[8]) was selected by JIM and provided to HP as a 'blind test' to determine the identification of the dung producer. All work done prior to indexing amplification was completed in the dedicated ancient DNA clean rooms of the McMaster Ancient DNA Centre (Hamilton, ON). All post amplification work was done in the McMaster Ancient DNA Centre's BioBubble, a post-amplification clean room in a different building.

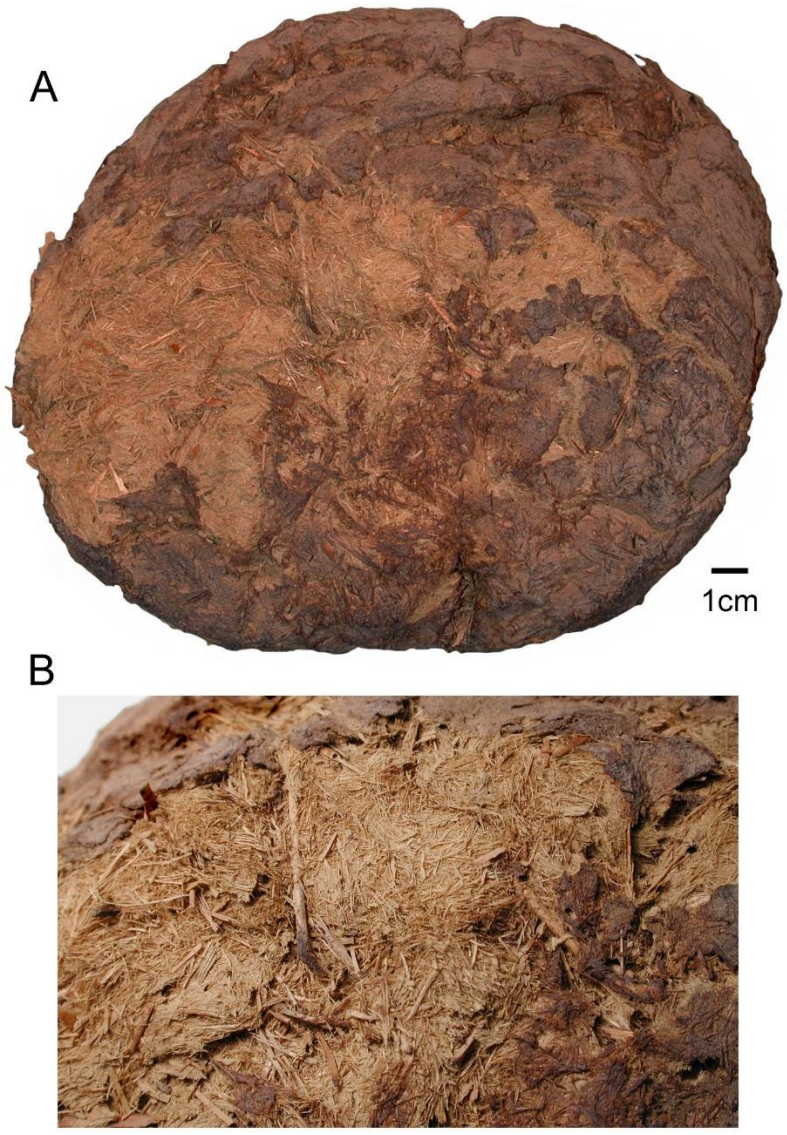


Figure 2: Bechan Cave dung. a, Single dung bolus presumed to be *Mammuthus* from Bechan Cave. **b,** Close-up of bolus showing the size of poorly-chewed contents.

Subsampling and Extraction

We took two subsamples (10 mg and 35 mg) from the bolus (labelled sample #SP442) and as a control, a subsample of *Mylodon darwinni* bone (40 mg) as well as an extraction control (a tube with no sample) to act as a carrier and extraction blank respectively (Appendix A: Subsampling). All samples were washed with 500 µl of 0.5 M EDTA for 1.5 hours at 400 rpm and room temperature in an Eppendorf ThermoMixer followed by an 18 hour demineralization with 1 ml of 0.5 M EDTA under identical conditions. Samples were then spun down, supernatants removed and then digested using 500 µl of a proteinase K buffer (Table A.1) at 50°C with rotation in a hybridization oven for 3 hours.

Supernatants were extracted with 900 µl of PCI (phenol/chloroform/ isoamyl alcohol) (pH 8) and 600 µl of chloroform. Demineralized supernatants underwent an additional round of PCI extraction due to substantial discoloration of the organic fraction. Both aqueous phases were then concentrated over a 30kDa Amicon Ultra 0.5 Centrifugal Filter tubes (EMD Millipore) and washed three times with 1xTE (pH 8–8.5). 15 µl of each extract was then additionally purified through MinElute PCR Purification Kit spin columns (Qiagen) using 6:1 volumes of Buffer PB, two washes with 750 µl of PE, and eluted in 15 µl of TEB (Buffer EB with 1 mM EDTA) and stored at -20°C.

Library Preparation

10 µl of each MinElute purified extract were converted into UDG-treated Illumina sequencing libraries using double-stranded methods of library preparation described previously[19,20] (Appendix A: Library Preparation). Libraries were amplified using unique P5 and P7 indexing primers[20] and purified over MinElute using the manufacturer's protocol eluting in 13 µl EBT. Exact concentrations for each step of library preparation and indexing can be found in Tables A.2 through A.6.

Proboscidean Enrichment and Sequencing

In-solution enrichment was carried out with 10 µl of indexed library using a previously designed bait set[21]. The bait set comprises five *Mammuthus* mitochondrial genomes (GenBank Accession #NC015529, EU153447, EU153453, EU153456, and KX027526) and a single *Mammut americanum* mitochondrial genome (NC009547). Samples and blanks underwent two rounds of enrichment and sequencing using a modified version of the MYbaits protocol (MYcroarray) (Appendix A: Enrichment).

Size selected libraries (150 bp to 600 bp) were sequenced on an Illumina HiSeq 1500, with a 2x85 bp (enrichment round 1) or 2x90 bp (enrichment round 2) paired-end

double-index protocol at the Farncombe Metagenomics Facility (McMaster University, ON). Sequencing data was demultiplexed using CASAVA v1.8.2.

Sequence Curation and Analysis

Demultiplexed reads from both enriched libraries were pooled, trimmed, merged, and all reads greater than or equal to 24 bp were mapped against four mitochondrial reference genomes, those of a woolly mammoth (*Mammuthus primigenius* - NC007596), Columbian mammoth (*Mammuthus columbi* - NC015529), American mastodon *(Mammut americanum* - NC009574), and mylodon ground sloth (*Mylodon darwinii* - KR336794) using a custom pipeline (Appendix A: Mapping Pipeline). Mapped files were imported into Geneious v6.1.5 and screened for any insertions or deletions (indels), by checking for their presence or absence in all reads mapped to a specific region. Indels which did not appear in all reads mapped to a particular region were removed from the consensus. Regions identified with less than 3x coverage were masked with N's. We obtained consensus sequences via mapping against each reference (>50% SNP support or a minimum of 2 out of 3 reads in low-coverage regions) and aligned them with 23 representative complete mammoth mitochondrial genomes spanning all major haplogroups across North America and Siberia (Table A.15) using MUSCLE v3.8.31[22]. The variable number tandem repeat (VNTR) of each sequence was masked with 10 N's. We used jModelTest v2.1.4[23] using the corrected Akaike information criterion (AICc) to choose an appropriate substitution model for all subsequent analyses.

Maximum likelihood phylogenies were generated using iqtree v0.9.6[24] for each consensus sequence generated via mapping to our four references separately as well as pooled together, using both TN+G4 and TIM+G4 substitution models and 200 bootstraps. Five BEAST v1.8.0[25] runs of 10 million generations each were performed using the *Mammuthus columbi* mapped consensus sequence along with the same 23 representative *Mammuthus* mitochondrial genomes. Individual runs were visualized with tracer v1.6.0, before being combined through logcombiner v1.8.0 (Table A.16). Maximum clade credibility trees were generated using TreeAnnotater v1.8.0 with a 10% burn-in. An Asian elephant mitochondrial genome (EF588275) was used as an outgroup.

In addition, we analyzed the total (post trimming and merging) and mapped fractions (>23 bp; *Mammuthus columbi* reference) of our sequence data via BLASTN v2.3.0[26] taking the top 5 hits. The output was visualized in MEGAN v5.11.3[27] and KRONA v2.6[28].

Xenarthran Enrichment

As an additional control, the remaining library from the samples and blanks was indexed and subject to two rounds of in-solution enrichment using a previously designed xenarthran bait set[29]. The xenarthran bait set was designed using representatives of modern xenarthran mitochondrial genomes and ancestrally inferred mitochondrial genomes to maximize the capture potential from extinct taxa[29]. Enrichments were carried out with 5 µl of indexed library, using an optimized protocol (Appendix A: Xenarthran Enrichment). All remaining preparation and analysis steps were identical to those used for the proboscidean enriched libraries, with the exception that sequencing was done only once using a 2x90 bp run on an Illumina HiSeq 1500, and data was demultiplexed using bcl2fastq v2.17.1.14 (Illumina).

Results

Proboscidean Enriched Libraries

We obtained a total of 5,605,400 raw sequencing reads across both subsamples and enrichment rounds (2,864,055 reads post trimming and merging). The 35 mg subsample performed better during enrichment, as seen via mapping, in almost every case except the first enrichment round mapping against *Mammut americanum* (Table 1).

Mapping of the total read dataset against the woolly or Columbian mammoth reference genomes (*Mammuthus primigenius* NC007596] and *Mammuthus columbi* [NC015529]) produced more uniquely mapped reads than they did when mapped to the mastodon (*Mammut americanum* [NC009574]) reference genome (3822 and 3825 vs. 2119 respectively) (Table 1). Mapping against the *Mammuthus columbi* reference slightly outperformed mapping against *Mammuthus primigenius*, with slight increases in reference coverage and depth (81.6% at 8.1x and 80.1% at 8.0x respectively). Mean unique mapped fragment lengths followed the same trend, increasing slightly from 34 bp against *Mammut americanum* to 35.8 bp against both the *Mammuthus columbi* and *Mammuthus primigenius* references (Fig. A.1).

Following enrichment with the proboscidean bait set, only 392 reads (with a mean fragment length of 31.2 bp) of the entire dataset mapped to a *Mylodon darwinii* reference (KR336794) producing a mean coverage of about 0.7x and covering ~7.0% of the reference (Table 1). Not surprisingly, the vast majority of these fragments aligned to conserved regions of the 16S and 12S rRNAs as well as conserved regions of the D-loop. In these spots reads stacked with mean coverage within stacks ranged from 10.2x to 26x (Fig. A.2).

Mapping of the carrier blank reads (7,543,298 raw sequencing reads) following enrichment with the proboscidean bait set, against each of the three proboscidean mitochondrial reference genomes, produced a similar trend with between 16,877 (*Mammut americanum*) and 17,627 (*Mammuthus columbi*) reads mapping (Table 1). The majority of these reads were concentrated into a few conserved regions (primarily D-loop, 12S rRNA, and 16S rRNA) where they formed stacks at very high coverage (540x to 1416.7x). Mapping of the carrier blank reads against the *Mylodon darwinii* reference resulted in 59,831 reads mapped, covering 80.6% of the reference with a mean depth of 224.5x, as would be expected given our choice of carrier. No sequence data was obtained for the extraction blank.

A BLAST analysis of the proboscidean enriched mapped carrier reads (against the *Mammuthus columbi* reference) further illustrated the non-specificity of these reads with only 23 of the total mapped reads being identified as Elephantidae. But upon closer examination, the 23 Elephantidae-identified carrier reads were found to have worse mapping quality (mean quality of 31.2) than the carrier reads mapped against the *Myldon darwinii* reference (mean quality 36.6) or the *Mammuthus columbi* mapped reads from the Bechan Cave sample (mean quality 36.9). Twenty-two of these reads were also found to stack in conserved regions such as the 16S rRNA (7 reads), tRNA-Leu (6 reads), and the ND5 CDS (9 reads). One read mapped to the COX3 gene. In comparison, even our worst-performing library (10 mg sample following enrichment round 2 [26 reads]) contained only two regions with two overlapping mapped reads each. Additionally, all the reads that mapped to the 16s rRNA and tRNA-Leu regions were found to map better (defined as fewer substitutions between the reads and the reference) to the *Mylodon darwinii* reference, while the COX3 read was found to map equally well. Finally, the 23 identified Elephantidae reads in the carrier blank comprise a relatively small proportion (0.0007% of unique carrier reads post trimming and merging) when compared with the Elephantidae-identified reads of the paleofecal sample (0.1% unique reads post trimming and merging). Therefore, while it is possible these reads comprise contamination (either cross between the dung sample and the blank or background airborne contamination) or have been incorrectly identified by BLAST (due to random similarity or database underrepresentation), it is likely that these and similar reads did not have any effect on the consensus we generated from our paleofeces.

Xenarthran Enriched Libraries

We obtained 11,837,066 raw reads combined from the 10 mg and 35 mg subsample libraries following enrichment with the xenarthran bait set. Mapping against each of the three proboscidean reference genomes produced between 872 (*Mammut americanum*) and 1132 reads (*Mammuthus primigenius* and *Mammuthus columbi*), and 609 reads when mapped against the *Mylodon darwinii* reference (Table 2). The vast

majority of these reads tend to be concentrated in conserved regions, primarily within rRNA, tRNA, and intragenic regions (Fig. A.3). Notably, 58% to 65% of the mapped reads to *Mammut americanum* reference and both *Mammuthus* references were still identified as Elephantidae (Fig. A.4 to A.6) through BLAST analysis. Although the number of Elephantidae identified reads falls to 27% when mapped against the *Mylodon darwinii* reference, it still represents the most abundant family-level contributor (Fig. A.7).

In comparison the carrier blank (14,538,132 raw reads) had significantly more mapped reads against each of the proboscidean references, but they were again found primarily within stacks in the same conserved regions as before. However, when mapped against the *Mylodon darwinii* reference we obtained about 725,271 reads mapping, resulting in approximately 87.5% coverage of the reference with a mean depth of 2788.2x.

Mapped Read BLAST

To be certain of the identity (origin) of our reads, we analyzed all 3825 reads that mapped to the *Mammuthus columbi* reference from the combined dataset using the BLASTN algorithm at NCBI. Forty-nine percent of all reads returned hits as belonging to at least *Mammuthus* (Fig. A.8). When considering mapped reads less diagnostic between proboscidean species and could only be identified confidently at the family-level (Elephantidae) this ratio increases to 89%.

Total Enriched Reads BLAST

Reads that could be confidently identified at the familial level (Elephantidae) formed a very small fraction of the total sequenced data from both samples, even after two rounds of enrichment. Following the first enrichment round, the 10 mg and 35 mg subsamples had only 0.08% and 0.12% of the total reads identified as Elephantidae respectively (Figs A.9 and A.10). Following the second round of enrichment, the 10 mg subsample produced 0.29% Elephantidae-identified reads representing a 3.63 fold increase, whereas the 35 mg subsample increased to ~2.2% Elephantidae-identified reads representing an 18.3 fold increase (Figs A.11 and A.12). The majority of all identified reads were bacterial in origin, constituting 83% to 76% of the 10 mg subsample and 78% to 76% of the 35 mg subsample (enrichment rounds 1 and 2 respectively).

Table 1: **Proboscidean Enrichment**. Mapping statistics of both proboscidean enrichment and sequencing rounds for libraries generated from the SP442 subsamples and from the carrier blank. The number of unique mapped reads 24 bp or greater from each subsample or carrier that mapped against each mitochondrial reference genome is shown following enrichment rounds 1 and 2. The Total column contains the total number of unique mapped reads 24 bp or greater that mapped against each reference combing across subsamples (for the 10 mg and 35 mg libraries) and enrichment rounds. The % coverage of each reference genome and the average depth of coverage when all regions with <3x coverage were removed from the consensus and masked with N's of the total dataset, is shown in the coverage column.

Proboscidean Enrichment

Reference Genome	Unique Mapped Reads (>23bp)								Coverage (>=3x)	
	Enrichment Round 1			Enrichment Round 2			Total			
	SP442 Bolus		Carrier	SP442 Bolus		Carrier	SP442	Carrier	SP442	Carrier
	10 mg	35 mg		10 mg	35 mg					
Mammut americanum (NC009574)	129	74	198	11	2014	16,823	2119	16,877	42.7% / 4.1x	15.8% / 48.8x
Mammuthus primigenius (NC007596)	87	134	208	26	3717	17,374	3822	17,430	80.1% / 8.0x	18.3% / 49.1x
Mammuthus columbi (NC015529)	86	134	211	26	3625	17,570	3825	17,627	81.6% / 8.1x	18.1% / 51x
Mylodon darwinii (KR336794)	8	18	1029	4	365	59,229	392	59,831	7.0% / 0.7x	80.6 % / 224.5x

Table 2: **Xenarthran Enrichment.** Mapping statistics of for the SP442 subsamples and the carrier blank following enrichment with the xenarthran bait set. The number of unique mapped reads 24 bp or greater that mapped against each mitochondrial reference genome is shown both individually and combined across subsamples. The % coverage of each reference genome and the average depth of coverage when all regions with <3x coverage were removed from the consensus and masked with N's, is shown in the coverage column.

Xenarthran Enrichment								
Reference Genome	Unique Mapped Reads (>23bp)				Coverage (>=3x)			
	SP442 Bolus			*Mylodon* Carrier	SP442 Bolus			*Mylodon* Carrier
	10 mg	35 mg	Combined (10 mg + 35 mg)		10 mg	35 mg	Combined (10 mg + 35 mg)	
Mammut americanum (NC009574)	222	662	872	28,435	6.3% / 0.4x	12.9% / 1.3x	14.5% / 1.7x	29.4% / 75.0x
Mammuthus primigenius (NC007596)	264	881	1132	29,911	0.4x / 7.5%	20.1% / 1.7x	22.7% / 2.3x	28.2% / 77.4x
Mammuthus columbi (NC015529)	264	880	1132	29,911	7.6% / 0.4x	20.3% / 1.8x	23.0% / 2.3x	28.7% / 80.1x
Mylodon darwinii (KR336794)	173	463	609	725,271	4.6% / 27.5x	8.3% / 29.7x	9.5% / 30.1x	87.5% / 2788.2x

Phylogeny

Our maximum likelihood analyses firmly placed all Bechan Cave consensus sequences within Clade 1, haplotype F/C[21,30], regardless of the reference genome used for mapping, substitution model, or when tested individually or together (Figs A.13 to A.20). This result was further corroborated through our BEAST analyses, with the *Mammuthus columbi* reference and the Bechan Cave consensus clustering within Clade 1, haplogroup F/C with high posterior probability (Fig. 3). Notably, this haplogroup contains all currently described diversity of *Mammuthus columbi* mitochondrial DNA[21].

Conclusions

We have reconstructed a nearly complete mammoth mitochondrial genome from the DNA of a paleofecal bolus from Bechan Cave, Utah. Our results corroborate the morphological analyses of Mead *et al.*[8] in that the dung blanket of Bechan Cave was likely predominantly produced by mammoths and not other proboscideans nor other megafaunal species endemic to the area during the late Pleistocene. The placement of the Bechan Cave consensus sequence within Clade 1, haplogroup F/C suggests that the inhabitants of Bechan Cave were corresponding well with the known phylogeography of North American *Mammuthus* species[21,31,32].

These results further illustrate the utility in working with paleofecal remains for ancient DNA analyses. Paleofeces can be a source of endogenous DNA from the target organism[10,14,33–35], in circumstances where skeletal remains of the defector are absent, such as Bechan Cave. While the high abundance of microbial DNA[36] in palaeofeces restricts the accessibility of target eukaryotic DNA, advances in ancient DNA methodologies (such as targeted enrichment) have allowed for those minute fractions to answer long standing paleontological questions.

Data Availability

The Bechan Cave consensus sequence generated against the *Mammuthus columbi* mitochondrial reference genome has been uploaded to NCBI under the GenBank accession #KX712146.

Conflict of Interest

All authors declare no conflict of interest.

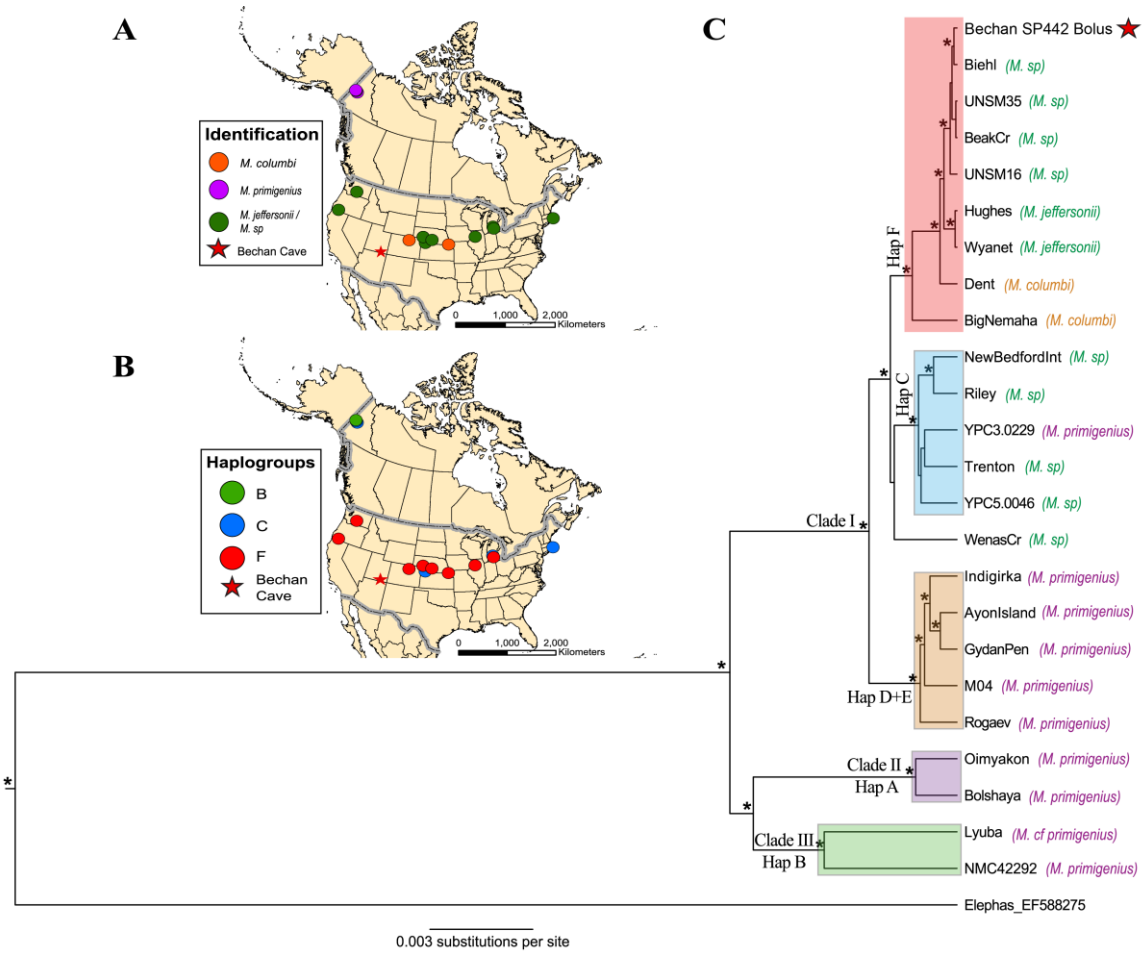


Figure 3: Phylogenenetics of Bechan Cave mammoths. Maps showing the distribution and (**a**) paleontological identification or (**b**) mitochondrial haplotype of representative North American mammoth specimens used in this study. The approximate location of the Bechan Cave specimen is indicated by a red star in both maps. **c,** Combined maximum clade credulity tree with a Bechan consensus generated against the *Mammuthus columbi* reference genome and 23 other representative mammoth genomes from five independent BEAST runs. An Asian elephant whole mitochondrial genome was included as an outgroup. Nodes with posterior support greater than 0.95 are indicated with an asterisks (*). Haplogroup and species identification coloring corresponds to the coloring in (A) and (B). Haplogroups are defined as in Enk *et al.*[21]; however in some studies haplogroups F and C are collapsed (into just haplogroup C) due to the paraphyletic nature of this clade.

Acknowledgments

The authors profusely thank the late Larry D. Agenbroad for the many years of friendship and collaboration on so many research projects, but especially those involving Bechan Cave. Without Larry's expertise, enthusiasm, and energy, the Bechan Cave project would not have happened. It was Larry, who upon seeing the large boluses roll out

of the test pit wall and observing that the contents were only slightly chewed long, coarse grasses and twigs, first identified the dung as ‘screamers’. We would also like to thank all members of the McMaster Ancient DNA Centre, past and present, for their help and guidance in processing and analyzing our work. In particular, we would like to extend a special thank you to Dr. Jacob Enk for his work at the conception of this project, and his guidance throughout.

This work was supported by an NSERC Discovery Grant, Accelerator Grant and CRC to HP.

Chapter 2 Bibliography

1. Mead, J. I. & Swift, S. L. Late Pleistocene (Rancholabrean) Dung Deposits of the Colorado Plateau, Western North America. New Mex. *Museum Nat. Hist. Sci. Bull.* **57**, 337–342 (2012).

2. Eames, A. J. Report on ground-sloth coprolite from Dona Ana County, New Mexico. *Am. J. Sci*. **5**, 353–356 (1930).

3. Lull, R. S. The ground sloth, Nothrotherium. *Am. J. Sci*. **344**–352 (1930).

4. Laudermilk, J. D. & Munz, P. A. Plants in the dung of *Nothrotherium* from Gypsum Cave, Nevada. in *Contributions to Paleontology IV* 29–37 (Carnegie Institution of Washington, 1934).

5. Martin, P. S., Sabels, B. E. & Shutler, D. Rampart Cave coprolite and ecology of the shasta ground sloth. *Am. J. Sci.* **259**, 102–107 (1961).

6. Long, A. & Martin, P. S. Death of American ground sloths. *Science* **186**, 638–640 (1974).

7. Spaulding, W. G., Martin, P. S., Genoways, H. H. & Baker, R. J. Ground sloth dung of the Guadalupe Mountains. in *Biological Investigations in the Guadalupe Mountains National Park, Texas* **4**, 259–269 (Natl. Park Service Washington, DC, 1979).

8. Mead, J. I., Agenbroad, L. D., Davis, O. K. & Martin, P. S. Dung of *Mammuthus* in the arid Southwest, North America. *Quat. Res*. **25**, 121–127 (1986).

9. Mead, J. I. & Spaulding, W. G. Pika (Ochotona) and paleoecological reconstructions of the Intermountain West, Nevada and Utah. *Late Quat. Environ. Deep Hist. a Tribut. to Paul S. Martin. Mammoth Site Hot Springs, South Dakota Sci. Pap*. **3,** 165–186 (1995).

10. Poinar, H. N. *et al.* Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* **281**, 402–406 (1998).

11. Hofreiter, M., Mead, J. I., Martin, P. & Poinar, H. N. Molecular caving. *Curr. Biol*. **13**, R693–R695 (2003).

12. Kropf, M., Mead, J. I. & Scott Anderson, R. Dung, diet, and the paleoenvironment of the extinct shrub-ox (*Euceratherium Collinum*) on the Colorado Plateau, USA. *Quat. Res*. **67**, 143–151 (2007).

13. Campos, P. F. *et al.* Clarification of the taxonomic relationship of the extant and extinct ovibovids, *Ovibos*, *Praeovibos*, *Euceratherium* and *Bootherium*. *Quat. Sci. Rev.* **29**, 2123–2130 (2010).

14. Campos, P. F., Willerslev, E., Mead, J. I., Hofreiter, M. & Gilbert, M. T. P. Molecular identification of the extinct mountain goat, *Oreamnos harringtoni* (Bovidae). *Boreas* **39**, 18–23 (2010).

15. Davis, O. K., Agenbroad, L., Martin, P. S. & Mead, J. I. The Pleistocene dung blanket of Bechan Cave, Utah. *Carnegie Museum Nat. Hist. Spec. Publ*. **8**, 267–282 (1984).

16. Mead, J. I., Agenbroad, L. D., Martin, P. S. & Davis, O. K. The mammoth and sloth dung from Bechan Cave in southern Utah. *Curr. Res. Pleistocene* **1**, 79–80 (1984).

17. Agenbroad, L. D., Mead, J. I., Mead, E. M. & Elder, D. Archaeology, Alluvium, and Cave Stratigraphy: The Record from Bechan Cave, Utah. *KIVA* **54**, 335–351 (1989).

18. Mead, J. I. & Agenbroad, L. D. Isotope Dating of Pleistocene Dung Deposits From the Colorado Plateau, Arizona and Utah. *Radiocarbon* **34**, 1–19 (1992).

19. Meyer, M. & Kircher, M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb. Protoc*. **2010**, 1–10 (2010).

20. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res*. **40**, 1–8 (2012).

21. Enk, J. *et al. Mammuthus* population dynamics in Late Pleistocene North America: Divergence, Phylogeogrpaphy and Introgression. *Front. Ecol. Evol*. **4**, 1–13 (2016).

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

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

24. Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol*. **30**, 1188–95 (2013).

25. Drummond, A. J., Suchard, M. a, Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol*. **29**, 1969–73 (2012).

26. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol*. **215**, 403–410 (1990).

27. Huson, D., Mitra, S. & Ruscheweyh, H. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* **21**, 1552–1560 (2011).

28. Ondov, B. D., Bergman, N. H. & Phillippy, A. M. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* **12**, 1–9 (2011).

29. Delsuc, F. *et al.* The phylogenetic affinities of the extinct glyptodonts. *Curr. Biol*. **26**, R155–R156 (2016).

30. Chang, D. *et al.* The evolutionary and phylogeographic history of woolly mammoths: a comprehensive mitogenomic analysis. *Sci. Rep*. **7**, 44585 (2017).

31. Agenbroad, L. D. North American Proboscideans: Mammoths: The state of knowledge. *Quat. Int*. **126–128**, 73–92 (2005).

32. Enk, J. *et al.* Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. *Genome Biol*. **12**, R51 (2011).

33. Poinar, H., Kuch, M., McDonald, G., Martin, P. & Paabo, S. Nuclear Gene Sequences from a Late Pleistocene Sloth Coprolite. *Curr. Biol*. **13**, 1150–1152 (2003).

34. Hofreiter, M. *et al.* A molecular analysis of ground sloth diet through the last glaciation. *Mol. Ecol*. **9**, 1975–1984 (2000).

35. Bon, C. *et al.* Coprolites as a source of information on the genome and diet of the cave hyena. *Proc Biol Sci* **279**, 2825–2830 (2012).

36. Kirillova, I. V. *et al.* The diet and environment of mammoths in North-East Russia reconstructed from the contents of their feces. *Quat. Int*. **406**, 147–161 (2016).

CHAPTER 3

American mastodon mitochondrial genomes suggest multiple dispersal events in response to Pleistocene climate oscillations

Emil Karpinski[1,2], Dirk Hackenberger[1,3], Grant Zazula[4,5], Chris Widga[6], Ana T. Duggan[1,7], G. Brian Golding[2], Melanie Kuch[1], Jennifer Klunk[1,8], Christopher N. Jass[9], Pam Groves[10], Patrick Druckenmiller[11,12], Blaine W. Schubert[6], Joaquin Arroyo-Cabrales[13], William F. Simpson[14], John W. Hoganson[15], Daniel C. Fisher[16], Simon Y.W. Ho[17], Ross D.E. MacPhee[18], Hendrik N. Poinar[1,3,7.]

[1] McMaster Ancient DNA Centre, Departments of Anthropology and Biochemistry, McMaster University, Hamilton, ON, L8S 4L9, Canada.
[2] Department of Biology, McMaster University, Hamilton, ON, L8S 4L8, Canada.
[3] Department of Biochemistry, McMaster University, Hamilton, ON, L8S 4L8, Canada.
[4] Yukon Palaeontology Program, Department of Tourism and Culture, Government of Yukon, Whitehorse, YT, Y1A 2C6, Canada.
[5] Research and Collections, Canadian Museum of Nature, Ottawa, ON, K2P 2R1, Canada.
[6] Center of Excellence in Paleontology and Department of Geosciences, East Tennessee State University, Johnson City, TN, 37614, USA.
[7] Department of Anthropology, McMaster University, Hamilton, ON, L8S 4L9, Canada.
[8] Arbor Biosciences, Ann Arbor, MI, 48103, USA.
[9] Quaternary Palaeontology Program, Royal Alberta Museum, Edmonton, AB, T5J 0G2, Canada.
[10] Institute of Arctic Biology, University of Alaska Fairbanks, AK, 99775, USA.
[11] Department of Geosciences, University of Alaska Fairbanks, AK, 99775, USA.
[12] University of Alaska Museum, University of Alaska Fairbanks, AK, 99775, USA.
[13] Laboratorio de Arqueozoologia, SLAA, Instituto Nacional de Antropología e Historia, Ciudad de México, 06600, México.
[14] Gantz Family Collections Center, Field Museum of Natural History, Chicago, IL, 60605, USA.
[15] North Dakota Geological Survey, Bismarck, ND, 58505, USA.
[16] Museum of Paleontology and Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, 48109, USA.
[17] School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, 2006, Australia.
[18] Department of Mammalogy/Vertebrate Zoology, American Museum of Natural History, New York, NY, 10024, USA.

A version of this work has previously been published in Nature Communications (Volume 11, article 4048). The main text of the study is presented herein. Please refer to the published work for access to Supplementary Materials.

Karpinski E, Hackenberger D, Zazula G, Widga C, Duggan AT, Golding GB, Kuch M, Klunk J, Jass CN, Groves P, Druckenmiller P, Schubert BW, Arroyo-Cabrales J, Simpson WF, Hoganson JW, Fisher DC, Ho SYW, MacPhee RDE, and Poinar HN. 2020. American mastodon mitochondrial genomes suggest multiple dispersal events in response to Pleistocene climate oscillations. Nature Communications 11: 4048. doi: https://doi.org/10.1038/s41467-020-17893-z

Author Contributions

E.K., G.Z., H.P., and R.D.E.M conceived the study. G.Z., C.W., C.N.J., P.G., P.D., B.W.S., J.A-C., W.F.S., J.W.H., D.C.F., and R.D.E.M. subsampled mastodon material and collected palaeontological information. E.K., D.H, M.K., and J.K. conducted wet-lab analysis. E.K., A.T.D., G.B.G, and S.Y.W.H performed all bioinformatics analyses. E.K., G.Z., H.P., S.Y.W.H and R.D.E.M drafted the manuscript. All authors contributed to manuscript revisions.

Abstract

Pleistocene glacial-interglacial cycles are correlated with dramatic temperature oscillations. Examining how species responded to these natural fluctuations can provide valuable insights into the impacts of present-day anthropogenic climate change. Here we present a phylogeographic study of the extinct American mastodon (*Mammut americanum*), based on 35 complete mitochondrial genomes. These data reveal the presence of multiple lineages within this species, including two distinct clades from eastern Beringia. Our molecular date estimates suggest that these clades arose at different times, supporting a pattern of repeated northern expansion and local extirpation in response to glacial cycling. Consistent with this hypothesis, we also note lower levels of genetic diversity among northern mastodons than in endemic clades south of the continental ice sheets. The results of our study highlight the complex relationships between population dispersals and climate change, and can provide testable hypotheses for extant species expected to experience substantial biogeographic impacts from rising temperatures.

Introduction

Anthropogenic climate change is causing considerable increases to the earth's mean surface temperatures[1] and ecological stress[2]. As a consequence, many species are experiencing demographic declines or extinction, or are shifting their ranges into regions that have become more habitable given new environmental conditions[3–6]. Although human practices have been the primary cause of global temperature changes over the past century, the phenomenon of large-scale, climatically driven environmental change has occurred numerous times, at varying temporal scales, during the Quaternary period the last 2 Ma. The largest changes involved the alternation of glacial and interglacial intervals, which for the past 800 thousand years (ky) have operated on cycles of ~100 ky[7]. These cycles resulted in periodic ice-sheet expansion across ~50% of the habitable land in North America[8] and fluctuations in global mean annual temperatures of greater than 10 °C[9–11].

The high-magnitude climate changes associated with glacial-interglacial cycles also resulted in dramatic rearrangement of North American terrestrial ecosystems and vegetation zonation[12]. Furthermore, these climate changes had substantial impacts on the amount of habitable land available as glacial formation and advance coupled with global sea level drop enabled new continental shelf lands to emerge during Pleistocene cold periods. Inversely, periods of climatic warming, such as those during previous interglaciations, resulted in climatic and biogeographic configurations across the continent similar to those experienced today. Examining the phylogeographic and demographic impacts of these major climatic oscillations on ancient populations can inform our understanding of how species respond to this scale of change and aid in the construction of predictive frameworks.

Ancient DNA recovered from fossil bones and teeth enables us to directly examine the genetics of extinct species over long periods of time. These methods can provide a nuanced understanding of responses (*e.g.* migration and extinction) to climatic stressors during the Pleistocene, and have already revealed demographic trends that are not easily recovered with traditional palaeontological techniques[13–16]. Most phylogeographic studies of North American Pleistocene taxa have focused on species adapted to grassland or steppe-tundra, and their responses to the arrival of humans and terminal Pleistocene warming[13,17–19]. Yet past climate change, particularly intervals of sharply increasing temperatures such as that which occurred during Marine Isotope Stage (MIS) 5e (Sangamonian interglaciation) ~125,000 years ago (kya), is also likely to have had substantial impacts on megafaunal populations[14,20,21]. This kind of climate-driven pressure would have especially applied to species adapted to forests and mixed woodland habitats, because these biomes, which greatly expanded during warmer interglacial intervals, were subsequently replaced or rendered inaccessible during subsequent glaciations[12,20].

American mastodons (*Mammut americanum*) were an iconic part of wooded and swampy habitats in Pleistocene North America[22–24], with remains recovered from the Central American subtropics to the Arctic latitudes of Alaska and Yukon[20,25]. Stable isotope data, dental morphology, and microwear analyses reveal some regional and chronological variation or flexibility in diet, although C_3 woody browse vegetation (*e.g.* spruce trees) seems to have been preferred[25–28]. Like most proboscideans, the mastodon was a keystone species and served an important role in maintaining the integrity and diversity of its preferred habitats[25,29,30].

According to recent palaeontological investigations, mastodons and mammoths displayed contrasting responses to cyclical glacial-interglacial climatic shifts. Temporal analyses of mastodon distribution patterns within the American midcontinent[31] and eastern Beringia (present-day unglaciated areas of Alaska and Yukon)[20] have inferred that American mastodons briefly expanded into high latitudes during the last interglaciation (MIS 5), but underwent regional extirpation when climates became much colder during the last glaciation (*i.e.*, MIS 4 to MIS 2), surviving thereafter only in lower-latitude temperate regions in North America. These extirpations were likely caused by climate-driven changes in vegetation at the onset of glaciation, which, by contrast, favoured the spatial expansion of mammoths and other grazing species adapted to steppe-tundra. However, these arguments remain difficult to test empirically. This is particularly the case for eastern Beringia where there is often little or no chronostratigraphic context for mastodon fossils, and which regularly return age estimates greater than the effective limit of radiocarbon (^{14}C) analysis (*i.e.*, >50,000 years ago)[20]. Alternative dating methods, such as optically stimulated luminescence, may eventually prove useful but have yet to be applied to questions like these.

Here we present an alternative approach to testing models of expansion-extirpation due to glacial-interglacial cycling, using detailed phylogeographic analysis and Bayesian clock dating of American mastodon mitochondrial genomes. Our findings suggest that American mastodons repeatedly expanded into northern latitudes in response to interglacial warming. However, we also note that northern clades have extremely low levels of genetic diversity, highlighting an important consideration for the conservation of modern species exhibiting similar dispersal patterns.

Results and discussion

Mastodon Phylogeography

Subsamples from fossil bones and teeth of American mastodons were obtained from museums, universities, and government institutions across North America (Supplementary Data 1). Complete mitochondrial genomes were sequenced from 33 of 122 specimens (~27% success rate), with completeness defined as >80% sequence

coverage of the *M. americanum* mitochondrial reference (GenBank accession NC_035800), at a minimum coverage depth of 3× (Fig. 1A; Supplementary Fig. 2). Alignments contained 1,492–51,113 uniquely mapped reads, with the short inserts (mean fragment sizes 37.26–76.63 bp) and terminal cytosine deamination that are characteristic of authentic ancient DNA (Supplementary Methods – Sequence Authenticity and Map Damage). Partial sequences were also obtained from another 12 specimens, ranging from 46 to 580 uniquely mapped reads, but these were excluded from all subsequent analyses (Supplementary Table 31).

We identified five well supported major clades in the mitogenomic phylogeny (Fig. 1B), through a combination of maximum-likelihood and Bayesian methods (Supplementary Methods – Phylogenetic Analyses). We named the five clades by the approximate geographic provenance of their constituent specimens (*i.e.*, A = Alaska; Y = Yukon; G = Great Lakes; M = Mexico; L = Alberta/Missouri). We included two specimens from Virginia (southeastern USA; ETMNH 19334 and ETMNH 19335) in Clade G, due to the lower support for their monophyly. We also tentatively assigned a single specimen, NSM092GF182.011 from Nova Scotia (eastern Canada), dated to 74.9 ± 5.0 ky[32], to a separate Clade N. This specimen is geographically and temporally distinct; given the deep divergence that we estimate for this lineage, it is likely to represent a separate group of east coast mastodons from the Sangamonian interglaciation of MIS 5.

The phylogenies inferred by Bayesian and maximum-likelihood methods were consistent, except for the placements of the two specimens in Clade M (DP1296 and RAM P97.7.1). When the tree is rooted with a *Mammuthus* (mammoth) outgroup, Clade M is rendered paraphyletic, albeit with low bootstrap support (79%). However, the monophyly of this clade is supported by midpoint-rooted trees inferred using maximum likelihood, as well as by Bayesian methods (Fig. 1B; Supplementary Methods – Phylogenetic Analyses).

We find evidence of broad phylogeographic structure, with mastodons from neighbouring localities generally being more closely related. This trend is also observed in North American mammoths[33], as well as in African[34] and Asian elephants[35], and is due to the matrilocal nature of proboscidean herds. A matrilocal herd structure for mastodons has also previously been argued based on difference in tusk growth profiles between males and females upon reaching sexual maturity[36], and relationships in preserved trackways[37]. Notably, female proboscidean philopatry also results in deep divergences between clades[38] and possibly explains the deep divergences that we observe in mastodons.

Despite the limited geographic dispersal expected within proboscidean matrilines, we identify two independent and genetically divergent clades (A and Y) that consist primarily of specimens from eastern Beringia. Clade Y is grouped with Clades G, L, and N, and diverged from Clade A between 1.37 million years ago (95% HPD interval: 857–

1,881 kya, when the ages of all undated ancient samples are estimated simultaneously in a "Joint" analysis) and 609 kya (95% HPD interval: 335–998 kya, when the ages of undated ancient samples are "Individually Dated").

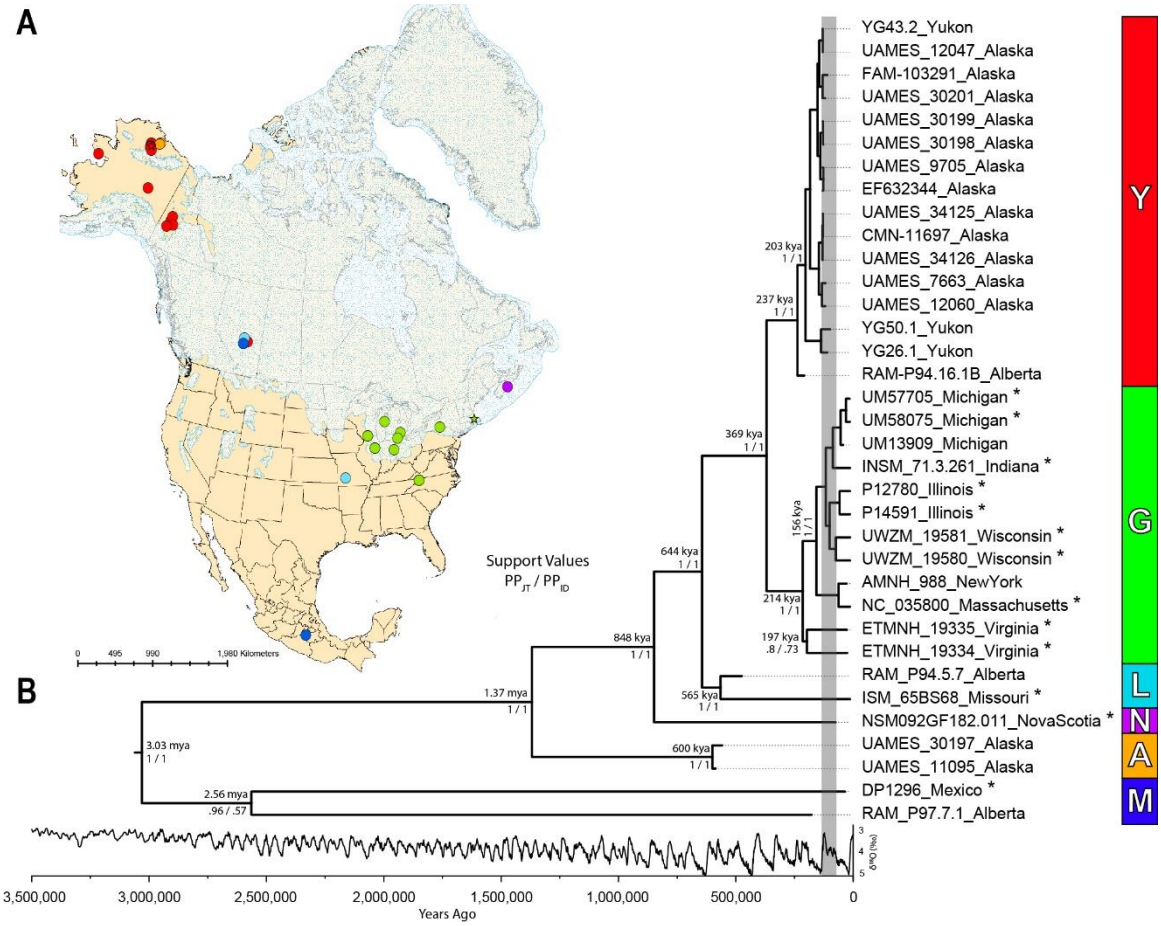


Fig. 1: Phylogeographic relationships of American mastodons. a, Locations of specimens included in this study. Circles indicate specimens for which complete mitochondrial genomes were obtained, coloured according to their clade assignment. Stars indicate previously sequenced specimens (EF632344, Alaska; NC_035800, Massachusetts). The locations of specimens from Alberta and eastern Beringia have been jittered to aid visualization (see Supplementary Fig. 2 for an unmodified version). **b,** Phylogenetic tree inferred by the Joint analysis. Median posterior ages are given for major nodes in the tree. Support values for each node represent posterior probabilities from the Joint (PP_{JT}) and Individually Dated (PP_{ID}) analyses. The $\delta^{18}O$ record for the last 3.5 million years is shown below the tree[7], with the approximate extent of the MIS 5 interglacial period shaded in grey. Clades identified in this study are designated by colour and named where possible by the geographic provenance of their members (*i.e.*, A = Alaska; Y = Yukon; G = Great Lakes; M = Mexico; N = Nova Scotia; L = Alberta/Missouri). Specimens with finite radiometric or geological ages are indicated with asterisks.

Specimens from Alberta are found in three of the five well-defined clades (L, M, and Y; Fig. 1), highlighting a complex ecological and biogeographical history that could not have been recovered from the palaeontological record alone. Notably, central Alberta was the site of maximum convergence of the Laurentide and Cordilleran ice sheets, and contained the earliest deglaciated corridor connecting areas north and south of the ice sheets following latest Pleistocene deglaciation[8]. Previous work on *Bison* has shown that this region was also the site of dynamic range changes in response to this last deglaciation episode, with rapid population expansions into the region from both Beringian and southern areas coinciding with glacial retreat[16,19]. While further research is required to determine whether these findings also apply to other taxa and time periods, the phylogenetically divergent placements of Alberta mastodons suggest that this region was one of immense biological fluidity for this species as well.

Mastodon Sample-Age Estimation

To explain the extirpation of American mastodons in eastern Beringia during the last glaciation, Zazula *et al.*[20] proposed a palaeoecological model that tied mastodon occupation generally to the MIS 5 interglacial, when the regional vegetation would have been dominated by mixed boreal forests and wetlands[39]. Assuming that mastodon presence in Beringia varied with vegetation type, the question becomes whether there was a repeated pattern of mastodon expansion and extirpation corresponding to the ~100 ky glacial-interglacial cycle (Fig. 2). Evidence for extending the palaeoecological model in this way should include both a temporal signal (*i.e.*, high-latitude mastodon presence should correlate with known interglacial intervals) and a biogeographic signal (*i.e.*, high-latitude populations ought to display lower levels of genetic diversity than southern populations, as a consequence of restricted subsets of matrilocal herds expanding northward as environmental conditions ameliorated).

Mastodon specimens from eastern Beringia and Alberta were determined by radiocarbon analysis to be greater than 50,000 years old, or analytically non-finite, and are often found out of stratigraphic context[20,40]. In the absence of any other applicable direct-dating method associated with a large date archive, we used a molecular clock analysis[41,42] to estimate the ages of high-latitude specimens. Although molecular dating tends to produce date estimates that are less precise than those of some radiometric or geological methods, its accuracy has been demonstrated in studies of simulated and molecular data[41] as well as morphological data[42]. Additionally, in cases where little to no temporal information is available, estimating the ages of undated specimens allows the inclusion of these specimens when they would otherwise need to be excluded from molecular dating analyses. We used two separate approaches to estimate the ages of the undated specimens: one where the dates of all specimens were estimated simultaneously

(Joint; JT), and one where we estimated the dates of specimens individually before analysing all of the samples together (Individually Dated; ID).

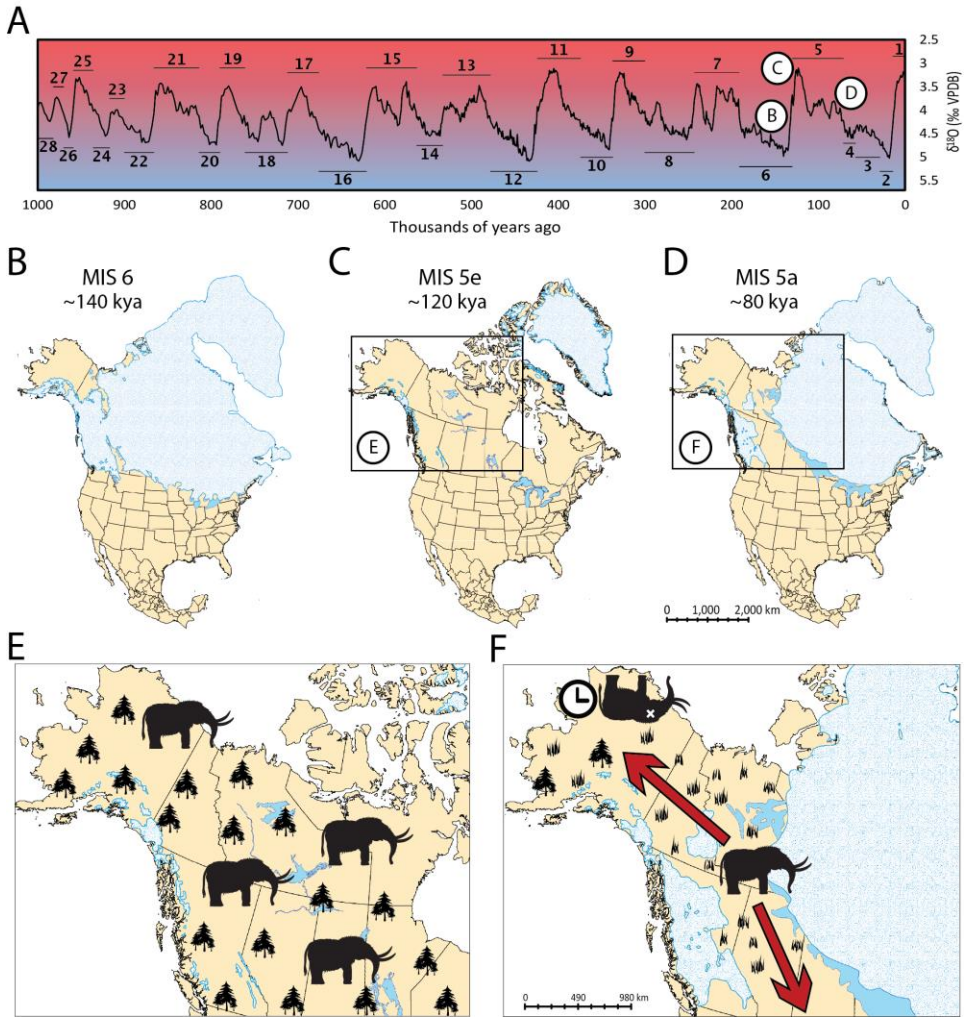


Figure 2: Model of mastodon extirpation and expansion in response to glacial cycles. a, Global stack of benthic foraminifera $\delta^{18}O$ for the last 1 million years, which tracks changes in deep-water temperature and global ice volume. The y-axis has been inverted so that periods of low ice buildup (and higher temperatures – red) are at the top of the graph, and periods of greater ice buildup (and lower temperatures – blue) are at the bottom. Marine Isotope Stage (MIS) extents are indicated with black bars above (interglacials) or below (glacials) the $\delta^{18}O$ record. $\delta^{18}O$ values and MIS terminations can be found in Lisiecki & Raymo[7]. One full glacial cycle is represented, showing the change from glacial (**b**) to interglacial (**c**) conditions, followed by a fall back into another glaciation (**d**). North American ice-sheet cover at each stage (**c-d**) is approximated from recorded $\delta^{18}O$ to similar conditions during the transition out of the last glaciation[8], or from published simulations where available (**b**)[65]. The ecological implications of these transitions are summarized in **e-f**, with mastodons being able to occupy most of eastern Beringia and Canada during interglacials (**e**), but progressively extirpated from these regions as conditions descend into the next glacial period (**f**). Populations would either need to retract to unglaciated regions south of the ice sheets or north to temporarily unglaciated refugia which would be unlikely to support mastodon populations throughout long glaciations.

Median posterior ages for the east Beringian Clade Y mastodons ranged between 98–130 kya (JT) and 74–91 kya (ID) (Fig. 3), falling within the boundaries of the MIS 5 interglacial, the last major extended warm period prior to the Holocene[7]. Although the 95% HPD intervals of the age estimates for individual samples are wide (combined Beringian mastodon 95% HPD range – JT: 50–219 ky; ID: 50–125 ky), their probability density is concentrated around times that correspond well with MIS 5, with the mode of each distribution also being located within the timespan of the MIS 5 interglaciation (Supplementary Table 42). Additionally, while the Joint analysis also includes a small subset of ages corresponding to the previous interglacial (MIS 7; ~191–243 kya)[7] within some specimens' 95% HPD interval, these ages are not recovered in the Individually Dated analysis. These findings strongly suggest that mastodon habitation in Eastern Beringia coincided with interglacial periods, as expected under our palaeocological model.

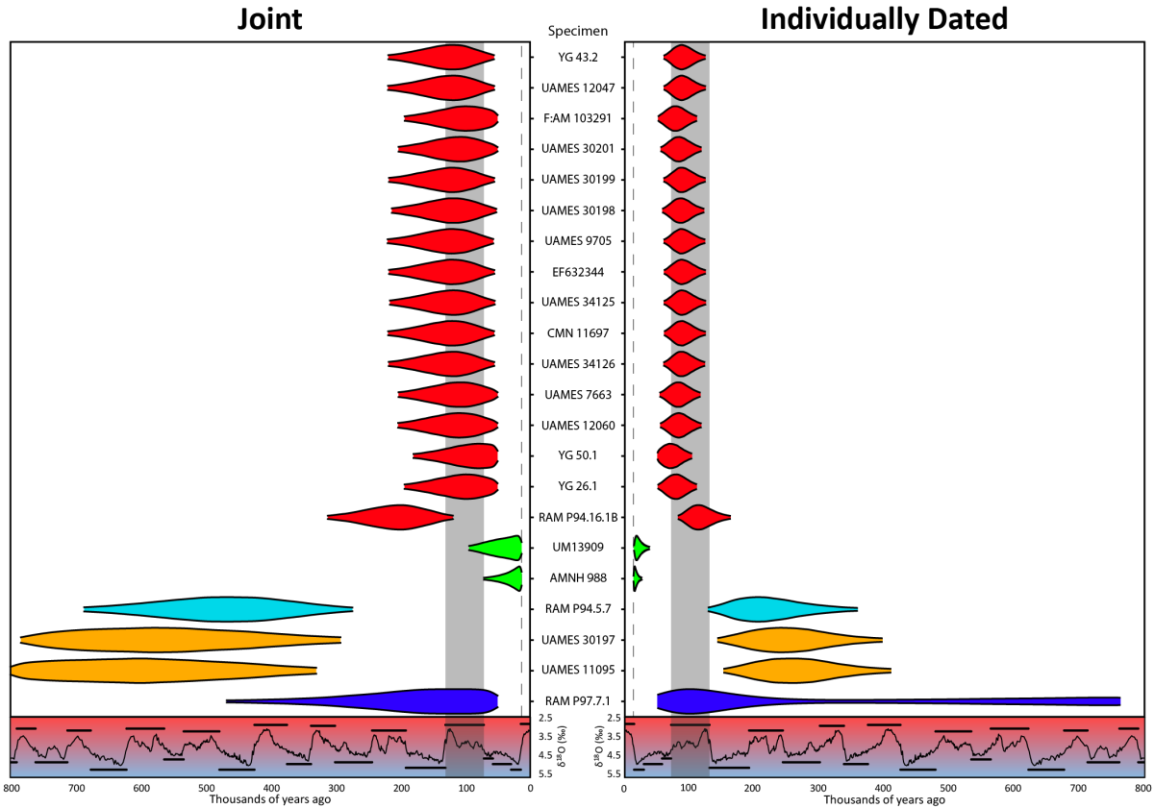


Figure 3: Marginal posterior densities of specimen ages in molecular dating analyses. Specimens are coloured by clade and arranged as in Fig. 1B. Each violin represents the 95% HPD interval of each specimen age estimated in the Joint (left) and Individually Dated (right) analyses. The $\delta^{18}O$ record for the last 800 thousand years is overlaid below both plots and corresponds to the one in Fig. 2A, with the approximate extent of the MIS 5 interglacial shaded in grey. The dashed grey line at 13 kya represents the lower bound imposed by the youngest specimen in the analyses.

East Beringian mastodons in Clade A had much older posterior age estimates than Beringian mastodons in Clade Y. Median ages for UAMES 11095 were 586 ky (JT 95% HPD interval: 329–800 ky) and 267 kya (ID 95% HPD interval: 152–410 ky), and median ages for UAMES 30197 were 558 ky (JT 95% HPD interval: 292–784 ky) and 254 ky (ID 95% HPD interval: 142–397 ky). However, the 95% HPD intervals of the ages of these specimens were much wider than those of specimens in Clade Y, spanning many more glacial and interglacial periods, and making it difficult to tie them to any specific marine isotope stage. Nevertheless, the ages of both specimens have 95% HPD intervals that do not overlap with those of east Beringian mastodons in Clade Y, which suggests that mastodons in these two clades are temporally distinct. Combined with their divergent positions in the phylogeny, these results are consistent with Clade A mastodons being part of a separate colonization event during an earlier interglaciation.

The Joint and Individually Dated analyses both estimate young ages for two mastodon individuals in Clade G. AMNH 988 has a median posterior age of 28 ky (JT 95% HPD interval: 13–71 ky) and 17 ky (ID 95% HPD interval: 13–27 ky), while UM13909 has median posterior ages of 43 ky (JT 95% HPD interval: 13–94 ky) and 21 ky (ID 95% HPD interval: 14–38 ky). In all analyses, however, the posterior age densities for these specimens have greater probability mass toward younger values and abut the lower bound at 13 kya, a hard limit based on the age of the youngest specimen in the dataset, INSM 71.3.261. These results are broadly consistent with the expected ages of these two specimens, given the radiometric and geological ages of other mastodons in this clade. Nevertheless, the shapes of the posterior age distributions suggest that these specimens might actually be younger than 13 ky.

Alberta mastodons were estimated at a range of ages, congruent with an interpretation of highly dynamic biogeographic landscape characterized by population turnover. However, we note that the limited number of specimens from this region, the wide 95% HPD intervals of their ages, and their scattered positions across the phylogeny make it difficult to tie them to specific periods. Specimen RAM P94.16.1B had a greater median posterior age (JT: 208 ky; ID: 117 ky) than other mastodons within its clade, although its 95% HPD interval overlapped those of other Clade Y mastodons (JT: 119–311 ky; ID: 82–163 ky). The widths of the 95% HPD intervals and their overlap also varied between the two analyses, and in their association with either the MIS 5 or MIS 7 interglacial. However, should this specimen ultimately be shown to date to MIS 7, it would suggest successive colonization events from the same or a similar source population.

The Alberta sample RAM P94.5.7 had a median posterior age estimate similar to those of mastodons in Clade A (JT: 474 ky; ID: 221 ky), but also with a wide 95% HPD interval that makes it difficult to associate with any particular marine isotope stage. However, unlike the separation between Beringian mastodons in Clade Y and Clade A,

the 95% HPD intervals of the ages of RAM P94.16.1B and RAM P94.5.7 do overlap by 37 ky (JT) and 35 ky (ID), making their separation more uncertain.

The posterior density of the age of sample RAM P97.7.1 had a mode within the MIS 5 interglacial age boundaries (Supplementary Table 42), but with a very wide 95% HPD interval (JT: 50–467 ky; ID: 50–763 ky). As with mastodons in Clades A and L, this pattern is likely to be due to its phylogenetic position and deep divergence from the majority of calibration points in the dataset.

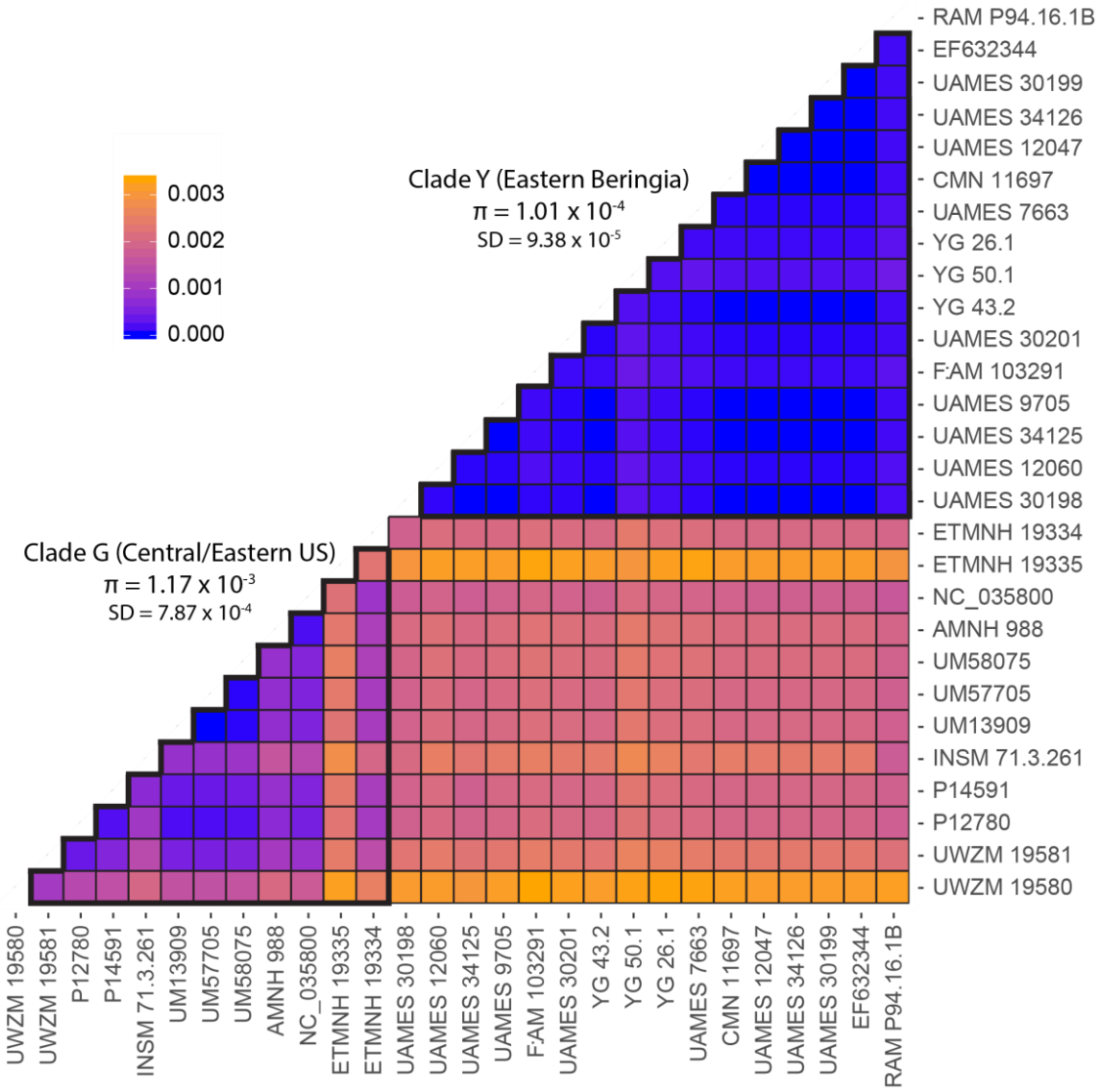


Figure 4: Pairwise genetic distance heatmap for all samples in Clades Y and G. Pairwise genetic distances for all samples in Clade G (including ETMNH 19335 and ETMNH 19334) and Clade Y (including RAM P94.16.1B). Genetic distances are calculated using the F84 model of nucleotide substitution. Nucleotide diversity within each clade (outlined with the thick black border) is indicated. All values are given in substitutions per site.

Genetic Diversity within Mastodon Clades

Under a model of repeated expansion and extirpation, northern clades of mastodons would be expected to have lower levels of genetic diversity. This pattern would be consistent with repeated expansion of small founder matriarchal herds in response to climatic warming during interglaciations, and the transient nature of their occupation of northern latitudes. Genetic diversity is expected to be higher among samples from regions south of the continental ice sheets that were likely to have been inhabited by populations of mastodons throughout the Pleistocene. We examined levels of nucleotide diversity within our dataset to test this hypothesis.

Clade Y had a low nucleotide diversity (π) of $\pi = 8.79\times10^{-5}$ substitutions per site (standard deviation (SD) = 8.9×10^{-5} substitutions per site) or 1.01×10^{-4} substitutions per site (SD = 9.38×10^{-5} substitutions per site) when including the potentially temporally distinct Alberta specimen (RAM P94.16.1B) (Fig. 4). The genetic distance between the two mastodons in clade A was also quite small (1.24×10^{-4} substitutions per site). By comparison, there were much higher levels of nucleotide diversity in clade G, which contains mastodons from south of the ice sheets, their endemic range. This was the case when either including ($\pi = 1.17\times10^{-3}$ substitutions per site; SD = 7.87×10^{-4} substitutions per site) or excluding ($\pi = 8.09\times10^{-4}$ substitutions per site; SD = 5.31×10^{-4} substitutions per site) the two mastodons from Virginia. This is consistent with expectations of small numbers of matrilocal mastodons expanding northward in response to glacial retreat, and also supports previous palaeoecological models of environments inhabited by northern mastodons.

Conclusions

Our sequencing and analysis of 33 mitochondrial genomes of the American mastodon, *Mammut americanum*, provides a framework for interpreting the genetic diversity of the species across space and time. We have identified six mitochondrial clades that span nearly the entire North American continent, from Alaska to Mexico. The two predominantly eastern Beringian clades are likely to have originated from separate expansions of mastodons into the region. We also note that the individuals from Alberta are distributed across the phylogenetic tree, highlighting the dynamic nature of American mastodon dispersal between southern and northern latitudes. Our proposed clade nomenclature may need revision as further mitochondrial genomes are sequenced, and as geographic and temporal gaps are filled. This applies especially to specimens in Clade M, which have a deep coalescence time and may stand as sole representatives of multiple lineages that are deeply divergent from the remaining mastodons.

Our analyses add further support for the model[20] that proposed that American mastodons only occupied higher latitudes (*i.e.*, Canada and Alaska) during interglacials, when prevailing warm climatic conditions supported the establishment of forests and wetlands. The presence of temporally distinct clades in Alaska and Yukon indicates that the inferred pattern of expansion during warm interglaciations, followed by local extirpations and range contraction to the south during colder times, was likely to have been a recurring scenario. We infer that this was a major, and perhaps widespread, biological response to global glacial-interglacial cycling that affected many species in eastern Beringia (*e.g.* western camel *Camelops hesternus*[21]; giant beaver *Castoroides ohioensis*[43]). Similar processes presumably occurred in Eurasia, with warm-adapted species such as hippopotamuses and hyaenas episodically expanding their ranges northward during earlier interglaciations into previously ice-dominated areas like the British Isles and Scandinavia[44–46]. However, this pattern also poses further questions: for example, why were species that had managed to repeatedly expand into the northernmost parts of North America during previous interglacials unable to do so following the return to interglacial conditions after the last glacial maximum (~21 kya)? Were they already in severe decline? More critically, will similar trends be seen in extant browsers?

At present, numerous bird[5], fish[4], and mammal communities[47,48] in northern North America are undergoing rapid restructuring in response to climatic warming. Moose[47] and beavers[48], iconic members of present-day northern boreal forest, have expanded their ranges northward by hundreds of kilometres in the last few decades alone. Our data suggest that regional expansion of at least some southern, temperate populations into northern latitudes is a probable outcome of the warmer and wetter conditions of today. However, populations at the expansion front are likely to be a subset of the current global diversity of these species, leaving them vulnerable if more genetically diverse southern populations are eventually lost. The phylogeographical history of Pleistocene megafauna can serve as a useful example for understanding the ecological responses of present-day species, and can generate testable hypotheses about the consequences of anthropogenic environmental impacts.

Methods

Sample Acquisition and Subsampling

Mastodons were subsampled at each of the institutions that contributed specimens to this study (Supplementary Data 1). Subsamples were then sent to the McMaster University Ancient DNA Centre, with all further processing conducted in dedicated ancient DNA clean rooms.

DNA Extraction and Processing

Specimens were processed using a variety of wet-lab methodologies in an attempt to maximize the probability of successful DNA extraction, and as new techniques were developed and modified to overcome issues with DNA recovery, inhibition, and poor endogenous preservation. Between 30 and 349.9 mg of material were demineralized and digested in successive rounds with 0.5 M EDTA or a Proteinase K digestion buffer. Supernatants were pooled and extracted using either organic or two different guanidinium-silica based extraction methods[49,50]. UDG-treated and non-UDG-treated libraries were prepared using double-stranded[51,52] or single-stranded methodologies[53,54], with some modifications from in-house optimization. Following indexing, all libraries underwent 1–2 rounds of in-solution enrichment with a comprehensive proboscidean bait set[33], to increase the relative abundance of the degraded endogenous fraction. Full methods for each sample are given in Supplementary Data 1 and the Supplementary Methods.

Sequence Mapping and Curation

Demultiplexed reads were trimmed and merged with leeHom[55], using the ancient DNA flag, and the double-stranded or single-stranded library adapter sequences as appropriate. Reads were then mapped against the *M. americanum* mitochondrial reference genome (NC_035800) using a network-aware version of BWA[56] (https://github.com/mpieva/network-aware-bwa) with common ancient DNA settings: maximum edit distance of 0.01 (-n 0.01), a maximum of two gap openings (-o 2), and seeding effectively disabled (-l 16500). Mapped reads that were either merged or properly paired were extracted using the retrieveMapped_single_and_ProperlyPair program of libbam (https://github.com/grenaud/libbam). Replication duplicates were then removed based on unique 5’ and 3’ positions (https://bitbucket.org/ustenzel/biohazard/src/master/), and filtered to remove sequences below a minimum length of 24 bp and mapping quality of 30. Specimens with multiple libraries were combined at this point, and underwent additional duplicate removal if they contained the same index pair (Supplementary Methods – Reference Guided Mapping).

Alignments were imported into Geneious v6.1.5 and manually curated to remove any sequencing artefacts. Regions below our requirement of 3× minimum depth coverage were masked with Ns. Due to stacking observed in conserved mitochondrial regions (*e.g.* 16S rRNA, D-loop), we further masked any positions with coverage depths greater than three standard deviations from the mean that displayed multi-allelic variants (Supplementary Methods – Sequence Curation). The variable number tandem repeat (VNTR) region of each sequence was masked with Ns in accordance with the

NC_035800 reference. Final consensus sequences were obtained using the majority base call at each position.

Model Selection and Phylogenetic Analyses

The 33 new mitochondrial genomes were aligned with the only two mastodon mitochondrial genomes previously published, MAS1 (NC_035800) and IK-99-237 (EF632344), using MUSCLE v3.8.31[57]. Model selection was performed using jModelTest v2.1.4[58] with the corrected Akaike information criterion. The HKY+G model was chosen for all subsequent analyses.

The phylogeny was inferred by maximum-likelihood analysis using IQ-TREE v1.6.6[59]. We used two approaches to root the tree, either by midpoint-rooting or by including an outgroup comprising two mammoth mitochondrial genomes (NC_007596 and NC_015529). In each case, node support was estimated using 1000 bootstrap replicates.

Both of the sequence alignments were also used for Bayesian phylogenetic inference in BEAST v.1.8.0[60] using an HKY+G4 model, a constant population size tree prior, a strict clock model, and default specifications for all priors.(Supplementary Methods – Phylogenetic Analyses). The analysis was run for 10 million steps sampling every 1000 states.

Sample-Age Estimation

We compared coalescent tree priors and clock models by calculating their marginal likelihoods using generalized stepping-stone sampling[61] (Supplementary Methods – Model Testing With GSS) in BEAST v.1.10.5[62]. We ran two separate age-estimation analyses. In the first approach, we jointly estimated the ages of all of the undated specimens in a single analysis (Joint; JT). We specified gamma prior distributions (shape = 1; scale = 200,000) for the ages of the undated samples. The ages of these samples were also bounded at 800 ky (the upper limit of successful ancient DNA recovery) and at 50 ky (the approximate limit of radiocarbon dating), with the exception of AMNH 988 and UM13909 which had a lower limit of 0 ky. In all cases, bounds were scaled relative to the age of the youngest specimen in the dataset (INSM 71.3.261 at 13 ky). Samples with known radiocarbon ages were calibrated using Calib v7.0.4[63], then the ages of these specimens were treated as point values based on their median ages.

In our second approach, we estimated the age of each specimen individually in an analysis that included the dated specimens. We then used the marginal posterior densities

of the ages of the individual samples to specify the priors for the ages of these samples in a combined analysis of all samples (Individually Dated; ID).

Both analyses were run with the HKY+G substitution model and empirical base frequencies. We chose a constant population size tree prior and a strict clock model, with uniform distributions for the associated population size prior (Uniform [1, 1×10^{6}]) and the substitution rate prior (Uniform [4×10^{-10}, 8×10^{-8}]), while all other remaining priors were left at their default distributions. To allow for more efficient sampling of unknown specimen ages, their weight was increased to 5. The chain length was increased to 500 million steps (sampling every 10,000), and all analyses were run in duplicate.

Nucleotide Diversity

Pairwise distances between all American mastodon mitochondrial genomes were calculated using the dist.dna() function in the R package ape[64]. Distances were calculated using the F84 model. Sites with missing data were deleted in a pairwise manner. For clades containing more than two mastodons, nucleotide diversity was calculated as the mean of all pairwise distances between specimens in that clade.

Data Availability

Mastodon specimens examined in this study were obtained from the external institutions listed in Supplementary Data 1. All requests for access to the material should be made to the external institution that houses the material.

Final consensus sequences for all complete mastodon specimens have been uploaded to NCBI with GenBank accessions MN616941–MN616973 (Supplementary Table 1). Raw sequencing reads for each complete mitochondrial genome were also uploaded to the SRA (BioProject: PRJNA578413). New radiocarbon dates are reported in Supplementary Table 44.

Two previously published American mastodon mitochondrial genomes (GenBank accessions NC_035800 and EF632344) were also analysed in this study. Two mammoth mitochondrial genomes (NC_007596 and NC_015529) were used in some phylogenetic analyses as outgroups.

Code Availability

No new software was generated during the course of this study. Custom scripts used to produce nucleotide diversity graphics are available are available at: https://github.com/ekarpinski/MastoScripts.

Acknowledgments

The authors profusely thank Christopher Beard (University of Kansas Museum of Natural History), Rodney Scheetz (Brigham Young University Museum of Paleontology), Deborah Skilliter (Nova Scotia Museum), Bruce J. MacFadden (Florida Museum of Natural History), Richard K. Stucky (Denver Museum of Nature and Science), Mary E. Thompson (Idaho Museum of Natural History), and all other people and institutions that donated samples to be analysed in this study. We extend a special thank you to the Yukon placer gold miners and the Vuntut Gwitch'in and Tr'ondëk Hwëtch'in First Nations communities for their continual support with the field collection of fossils in the Yukon. Thank you to Jacob Enk and Ben J. Evans, as well as all members of the Poinar, Golding, and Evans lab groups for their support in the direction and troubleshooting of various aspects of this project, as well as to Alberto Reyes for his suggestions and assistance with Pleistocene ice sheet reconstructions. Lastly, we extend our gratitude to Arbor Biosciences, who donated enrichment reagents and support for this project. This work was supported by an NSERC postgraduate scholarship (PGSD3-518942-2018) and the Dr. Larry Agenbroad Legacy Fund for Research (The Mammoth Site) to E.K; NSERC Discovery Grant to G.B.G. (grant No. 140221-10); and by an NSERC Discovery Grant (grant No. 4184-15), and CRC to H.P.

Conflict of Interest

Arbor Biosciences (current employer of J.K.) generously provided the in-solution baits and enrichment reagents for this study. All other authors declare no other competing interests.

Chapter 3 Bibliography

1. Collins, M. *et al. In Climate Change 2013—The Physical Science Basis* (ed. Intergovernmental Panel on Climate Change) 1029–1136 (Cambridge University Press, Cambridge, 2013).

2. Ackerly, D. D. *et al.* The geography of climate change: implications for conservation biogeography. *Divers. Distrib.* **16**, 476–487 (2010).

3. Bradshaw, W. E. & Holzapfel, C. M. Evolutionary Response to Rapid Climate Change. *Science* **312**, 1477–1478 (2006).

4. Chu, C., Mandrak, N. E. & Minns, C. K. Potential impacts of climate change on the distributions of several common and rare freshwater fishes in Canada. *Divers. Distrib.* **11**, 299–310 (2005).

5. Princé, K. & Zuckerberg, B. Climate change in our backyards: the reshuffling of North America's winter bird communities. *Glob. Chang. Biol.* **21**, 572–585 (2015).

6. Scheffers, B. R. *et al.* The broad footprint of climate change from genes to biomes to people. *Science* **354**, aaf7671 (2016).

7. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ 18 O records. *Paleoceanography* **20**, PA1003 (2005).

8. Dyke, A. S. An outline of the deglaciation of North America with emphasis on central and northern Canada. *Quat. Glaciat. Chronol. Part II North Am.* **2b**, 373-424 (2004).

9. Thompson, L. G. *et al.* Late Glacial Stage and Holocene Tropical Ice Core Records from Huascaran, Peru. *Science* **269**, 46–50 (1995).

10. Johnsen, S. J. *et al.* Oxygen isotope and palaeotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and NorthGRIP. *J. Quat. Sci.* **16**, 299–307 (2001).

11. Kawamura, K. *et al.* Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature* **448**, 912–916 (2007).

12. Dyke, A. S. Late Quaternary Vegetation History of Northern North America Based on Pollen, Macrofossil, and Faunal Remains. *Géographie Phys. Quat.* **59**, 211–262 (2005).

13. Froese, D. *et al.* Fossil and genomic evidence constrains the timing of bison arrival in North America. *Proc. Natl. Acad. Sci.* **114**, 3457–3462 (2017).

14. Palkopoulou, E. *et al.* Holarctic genetic structure and range dynamics in the woolly mammoth. *Proc. R. Soc. B Biol. Sci.* **280**, 20131910 (2013).

15. Debruyne, R. *et al.* Out of America: ancient DNA evidence for a new world origin of late quaternary woolly mammoths. *Curr. Biol.* **18**, 1320–1326 (2008).

16. Shapiro, B. *et al.* Rise and Fall of the Beringian Steppe Bison. *Science* **306**, 1561–1565 (2004).

17. Campos, P. F. *et al.* Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5675–5680 (2010).

18. Chang, D. *et al.* The evolutionary and phylogeographic history of woolly mammoths: a comprehensive mitogenomic analysis. *Sci. Rep.* **7**, 44585 (2017).

19. Heintzman, P. D. *et al.* Bison phylogeography constrains dispersal and viability of the Ice Free Corridor in western Canada. *Proc. Natl. Acad. Sci.* **113**, 8057–8063 (2016).

20. Zazula, G. D. *et al.* American mastodon extirpation in the Arctic and Subarctic predates human colonization and terminal Pleistocene climate change. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18460–18465 (2014).

21. Zazula, G. D. *et al.* A case of early Wisconsinan "over-chill": New radiocarbon evidence for early extirpation of western camel (*Camelops hesternus*) in eastern Beringia. *Quat. Sci. Rev.* **171**, 48–57 (2017).

22. Saunders, J. J. *et al.* Paradigms and proboscideans in the southern Great Lakes region, USA. *Quat. Int.* **217**, 175–187 (2010).

23. Oltz, D. F. & Kapp, R. O. Plant Remains Associated with Mastodon and Mammoth Remains in Central Michigan. *Am. Midl. Nat.* **70**, 339–346 (1963).

24. Dreimanis, A. Extinction of Mastodons in Eastern North America : Testing a New Climatic- Environmental Hypothesis. *Ohio J. Sci.* **68**, 257–272 (1968).

25. Shoshani, J. Understanding proboscidean evolution: a formidable task. *Trends Ecol. Evol.* **13**, 480–487 (1998).

26. Teale, C. L. & Miller, N. G. Mastodon herbivory in mid-latitude late-Pleistocene boreal forests of eastern North America. *Quat. Res.* **78**, 72–81 (2012).

27. Green, J. L., DeSantis, L. R. G. & Smith, G. J. Regional variation in the browsing diet of Pleistocene *Mammut americanum* (Mammalia, Proboscidea) as recorded by dental microwear textures. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **487**, 59–70 (2017).

28. Birks, H. H. *et al.* Evidence for the diet and habitat of two late Pleistocene mastodons from the Midwest, USA. *Quat. Res.* **91**, 792–812 (2019).

29. Owen-Smith, N. Pleistocene extinctions: the pivotal role of megaherbivores. *Paleobiology* **13**, 351–362 (1987).

30. Barnosky, A. D. *et al.* Variable impact of late-Quaternary megafaunal extinction in causing ecological state shifts in North and South America. *Proc. Natl. Acad. Sci.* **113**, 856–861 (2016).

31. Widga, C. *et al.* Late Pleistocene proboscidean population dynamics in the North American Midcontinent. *Boreas* **46**, 772–782 (2017).

32. Godfrey-Smith, D., Grist, A. & Stea, R. Dosimetric and radiocarbon chronology of a pre-Wisconsinan mastodon fossil locality at East Milford, Nova Scotia, Canada. *Quat. Sci. Rev*. **22**, 1353–1360 (2003).

33. Enk, J. *et al.* Mammuthus population dynamics in Late Pleistocene North America: Divergence, Phylogeogrpaphy and Introgression. *Front. Ecol. Evol*. **4**, 1–13 (2016).

34. Ishida, Y., Georgiadis, N. J., Hondo, T. & Roca, A. L. Triangulating the provenance of African elephants using mitochondrial DNA. *Evol. Appl*. **6**, 253–265 (2013).

35. Fernando, P., Pfrender, M. E., Encalada, S. E. & Lande, R. Mitochondrial DNA variation, phylogeography and population structure of the Asian elephant. *Heredity* **84**, 362–372 (2000).

36. Fisher, D. Extinction of Proboscideans in North America. *in The Proboscidea: Evolution and Paleoecology of Elephants and their Relatives* (eds. Shoshani, J. & Tassy, P.) 296–315 (Oxford University Press, 1996).

37. Fisher, D. C. Paleobiology of Pleistocene Proboscideans. *Annu. Rev. Earth Planet. Sci*. **46**, 229–260 (2018).

38. Rohland, N. *et al.* Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. *PLoS Biol*. **8**, 1–10 (2010).

39. Muhs, D. R., Ager, T. A. & Begét, J. E. Vegetation and paleoclimate of the last interglacial period, central Alaska. *Quat. Sci. Rev*. **20**, 41–61 (2001).

40. Jass, C. N. & Barrón-Ortiz, C. I. A review of Quaternary proboscideans from Alberta, Canada. *Quat. Int*. **443**, 88–104 (2017).

41. Shapiro, B. *et al.* A Bayesian Phylogenetic Method to Estimate Unknown Sequence Ages. *Mol. Biol. Evol*. **28**, 879–887 (2011).

42. Drummond, A. J. & Stadler, T. Bayesian phylogenetic estimation of fossil ages. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150129 (2016).

43. Plint, T., Longstaffe, F. J. & Zazula, G. Giant beaver palaeoecology inferred from stable isotopes. *Sci. Rep*. **9**, 7179 (2019).

44. Yalden, D. W. Pleistocene Mammals. in *The history of British mammals* 12–27 (T & A D Poyser Ltd, 1999).

45. Schreve, D. C. A new record of Pleistocene hippopotamus from River Severn terrace deposits, Gloucester, UK—palaeoenvironmental setting and stratigraphical significance. *Proc. Geol. Assoc*. **120**, 58–64 (2009).

46. Stoffel, C. *et al.* Genetic consequences of population expansions and contractions in the common hippopotamus (Hippopotamus amphibius) since the Late Pleistocene. *Mol. Ecol*. **24**, 2507–2520 (2015).

47. Tape, K. D., Gustine, D. D., Ruess, R. W., Adams, L. G. & Clark, J. A. Range Expansion of Moose in Arctic Alaska Linked to Warming and Increased Shrub Habitat. *PLoS One* **11**, e0152636 (2016).

48. Tape, K. D., Jones, B. M., Arp, C. D., Nitze, I. & Grosse, G. Tundra be dammed: Beaver colonization of the Arctic. *Glob. Chang. Biol*. **24**, 4478–4488 (2018).

49. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15758–15763 (2013).

50. Glocke, I. & Meyer, M. Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome Res*. **27**, 1–8 (2017).

51. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res*. **40**, 1–8 (2012).

52. Meyer, M. & Kircher, M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb. Protoc*. **2010**, 1–10 (2010).

53. Gansauge, M.-T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc*. **8**, 737–748 (2013).

54. Gansauge, M.-T. *et al.* Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, 1–10 (2017).

55. Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res*. **42,** e141–e141 (2014).

56. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–60 (2009).

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

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

59. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

60. Drummond, A. J., Suchard, M. a, Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–73 (2012).

61. Baele, G., Lemey, P. & Suchard, M. A. Genealogical Working Distributions for Bayesian Model Testing with Phylogenetic Uncertainty. *Syst. Biol.* **65**, 250–264 (2016).

62. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, 1–5 (2018).

63. Stuiver, M. & Reimer, P. J. Extended 14C database and revised CALIB radiocarbon calibration program. *Radiocarbon* **35**, 215–230 (1993).

64. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).

65. Colleoni, F., Wekerle, C., Näslund, J.-O., Brandefelt, J. & Masina, S. Constraint on the penultimate glacial maximum Northern Hemisphere ice topography (≈140 kyrs BP). *Quat. Sci. Rev.* **137**, 97–112 (2016).

CHAPTER 4

Mastodon mitochondrial genomes from American Falls, Idaho

Emil Karpinski[1,2*], Chris Widga[3], Andrew Boehm[4], Brandon R. Peecook[5,6], Melanie Kuch[1], Tyler Murchie[1,7], Hendrik N. Poinar[1,7,8*.]

[1] McMaster Ancient DNA Centre, Departments of Anthropology and Biochemistry, McMaster University, Hamilton, ON, L8S 4L9, Canada.
[2] Department of Biology, McMaster University, Hamilton, ON, L8S 4L8, Canada.
[3] Center of Excellence in Paleontology and Department of Geosciences, East Tennessee State University, Johnson City, TN, 37614, USA.
[4] University of Oregon Museum of Natural and Cultural History, University of Oregon, Eugene, OR, 97403.
[5] Idaho Museum of Natural History, Idaho State University, Pocatello, ID 83201.
[6] Department of Biological Sciences, Idaho State University, Pocatello, ID 83209.
[7] Department of Anthropology, McMaster University, Hamilton, ON, L8S 4L9, Canada.
[8] Department of Biochemistry, McMaster University, Hamilton, ON, L8S 4L8, Canada.

This work is in preparation for publication submission. The main text of the study is presented within this chapter. Supplementary figures and data referred to in this chapter are available in Appendix A.

Author Contributions

EK and HP conceived the study with feedback from CW. EK, MK, and TM conducted all genetic and EK completed all bioformatic analyses. BRP conducted morphological examination of mastodon mandibles. EK, CW, AB, and BRP interpreted the results. EK wrote the first draft of the manuscript. All authors helped in the revision of subsequent drafts.

Abstract

Palaeogenetics offers a powerful method to examine the evolutionary relationships between and within taxonomic groups. With sufficient data these analyses can be extended to provide information about the palaeodemography of extinct species. Recent mitochondrial genomes from American mastodons (*Mammut americanum*) have highlighted the roles glacial/interglacial cycles had on their range dynamics, but was geographically limited to a subset of American mastodon populations, with many lacking adequate representation. Here we extend previous mastodon ancient DNA analyses with two new mitochondrial genomes from the American Falls Reservoir in Idaho. Our phylogenetic analyses clearly place mastodons from this region as close relatives to other interglacial mastodons from eastern Beringia, and may represent a southward migration in response to glacial advance. We also attempt to contextualize our findings given what is currently assumed about the distribution of American and Pacific mastodons, and argue that a much more comprehensive examination of morphological variation within *Mammut* is likely required.

Introduction

North America was inhabited by two groups of morphologically distinct mastodons in the mid to late Pleistocene: the American mastodon (*Mammut americanum*) which begin to appear in the fossil record by at least 3.75 million years ago[1]; and the more recently described Pacific mastodon (*Mammut pacificus*)[2], known from Irvingtonian (~1.35 my – 195 ky) strata in California, Idaho, and Montana, and persisting into the Rancholabrean (~195 to 135 ky – 11 ky) in California and Idaho[2,3]. The delineation of these two *Mammut* species is based on five skeletal and dental traits, though most work on *M. pacificus* has largely focused only on specimens containing intact second and third molars[2,3].

Fluvial sediments of the lower American Falls formation are exposed around the margins of the American Falls Reservoir (AFR) of the Snake River in southeastern Idaho and have produced well over 6,000 late Pleistocene vertebrate fossils pertaining to a minimum of 51 taxa[4,5]. Amongst the medium to large mammals, the American Falls local fauna includes many taxa still present in the region today (*Ursus*, *Canis*, *Puma*, *Lynx*, *Ovis*, *Antilocapra*, *Odocoileus*, *Cervus*, *Rangifer*, etc.) alongside charismatic Rancholabrean megafauna: ground sloths (*Megalonyx*, *Paramylodon*), carnivorans (*Aenocyon*, *Arctodus*, *Smilodon*, *Homotherium*, *Panthera atrox*), a horse (*Equus scotti*), camelids (*Camelops*, *Hemiauchenia*), a peccary (*Platygonus*), a muskoxen (*Bootherium*), bison (*Bison latifrons*, *B. priscus*, and *B. alaskensis*), and proboscideans (*Mammut*, *Mammuthus*)[4,6]. The lower American Falls formation is bound in terms of absolute ages by the unconformably underlying Crystal Springs Basalt (K/Ar date of 210,000 years BP[7]) and the Cedar Butte basalt (K/Ar date of 72,000 years BP[8]), which dammed the Snake River. However, the lower American Falls formation is argued to preserve a fauna constrained to the climatically mild MIS 5 interglacial stage (125,000-75,000 years BP) due to the presence of *Panthera atrox* and the herpetofauna, and the lack of tundra adapted rodents (*e.g. Dicrostonyx*)[4]. Unpublished radiometric dates of several lower American Falls fossils have yielded ages of approximately 100,000 years BP, with one mammoth tooth at 125,000 years BP (Peecook, B.R., personal communication, June 10th 2021).

The thinner upper American Falls formation consists of lacustrine sediments much less extensively preserved around the AFR, but contains a similar fauna. Samples of peat from the bottom of the upper American Falls formation have been dated to 51,900 to 42,000 years BP (Peecook, B.R., personal communication, June 10th 2021), while unpublished radiometric dates of wood, bone, and enamel excavated from nearby landfills and potato cellars on the Snake River Plain yielded dates of 55,500 to 44,600 years BP. An in situ *Equus* incisor from the highest locality around AFR and within the American Falls formation (Duck Point) has been dated to 20,700 years BP (Peecook, B.R., personal communication, June 10th 2021).

American Falls has also played an important role in the establishment of the Pacific mastodon, being one of the first locations outside of California where a specimen of *M. pacificus* was described (USNM 13701[2]). Two additional *M. pacificus* molars have also been described from nearby Bingham county (<100 km from American Falls) from Irvingtonian strata[2].

Ancient DNA provides a powerful dimension where evolutionary relationships, both within species[9–11] as well as between species[12,13] can be interrogated with temporal depth. It is particularly useful when the full extent of morphological variation within a species is unknown, or in cases of extensive convergent evolution[9,14]. Recent aDNA work on American mastodons has begun to elucidate their mitochondrial diversity throughout the mid-to-late Pleistocene, and highlighted the important role glacial/interglacial cycles played on biogeography[15]. It also shows that while mastodons exhibit a high degree of female philopatry, certain clades encompass animals across a very wide range, and some regions display a large amount of heterogeneity in mastodon matrilines[15].

Here we describe our attempts to extract and analyze mitochondrial DNA from eight mastodon specimens found in the Lower American Falls formation. Our findings suggest that mastodons from American Falls are closely related to those that expanded into eastern Beringia likely during the previous interglacial (Marine Isotope Stage (MIS) 5). Furthermore, their phylogenetic and geographic position have interesting implications for understanding the palaeogeographic distribution and interactions between American and Pacific mastodons.

Results

Two of the eight mastodon specimens yielded sufficient data to reconstruct complete mitochondrial genomes: IMNH 50003/16643 and IMNH 52002/25939 (Table 1). DNA preservation within each specimen was variable, but in particular in IMNH 52002/25939 where the amount of recovered endogenous reads varied by over three orders of magnitude among generated libraries (Supplementary Table 1). Mean DNA fragment insert lengths for both specimens were short, and exhibited the expected increasing trend in cytosine deamination damage towards the 5’ and 3’ termini of the reads (Table 1; Supplementary Figures 2 and 3).

Extraction and library blanks contained no to very few reads that mapped to the *Mammut americanum* mitochondrial reference, suggesting minimal external contamination (Supplementary Table 1). Despite this we did notice a significant spike in coverage within the 16S rRNA region within specimen IMNH 50003/16643, containing reads primarily of bacterial origin (Supplementary Figures 4 and 5). To prevent any polymorphisms from this nonendogenous signal influencing phylogenetic reconstruction, we masked a 35 bp portion of this region in the IMNH 50003/16643 consensus.

Table 1: Mapping and coverage statistics. Sample information for the two IMNH for specimens which returned complete mitochondrial genomes.

Specimen	# Mapped Reads	Mean Frag. Size	Coverage	Age	Age Type
IMNH 50003/16643	11,372	39.7	99.7 % / 27.4 ×	-	-
IMNH 52002/25939	29,129	40.4	99.8 % / 71.5 ×	26500 +/- 3500	14C

Maximum likelihood phylogenies recover a well-supported (100% bootstrap support) monophyletic relationship for IMNH 50003/16643 and IMNH 52002/25939 and other mastodons in clade Y, a clade primarily composed of east Beringian *M. americanum* likely dating to around MIS 5 [15]. This pattern is consistent regardless of the inclusion of outgroups or substitution model chosen for the analysis, and is recovered with very high bootstrap support in every scenario (Supplementary Figures 6-11). Ultrametric phylogenies constructed with BEAST v1.10.5pre[16], recover a similar pattern, although place both Idaho mastodons as basal to other specimens within Clade Y (Supplementary Figures 12 and 13). This result is likely due to the large temporal range of samples across the tree and within this clade. When including prior information on estimated ages for each mastodon this pattern disappears (Supplementary Figures 15 and 17).

We also attempted to estimate the age of IMNH 50003/16643 in a molecular dating framework under two models: one in which the median radiocarbon age of IMNH 52002/25939 was used for calibration, and one in which both Idaho specimens were fit with broad priors spanning approximately the last 800 thousand years. When using IMNH 52002/25939 as a calibration point, the age of IMNH 50003/16643 produced a corrected median estimate of 24 ky (95% highest posterior density (95% HPD): 13–51 ky) (Supplementary Table 3). This estimated age is very close to the median calibrated radiocarbon age of IMNH 52002/25939 (30,810 y BP; 2σ range: 23365 – 38635), unsurprisingly given the low genetic distance between the two specimens (pairwise distance = 0.000122 substitutions/site; ~2 substations total). When the ages of both specimens are instead fit with broad priors, the estimated ages for both mastodons produce older estimates (IMNH 50003/16643 median: 46 ky; IMNH 52002/25939 median: 66 ky) (Supplementary Table 4). However the 95% HPD intervals for IMNH 50003/16643 (95% HPD: 13 – 115 ky) overlaps between both analyses.

Regardless of the analysis, estimates for IMNH 52002/25939 and IMNH 50003/16643 produce younger median age estimates and have much more of their

posterior probability density concentrated around younger ages than all other mastodons within clade Y. However, the 95% HPD intervals do overlap by ~ 1.5 ky with those other clade Y mastodons when IMNH 52002/25939 is used to calibrate the analysis (Supplementary Figures 14 and 16).

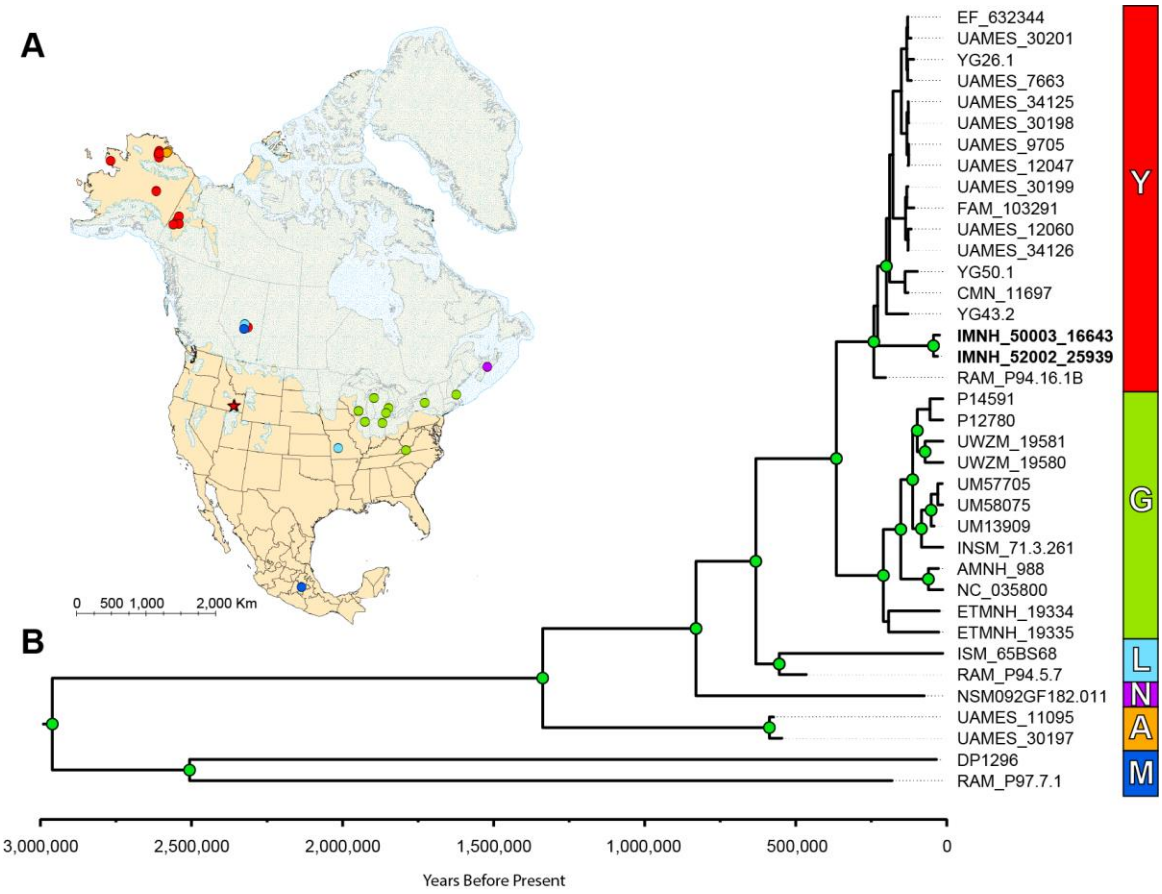


Figure 1: Phylogeny of IMNH mastodons. a, Map showing the approximate location of specimens analyzed in this study. Specimen locations have been jittered to facilitate better visualization. Specimens are coloured based on their clade assignments. The positions of the two complete specimens from Idaho are indicated with stars. **b**, Tip-calibrated MCC tree with IMNH 52002/25939 used as a calibration point. New specimens generated as part of this study are indicated in bold. Nodes with posterior probability support greater than 0.9 are indicated with light-green circles.

Discussion

The younger age estimates, coupled with their position on the tree, and their geographic origins (*i.e.* south of the glacial extent), have interesting implications for understanding the responses of mastodons to glacial/interglacial cycling during the Pleistocene. Low posterior probabilities prevent intraspecific relationships between mastodons within this clade to be resolved, however strong (posterior probability for clade monophyly = 1) support for their position within this clade allows us to propose two

possible explanations for the phylogenetic position. First, they could be descendants of a lineage that migrated along the Rockies from Alberta to eastern Beringia during the last interglacial approximately 130 kya. Under this scenario, these specimens should fall basal to all other mastodons within clade Y, as the migrating population would have branched off prior to the remainder moving further north in response to deglaciation. This would suggest a founder event that led to a long-term, highly stable population that occupied this region for approximately 100 ky. Alternatively, these animals could represent a southward migration from northern populations as North America transitioned from warmer interglacial conditions of MIS 5 to glacial conditions in MIS 4-2 as previously suggested[15]. While detailed pre-LGM environmental reconstructions would likely be needed to fully interrogate this scenario, this pattern appears to be consistent with glacial positions at the time[17].

The relationships between the Idaho and other clade Y mastodons has interesting implications for better contextualizing the relationships between American and Pacific mastodons, in terms of their systematics and biogeography. Following its identification and description in 2019, *M. pacificus* has been identified in California, Idaho, and Montana and has been argued to be likely present in Oregon[2,3]. Furthermore, the habitation of *M. pacificus* in these regions was argued to have been stable and exclusive, having occupied these localities for large parts of the mid and late Pleistocene. The analysis presented here suggests at least a portion of these trends are incorrect, and instead highlight that the relationships and distributions of Pacific and American mastodons were likely much more complex.

While the degraded nature of the molars from our two Idaho mastodons prevent us from concretely assigning them to either species of mastodon, the genetics clearly show that these animals fall well within the currently described mitochondrial diversity of American mastodons. This finding has two important implications depending on whether we accept the designation of Idaho mastodons as belonging to *M. pacificus*.

If these and other contemporary mastodons from this region are identified as *M. pacificus*, our results would suggest there is little support, at least at the mitochondrial level, for a species-level delineation of the Pacific mastodon. Instead the patterns of morphological variation we see between these two groups, are likely to be morphotypes with localized adaptations to unique ecological niches. Similar patterns have recently been described in other Pleistocene taxa such as *Bison latifrons*, *B. priscus*, and *B. bison*, which exhibit tremendous morphological variation despite being conspecific based on mitochondrial phylogenies[9].

If instead we accept that these remains belong to members of *M. americanum*, this would suggest the spatial and temporal co-occurrence of both mastodons in Idaho during the Rancholabrean. This pattern contradicts the previously assumed distributions of *M. pacificus* and *M. americanum*, and also produces additional questions as to how these

species may have interacted, including whether they may have interbred, something our mitochondrial phylogeny cannot specifically address.

To further examine this possibility, we also looked at the presence of mandibular tusks in other mastodon mandibles from American Falls. While this character is highly variable across Mammutidae and has also been attributed to sex[18,19] or chronological age[20,21], mandibular tusks are absent in all specimens identified as *M. pacificus* and the lack of them is a stated character difference from *M. americanum* in Dooley *et al.*[2]. Of the *Mammut* mandibles from American Falls the trait is variable: USNM 13701 (listed in Dooley *et al.*[2]), IMNH 52002/25817, and 69002/23158 indeed lack mandibular tusks, but in IMNH 35009/18144 and 71004/27988 mandibular tusks are present. Regrettably, the IMNH mandibles are lacking diagnostic m3 characters to identify them as *M. pacificus* or not. Assuming this trait is actually fully penetrant, the co-occurrence of tusked and tuskless mandibles amongst coeval American Falls mastodons is consistent with the idea that both species of mastodons were present in the region during the late Pleistocene.

Conclusions

The results we present here expand our understanding of the genetic landscape of mastodons in Pleistocene North America, and have important implications for understating their biogeography and systematics. Our data supports published models which suggest that glacial/interglacial cycles played a key role in the distribution of mastodons and that these groups experienced high mobility in response to these cycles[15]. Our findings also question the true extent of the distribution and relationships between Pacific and American mastodons, suggesting that either these two groups are conspecifics exhibiting localized variation, or co-occurred within parts of North America. While there still remains a fair bit of uncertainty in how these results should be interpreted, they highlight that a more comprehensive look at mastodon morphological variation to ascertain the validity and consistency of described characters, as well as assumptions about the distribution of mastodon taxa are drastically required.

Methods

Sample Acquisition and Subsampling

Subsamples of each specimen were originally taken at the Idaho Museum of Natural history, then sent to the McMaster University Ancient DNA Centre for further processing.

DNA Extraction and Processing

Between 200 and 500 mg of material was sampled from each specimens' tooth and manually pulverized. Subsamples were washed with 300 µl 0.5 M EDTA for 20 minutes at 1000 RPM to remove dust, exogenous contaminants, and clean the outside of the subsampled remains. Subsampled were demineralized using 0.75-1 ml of 0.5M EDTA, incubating at room temperature for 24-92 hours with shaking (1000-2000 RPM). Digestion used 0.75-1 ml of a Proteinase K based digestion buffer (0.01 M Tris-CL (pH 9); 0.20% Sarcosyl; 0.25 ml/ml Proteinase K; 0.01 M $CaCl_2$), incubating at 45°C with either shaking (500 RPM) or rotation to keep tooth material suspended. Demineralization and digestion was done in an alternating fashion for a total of five rounds (*i.e.* round one demineralization, round one digestion, round 2 demineralization; etc.), and pooled for extraction. An extraction blank containing no material was treated in parallel and used to monitor for contamination throughout.

DNA was extracted using a 5 M guanidinium HCl buffer as described in Dabney *et al.*[22] with the following modifications: 1 ml of supernantant was combined with binding buffer and successively passed over the column to bind all the demineralization/digestion supernatant; DNA was eluted twice with 25 µl of EBT (Buffer EB + 0.05% Tween-20).

Extracted DNA was used to construct non-UDG-treated double-stranded libraries[23,24] using 10-20 µl of DNA as input and with the following modifications: NE Buffer 2 was replaced with NE Buffer 2.1 in the end-repair reaction; volumes of all reactions were changed to 40 µl; MinElute clean-ups between reactions included two wash steps with 750 µl PE Buffer, and an additional 60 second dry spin to remove residual ethanol; libraries were heat-deactivated at 80°C for 20 minutes following adapter fill-in instead of purified over MinElute columns. 12.5 µl of each library was combined with unique P5 and P7 adapters in 40 µl indexing reactions (1X KAPA SYBR Fast qPCR Master Mix; 750 nM each P5/P7 adapter). Indexing reactions were carried out for a maximum of 20 cycles (denaturing at 95°C for 30 seconds; annealing at 60°C for 45 seconds), although specimens were monitored and removed at earlier points if observed to have entered the exponential phase. Multiple libraries were prepared from some specimens which were suspected of having higher DNA preservation (Supplementary Table 1). Indexing reactions were purified over MinElute columns using the same modifications as during library preparation, and eluted in 13 µl EBT.

To increase the recovery of endogenous DNA, indexed libraries were also subject to an in-solution enrichment using a proboscidean mitochondrial bait set[25]. Enrichment was carried out according to the myBaits v5 High Sensitivity protocol for one round using 9.05 µl of indexed library as input with the following modifications: custom xGen oligos were used during the Blockers Mix setup instead of Block X; beads were resuspended in 18.8 µl of EBT and used directly as input for reamplification (40 µl final reaction; 1X

KAPA SYBR Fast qPCR Master Mix; 150 nM each forward/reverse primer). Amplified post-capture libraries were purified over MinElute columns with the modifications described previously and eluted in 10.5 µl of EBT.

Enriched libraries were pooled to approximately equimolar concentrations and size selected for fragments between ~150-500 bp (3% Nusieve GTG Agarose Gel; 100 V for 35 minutes). Gel plugs were purified using QIAquick Gel Extraction Kit with the following modifications: two wash steps with 700 µl PE Buffer; an additional dry spin at max speed to remove residual ethanol; elution in 20 µl with Buffer EB. Libraries were sequenced on an Illumina HiSeq 1500 using 2x90 bp read chemistry.

Sequence Mapping and Curation

Demultiplexed reads were trimmed and merged using the ancient DNA settings in leeHom[26]. A modified *M. americanum* mitochondrial reference genome (NC_035800) was constructed by appending ~20 bp from the start/end of the reference to the opposite end to facilitate better recovery towards the tails of the reference. Reads were mapped to this modified reference using a network-aware version of BWA[27] (https://github.com/mpieva/network-aware-bwa) with common ancient DNA settings: maximum edit distance of 0.01 (-n 0.01), a maximum of two gap openings (-o 2), and seeding effectively disabled (-l 16500). Reads which mapped and were either merged or properly paired were extracted using the retrieveMapped_single_and_ProperlyPair program of libbam (https://github.com/grenaud/libbam), and collapsed based on unique 5' and 3' positions to remove PCR duplicates (https://bitbucket.org/ustenzel/biohazard/src/master/). Reads were further filtered to a minimum mapping quality of 30 and a minimum length of 24 bp using SAMtools[28]. Specimens for which multiple libraries were sequenced were combined into a single alignment. Sequencing data from Karpinski *et al.*[15] was also included for specimens which returned partial mitochondrial genomes (IMNH 50003/16643 and 52002/25939).

Alignments were imported in Geneious v11.0.9 and manually curated to remove sequencing artefacts and insertions not supported by a majority of reads. Positions with less than 3× coverage were masked with Ns. For all other positions the consensus was determined by a strict majority. We also masked a 35 bp fragment of the 16S rRNA region within the IMNH 50003/16643 alignment, as this region had a coverage depth greater than 10 standard deviations from the mean, and contained a large number of stacked reads, likely of bacterial origin (Supplementary Figures 4 and 5). The variable number tandem repeat (VNTR) region of each consensus sequence was also masked as in the *M. americanum* mitochondrial reference (NC_035800). The 20 bp regions appended to the start and end of the original reference were also removed.

To ensure the data was consistent with expectations for authentic ancient DNA libraries, we also examined cytosine deamination rates towards the termini of our molecules using MapDamage2.0[29].

Model Selection and Phylogenetic Analyses

Consensus sequences for IMNH 50003/16643 and 52002/25939 were aligned against all other complete *Mammut americanum* mitochondrial genomes currently available (n=35), and also separately including two mammoth mitochondrial genomes (NC_007596 and NC_015529) as outgroups. Substitution models were selected using the ModelFinder function of IQ-TREE v1.6.6[30,31] selecting the model that produces the smallest AICc estimate for the outgroup containing (TIM3+F+I+G4) and no-outgroup (TPM3u+F+I) datasets. Maximum likelihood phylogenies were subsequently generated for both datasets using the selected model for each, as well as the model other datasets model, and an HKY+F+G4 model used during BEAST analysis. All phylogenies were run with 1000 bootstrap replicates to assess branch support. Both datasets (with and without outgroups) were also investigated in BEAST v.10.5pre[16]. Chains were run using an HKY+F+G4 model and with a constant population size tree prior, for a total of 10 million generations (sampling every 1000). Three independent chains were run to monitor for convergence, and combined for analysis.

We also attempted to estimate the age of the Idaho mastodons, as well as their divergence from other mastodons in Clade Y. The dataset was analyzed as per the Joint analysis in Karpinski *et al.*[15], except that a radiocarbon date for ISNM 71.3.26 (11,450 ± 110 (Lab ID: AA100650)[32]) was calibrated using Calib v8.2[33] and the median age (13,326 ky) was used to calibrate the analysis instead of an estimate of 13 ky. Specimen IMNH 50003/16643 was fit with a broad gamma distribution like other possibly finite mastodons in the tree. Specimen IMNH 52002/25939 was treated two ways: in the first we calibrated the radiocarbon age and used the median as a point estimate to help calibrate the model; in the second we fit it with an identical prior as 50003/16643. Three independent chains were run for 500 million generations (sampling every 10,000) to monitor for divergence, and combined for the final analysis. Sample XMLs for both runs can be found in the supplementary online materials.

Distance Measurements

Pairwise distance and counts for mastodons were calculated using the dist.dna() function of the ape package in R[34]. Distances and counts were estimated using raw and N models respectively, with ambiguous sites removed in a pairwise manner.

Data Availability

Mastodon specimens examined in this study were obtained from Idaho Museum of Natural History. All requests for access to the material should be made to them.

Final consensus sequences for both complete mastodon mitochondrial genomes will be uploaded to GenBank and the SRA upon manuscript acceptance.

Acknowledgments

The authors would like to thank the Idaho Museum of Natural History and the Department of the Interior Bureau of Reclamation for access to the specimens, as well as to Mary E. Thompson who originally sent these samples for analysis. We would also like to thank Brian Golding for computional resources used in the analysis of this data, as well as G. Brian Golding, Ben J. Evans, and the members of the Poinar, Golding, and Evans lab groups for support and feedback on various aspects of this project. This work was supported by an NSERC postgraduate scholarship (PGSD3-518942-2018) E.K; and by an NSERC Discovery Grant (grant No. 4184-15), and CRC to H.P.

Chapter 4 Bibliography

1. Saunders, J. J. North American Mammutidae. *in The Proboscidea - Evolution and Paleoecology of Elephants and their Relatives* (eds. Shoshani, J. & Tassy, P.) 271–279 (Oxford University Press, 1998).

2. Dooley, A. C. *et al. Mammut pacificus* sp. nov., a newly recognized species of mastodon from the Pleistocene of western North America. *PeerJ* **7**, e6614 (2019).

3. McDonald, A. T., Atwater, A. L., Dooley Jr, A. C. & Hohman, C. J. H. The easternmost occurrence of *Mammut pacificus* (Proboscidea: Mammutidae), based on a partial skull from eastern Montana, USA. *PeerJ* **8**, e10030 (2020).

4. Pinsof, J. D. The American falls local fauna: late Pleistocene (Sangamonian) vertebrates from southeastern Idaho. whereas... *Pap. Vertebr. Paleontol. Idaho Honor. John A. White* **1**, 121–145 (1998).

5. Stratton, A. E., Tapanila, L. & Stratton. Collecting history of vertebrate fossils at American Falls, Idaho: A reservoir of data to inform land-use policy. *Palaios* **29**, 393–400 (2014).

6. Pinsof, J. D. *The Late Pleistocene vertebrate fauna from the American Falls area, southeastern Idaho*. (Idaho State University, 1992).

7. Desborough, G. A., Raymond, W. H., Marvin, R. F. & Kellogg, K. S. *Pleistocene sediments and basalts along the Snake River in the area between Blackfoot and Eagle Rock, southeastern Snake River Plain, southeastern Idaho.* (1989).

8. Scott, W. E., Pierce, K. L., Bradbury, J. P. & Forester, R. M. Revised Quaternary stratigraphy and chronology in the American Falls Area, Southeastern Idaho. *Cenozoic Geol. Idaho: Idaho Bur. Mines Geol. Bull.* **26**, 581–595 (1982).

9. Froese, D. *et al.* Fossil and genomic evidence constrains the timing of bison arrival in North America. *Proc. Natl. Acad. Sci.* **114**, 3457–3462 (2017).

10. Debruyne, R. *et al.* Out of America: ancient DNA evidence for a new world origin of late quaternary woolly mammoths. *Curr. Biol.* **18**, 1320–1326 (2008).

11. Pečnerová, P. *et al.* Mitogenome evolution in the last surviving woolly mammoth population reveals neutral and functional consequences of small population size. *Evol. Lett.* **1**, 292–303 (2017).

12. Heintzman, P. D. *et al.* A new genus of horse from pleistocene North America. *Elife* **6**, 1–43 (2017).

13. Chang, D. *et al.* The evolutionary and phylogeographic history of woolly mammoths: a comprehensive mitogenomic analysis. *Sci. Rep.* **7**, 44585 (2017).

14. Delsuc, F. *et al.* Ancient Mitogenomes Reveal the Evolutionary History and Biogeography of Sloths. *Curr. Biol*. **29**, 2031–2042 (2019).

15. Karpinski, E. *et al.* American mastodon mitochondrial genomes suggest multiple dispersal events in response to Pleistocene climate oscillations. *Nat. Commun*. **11**, 4048 (2020).

16. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*. **4**, 1–5 (2018).

17. Gowan, E. J. *et al.* A new global ice sheet reconstruction for the past 80 000 years. *Nat. Commun*. **12**, 1199 (2021).

18. Barbour, E. H. The American Mastodon with Mandibular Tusks. *Bull. Nebraska State Museum* **19**, 163–170 (1931).

19. Von Koenigswald, W., Widga, C. & Göhlich, U. B. *New mammutids (Proboscidea) from the Clarendonian and Hemphillian of Oregon - a survey of Mio-Pliocene mammutids from North America* (2021). doi:http://dx.doi.org/10.13140/RG.2.2.26270.66881

20. Green, J. L. Chronoclinal variation and sexual dimorphism in Mammut americanum (American mastodon) from the Pleistocene of Florida. *Florida Museum Nat. Hist. Bull*. **46**, 29 (2006).

21. Cherney, M. D., Fisher, D. C. & Rountrey, A. N. Tusk pairs in the Ziegler Reservoir mastodon (*Mammut americanum*) assemblage: Implications for site taphonomy and stratigraphy. *Quat. Int*. **443**, 168–179 (2017).

22. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15758–15763 (2013).

23. Meyer, M. & Kircher, M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb. Protoc*. **2010**, 1–10 (2010).

24. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res*. **40**, 1–8 (2012).

25. Enk, J. *et al. Mammuthus* population dynamics in Late Pleistocene North America: Divergence, Phylogeogrpaphy and Introgression. *Front. Ecol. Evol*. **4**, 1–13 (2016).

26. Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141–e141 (2014).

27. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–60 (2009).

28. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

29. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).

30. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

31. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

32. Widga, C. *et al.* Late Pleistocene proboscidean population dynamics in the North American Midcontinent. *Boreas* **46**, 772–782 (2017).

33. Stuiver, M., Reimer, P. J. & Reimer, R. W. *CALIB 8.2* [WWW program]. (2021).

34. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).

Chapter 5: Conclusion

Main findings and contributions

The works contained within this thesis greatly expands our understanding of the genetic landscape of Pleistocene proboscideans, particularly from warmer locales and time periods. In the second chapter, I provide a genetic confirmation of the megafaunal inhabitants of Bechan Cave, Utah as *Mammuthus columbi*. This not only confirmed previous morphology-based hypotheses, but also contextualized the mammoths of Bechan Cave within the greater context of *Mammuthus* mithochondrial diversity.

However, this publication also represents a significant addition to the palaeofecal-aDNA subfield, as well as to the representation of North American mammoths. While the extraction of ancient DNA from palaeofeces was shown to be possible in 1998[1], most palaeofecal studies did not recover complete mitochondrial genomes[1–4]. In fact, to the best of my knowledge, there has been only one previous study describing the recovery of complete mitochondrial genomes from coprolites of Pleistocene cave hyena and red deer[5]. Consequently, this study represents a fairly novel result, being only the second outlining the recovery of mitochondrial genomes from palaeofecal remains, and recovering this genome from hotter latitudes. This warmer location also means that not only does this study represent the first mitochondrial genome recovered from mammoth palaeofeces, but at the time of writing still represents the southernmost complete mammoth mitogenome ever recovered, and only the second from Utah.

In the third chapter, I conduct the first large-scale mitochondrial analysis of American mastodons across almost their entire range. This study provides a broad overview of mastodon mitochondrial diversity across space and time, and identified 5-6 well supported, geographically structured clades. I also identified multiple lineages likely representing the repeated expansion of American mastodons into east Beringia in response to deglaciation, extending previous palaeontological models focusing on only the last interglacial transition, through the mid-Pleistocene. I also found that animals corresponding to these northward interglacial expansions exhibited a high degree of genetic similarity in comparison to populations from south of the ice sheets, consistent with the idea that these expansions were likely conducted by small matriarchal herds.

This study probably represents the single most important contribution of this thesis. It represents the first step towards understanding the dynamics of extinct megafaunal browsers during the Pleistocene, and what effect glacial-interglacial cycles had on their demography and biogeography. Additionally, the majority of the samples analyzed in the study were past the effective limits of radiocarbon dating (*i.e.* of non-finite radiocarbon age). This differs from many ancient studies DNA which primarily focus on specimens from the last fifty thousand years. As such, it provides a good case study on how to work with these temporally-uncertain specimens in phylogenetic

analyses and the important contributions they can make to understanding complex Pleistocene ecology. This wide temporal range, combined with the near continental distribution of the samples also provides an important framework for addressing other palaeogenetic questions about mastodons.

In the fourth chapter, I start to fill in some of the gaps remaining in the American mastodon dataset generated in chapter 3, and take the first steps towards addressing the systemic and ecological questions about the relationships between the pacific mastodons (*Mammut pacificus*) and the American mastodons (*Mammut americanum*). I produce mitochondrial genomes from two mastodons from the American Falls Reservoir, Idaho. Phylogenetic analysis suggested these specimens were most closely related to other mastodons which inhabited eastern Beringia during the previous interglacial, and may represent a southward migration in response to glaciation.

However, this study also begins to address evolutionary and ecological questions about a newly described species of mastodon – *M. pacificus*. American Falls was one of the first locations this species was identified outside of California, and was argued to be the only mastodon present in the region[6]. While poor morphological preservation prohibits us from confidently assigning our two mastodons to either species, my study instead suggests that these palaeontologically-assumed patterns are not as clear as they were originally made out to be. We identify two possible scenarios that explain the data: either *M. pacificus* and *M. americanum* are conspecific and any morphological variation likely represents localized adaptations to unique diets or environments; or both *M. pacificus* and *M. americanum* occupied Idaho during the last Pleistocene and likely interacted. Regardless of the eventual outcome, the genetic work shows that the palaeobiology of Pleistocene *Mammut* is more complicated than was previously proposed, and can provide important information for further palaeontological analyses.

This thesis represents a significant increase in our understanding of Pleistocene proboscideans, increases the genetic representation of mammoths from southern locales, and provides the first high-resolution glimpse in the history of North American Pleistocene browsers. Collectively, these works represent a cross-section of the palaeogenetic field, providing well executed case studies from three of the four primary ancient DNA categories.

Future Directions

The works contained within this thesis provides many avenues for future research. As seen in chapter 2, the Colorado plateau and parts of the American southwest contain a wealth of information in, still relatively unexplored, palaeofecal accumulations. Although the preservation in these coprolites appears to be fairly low, studies like mine and others[5,7] highlight their usefulness as a source of ancient DNA, particularly in cases where

bone material may also be degraded or absent. Given their abundance in the southwest, and the greater willingness of many curators to destructively sample them over skeletal material for molecular analyses, they represent a powerful opportunity to increase the genetic representation of many species.

Palaeofecal analysis also comes with the additional advantage of providing data on diet. While outside the scope of my work in chapter two, this could be incorporated into future projects, particularly as methodologies of extracting and sequencing ancient DNA improve and decrease in cost. For example, newly developed sedaDNA extraction protocols may improve on endogenous DNA recovery and an enrichment based approach may allow for better recovery of rare floral constituents[8]. Combining this information with mitochondrial or nuclear data might allow for the simultaneous examination of both megafaunal demography, as well as any environmental factors they may have been experiencing. This would be most powerful in caves with significant accumulations of dung and strong chronological controls.

As mentioned previously, the mastodon genetic framework constructed in chapter 3 is extensive, but has some geographic and temporal holes. This opens up the possibility of many localized studies addressing some of those shortcomings, akin to what we started in chapter 4. This comparative dataset allows for a much more detailed examination of the mastodons from any given region or time point, and their contexualization within the broader evolutionary history of the species. Of particular interest might be expanding along the American west coast to fully characterize the diversity of *M. pacificus*, as well as attempting more material from Nova Scotia and Mexico, both of which contain samples occupying unique positions in the mitochondrial phylogeny.

Outside of palaeogenomics, chapters 3 and 4 highlight a few additional avenues for research. Both chapters highlight the need for additional excavations to identify new material from many underrepresented locales, as well as closer re-examination of existing material to identify morphological trends consistent with DNA-informed phylogenies. For example, the Alberta mastodon record was identified as highly dynamic in chapter 3, but only thirteen mastodons have been recovered from the province, and currently only three mastodons have yielded mitochondrial DNA[9].

Additionally, a lot of the work in chapters 3 and 4 was hindered by poor temporal resolution for many mastodons in the phylogeny. While a significant portion of this can be attributed to these specimens being past the effective limits of radiocarbon dating, identifying new mastodons *in situ*, or exploring other radiometric dating methods may provide desperately needed calibration points to better resolve the age of many parts of the phylogeny. This will become especially important as more and more mastodons from deeper time points are added, to be able to separate out migrations during different interglacial, versus multiple migrating lineages during the same interglacial.

In chapter 3, I also highlighted that a similar pattern of expansion and contraction probably occurred in many other extinct Pleistocene taxa such as western camels and giant beavers[10,11]. The genetic analysis of these and other taxa should recapitulate similar biogeographic patterns as in mastodons, and may serve as a useful confirmation and identify periods of expansion not captured in the mastodon analysis. Extending these analyses to southern populations may also capture interesting biogeographic shifts in response to environmental restructuring as opposed to glacial progress. Additionally, we noted that many extant taxa are expanding northward in response to anthropogenic warming. If these taxa follow a similar expansion model to mastodons, northern populations may be genetically impoverished and present a false sense of security with respect to conservation.

Lastly, all the studies contained within this thesis deal with mitochondrial data. The mitochondrial genome's small size and high copy number in cells, as well as its usefulness for phylogenetic analyses and for species identification have made it a staple in ancient DNA analysis[12]. However, mitochondrial analyses also come with a suite of limitations primarily due to its nature as a maternally inherited molecule. Extending analyses such as those in chapters 3 and 4 to nuclear DNA would provide additional information as to the population history of mastodons, how different populations or species may have interacted, and if any functional adaptations may have played a role in the dispersal or persistence of various lineages. For example, a recent study of Pleistocene bears found evidence of admixture in northern populations likely derived from the post-LGM expansion[13].

Like all good science projects, this thesis opens up more questions and more possibilities than it solves. The ones presented above represent some of the questions I would be most interested in addressing, but are just a sample of possible avenues for future research, especially as methodologies continue to improve. The large amount of questions posed by this thesis and similar research, bode well for the overall health of the field, and will no doubt lead to many exciting discoveries in the years to come.

Chapter 5 Bibliography

1. Poinar, H. N. *et al.* Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* **281**, 402–406 (1998).

2. Hofreiter, M., Mead, J. I., Martin, P. & Poinar, H. N. Molecular caving. *Curr. Biol.* **13**, R693–R695 (2003).

3. Campos, P. F., Willerslev, E., Mead, J. I., Hofreiter, M. & Gilbert, M. T. P. Molecular identification of the extinct mountain goat, *Oreamnos harringtoni* (Bovidae). *Boreas* **39**, 18–23 (2010).

4. Campos, P. F. *et al.* Clarification of the taxonomic relationship of the extant and extinct ovibovids, *Ovibos, Praeovibos, Euceratherium* and *Bootherium*. *Quat. Sci. Rev*. **29**, 2123–2130 (2010).

5. Bon, C. *et al.* Coprolites as a source of information on the genome and diet of the cave hyena. *Proc Biol Sci* **279**, 2825–2830 (2012).

6. Dooley, A. C. *et al.* Mammut pacificus sp. nov., a newly recognized species of mastodon from the Pleistocene of western North America. *PeerJ* **7**, e6614 (2019).

7. Delsuc, F. *et al.* Ancient Mitogenomes Revisit the Evolutionary History and Biogeography of Sloths the evolutionary history and biogeography of sloths. *Curr. Biol*. **29**, 2031–2042 (2019).

8. Murchie, T. J. *et al.* Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. *Quat. Res*. **99**, 305–328 (2021).

9. Jass, C. N. & Barrón-Ortiz, C. I. A review of Quaternary proboscideans from Alberta, Canada. *Quat. Int*. **443**, 88–104 (2017).

10. Zazula, G. D. *et al.* A case of early Wisconsinan "over-chill": New radiocarbon evidence for early extirpation of western camel (*Camelops hesternus*) in eastern Beringia. *Quat. Sci. Rev*. **171**, 48–57 (2017).

11. Plint, T., Longstaffe, F. J. & Zazula, G. Giant beaver palaeoecology inferred from stable isotopes. *Sci. Rep*. **9**, 7179 (2019).

12. Paijmans, J. L. a, Gilbert, M. T. P. & Hofreiter, M. Mitogenomic analyses from ancient DNA. *Mol. Phylogenet. Evol*. **69**, 404–416 (2013).

13. Pedersen, M. W. *et al.* Environmental genomics of Late Pleistocene black bears and giant short-faced bears. *Curr. Biol*. 1–9 (2021). doi:10.1016/j.cub.2021.04.027

Appendix A

Supplementary Information for Mastodon mitochondrial genomes from American Falls, Idaho

Emil Karpinski[1,2*], Chris Widga[3], Andrew Boehm[4], Brandon R. Peecook[5,6], Melanie Kuch[1], Tyler Murchie[1,7], Hendrik N. Poinar[1,7,8*.]

[1] McMaster Ancient DNA Centre, Departments of Anthropology and Biochemistry, McMaster University, Hamilton, ON, L8S 4L9, Canada.
[2] Department of Biology, McMaster University, Hamilton, ON, L8S 4L8, Canada.
[3] Center of Excellence in Paleontology and Department of Geosciences, East Tennessee State University, Johnson City, TN, 37614, USA.
[4] University of Oregon Museum of Natural and Cultural History, University of Oregon, Eugene, OR, 97403.
[5] Idaho Museum of Natural History, Idaho State University, Pocatello, ID 83201.
[6] Department of Biological Sciences, Idaho State University, Pocatello, ID 83209.
[7] Department of Anthropology, McMaster University, Hamilton, ON, L8S 4L9, Canada.
[8] Department of Biochemistry, McMaster University, Hamilton, ON, L8S 4L8, Canada.

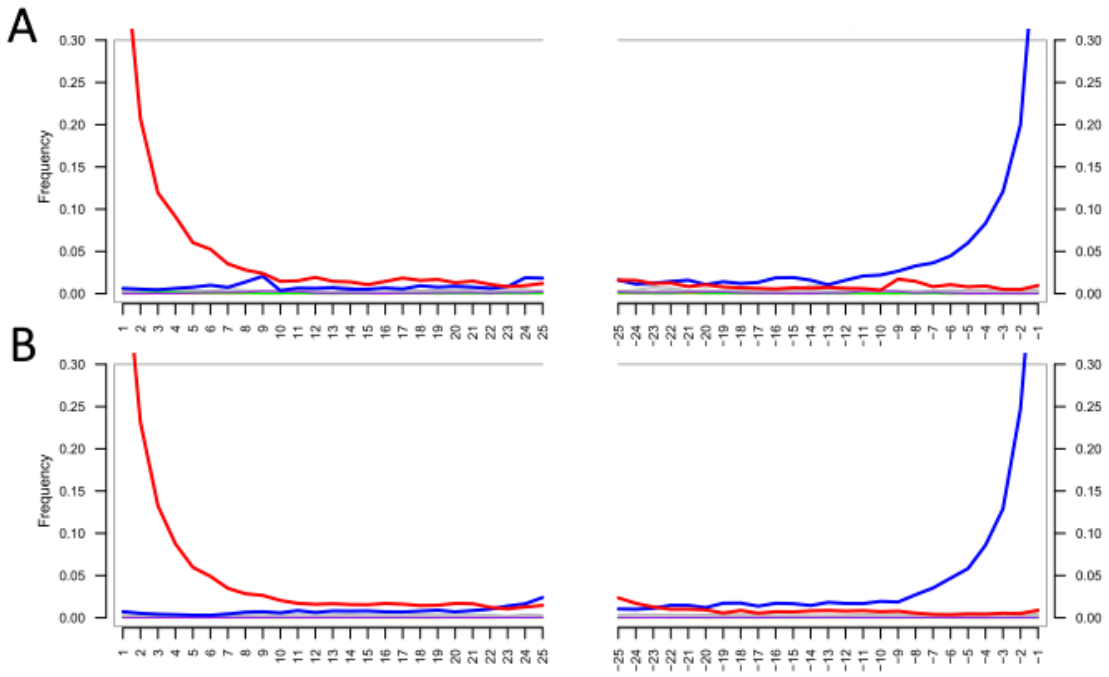
Supplementary Figure 1: Map of the American Falls Reservoir. The green box shows the approximate location of Lower American Falls where specimens were found. An inset showing the location of the American Falls Reservoir within the continental United States is shown in the lower right corner.

Supplementary Table 1: Extract, library, and sequencing data for specimens analyzed as part of this study. The amount of mapped reads indicated is shown after filtering for a minimum length of 24 base pairs and a map quality of 30 against the extended *M. americanum* reference.

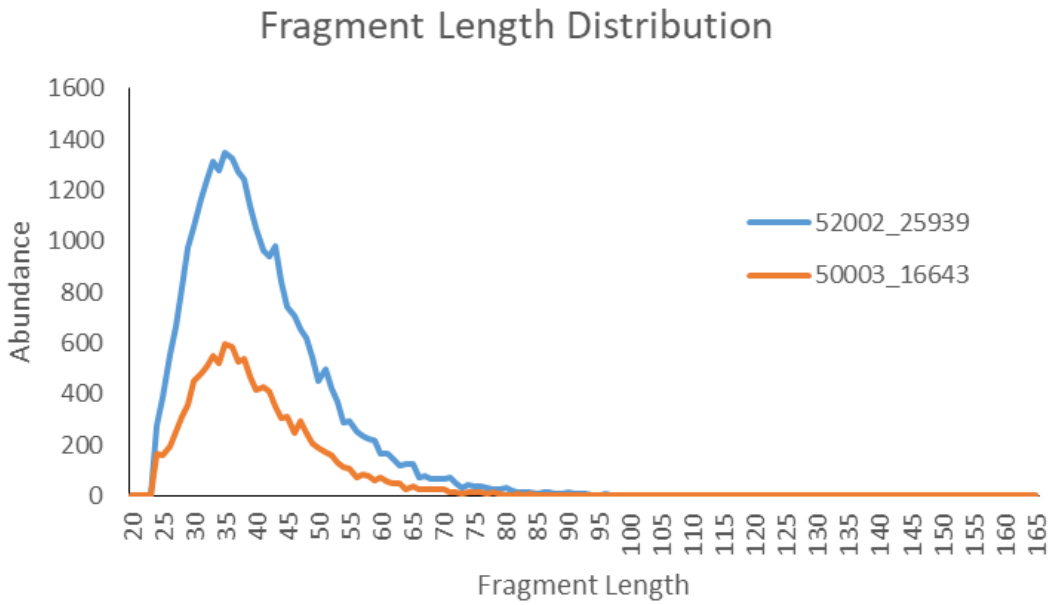
Specimen	Extract #	Amount Subsampled (mg)	Library	Indexed Library	Clusters	Mapped Reads
50003/16643	1118	381	L1118	L1118-I	2,641,734	1,752
				PP19-I	3,120,457	1,275
			L1118B	L1118B-I	3,075,586	1,156
				PP15-I	2,714,431	1,158
	1119	310	L1119	L1119-I	5,474,419	2,420
				PP20-I	2,182,755	1,089
			L1119B	L1119B-I	2,276,193	1,183
				PP16-I	4,329,002	1,058
52002/25939	1121	406	L1121	L1121-I	2,597,579	3,689
				PP21-I	3,062,495	11,676
			L1121B	L1121B-I	2,022,365	8,704
				PP17-I	933,827	4,323
	1122	476	L1122	L1122-I	2,722,203	35
				PP22-I	1,985,155	38
			L1122B	L1122B-I	1,751,077	47
				PP18-I	215,306	33
48001/1562	1115	224	L1115	L1115-I	2,540,017	71
48001/2282	1117	290	L1117	L1117-I	2,055,214	19
48001/279	1114	200	L1114	L1114-I	2,227,159	41
50001/1803	1116	235	L1116	L1116-I	1,899,471	0
71005/25867	1120	356	L1120	L1120-I	2,402,583	16
75008/39082	1123	500	L1123	L1123-I	13	0
	1124	373	L1124	L1124-I	6	0
Extraction Blank	1135	NA	L1135	L1135-I	1,630,667	17
				PP24-I	266,465	4
Library Blank	NA	NA	LBlankA	LBlankA-I	50,088	0

Supplementary Table 2: Reanalysis of previous data. Number of mapped reads obtained from reanalysis of data generated as part of Karpinski *et al.*[1]. The number of mapped reads shown is after filtering for a minimum length of 24 base pairs and mapping quality of 30.

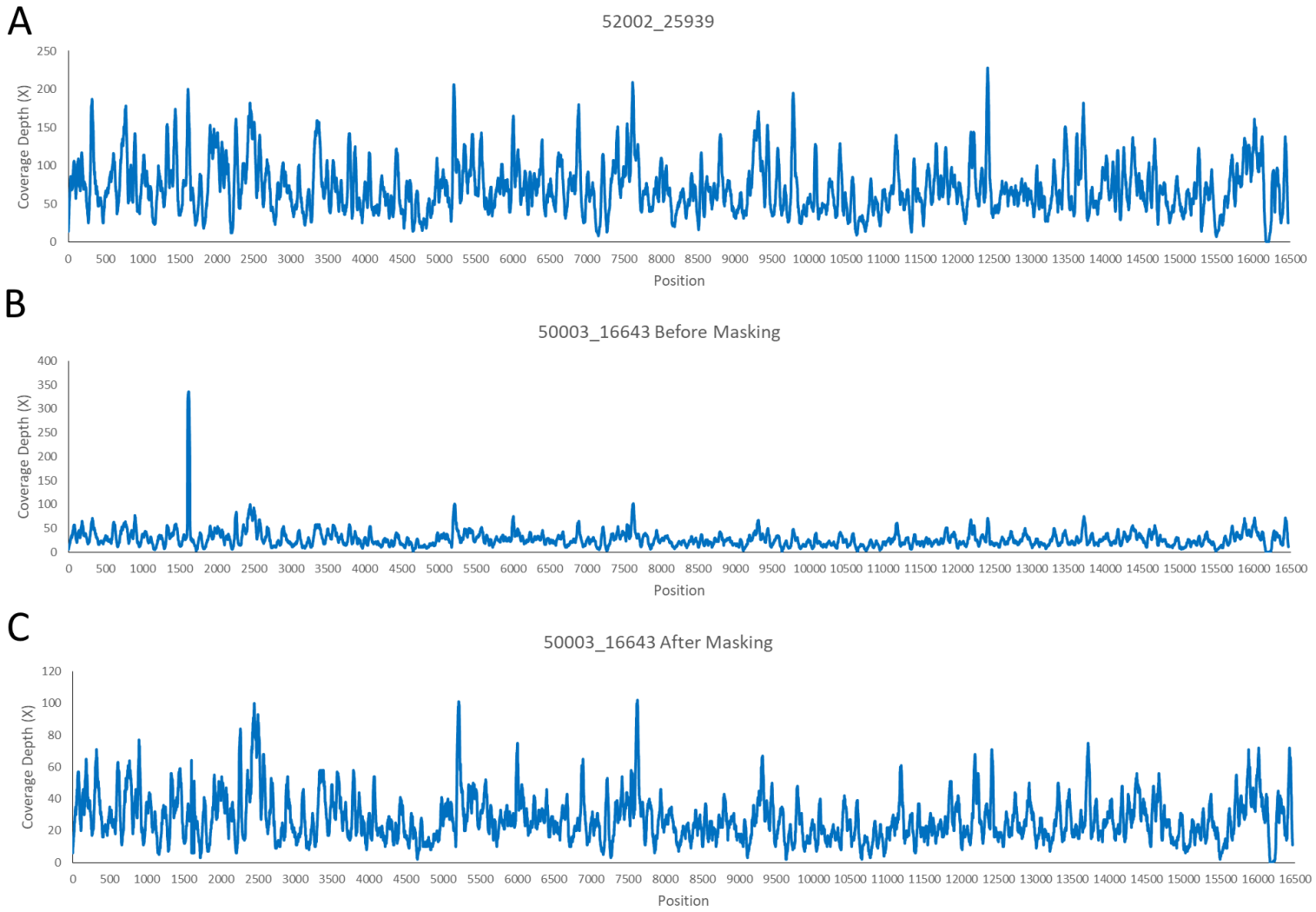
Specimen	Library	Mapped Reads
50003/16643	L956-E2B	235
	LFR3-E2	31
	LFR3-1P	17
	LFR3-2P	16
	LFR3-3P	9
	EEP38	6
	EEP79	208
52002/25939	L957-E2B	110
	LFR4-E2	73
	LFR4-1P	206
	LFR4-2P	147
	LFR4-3P	48



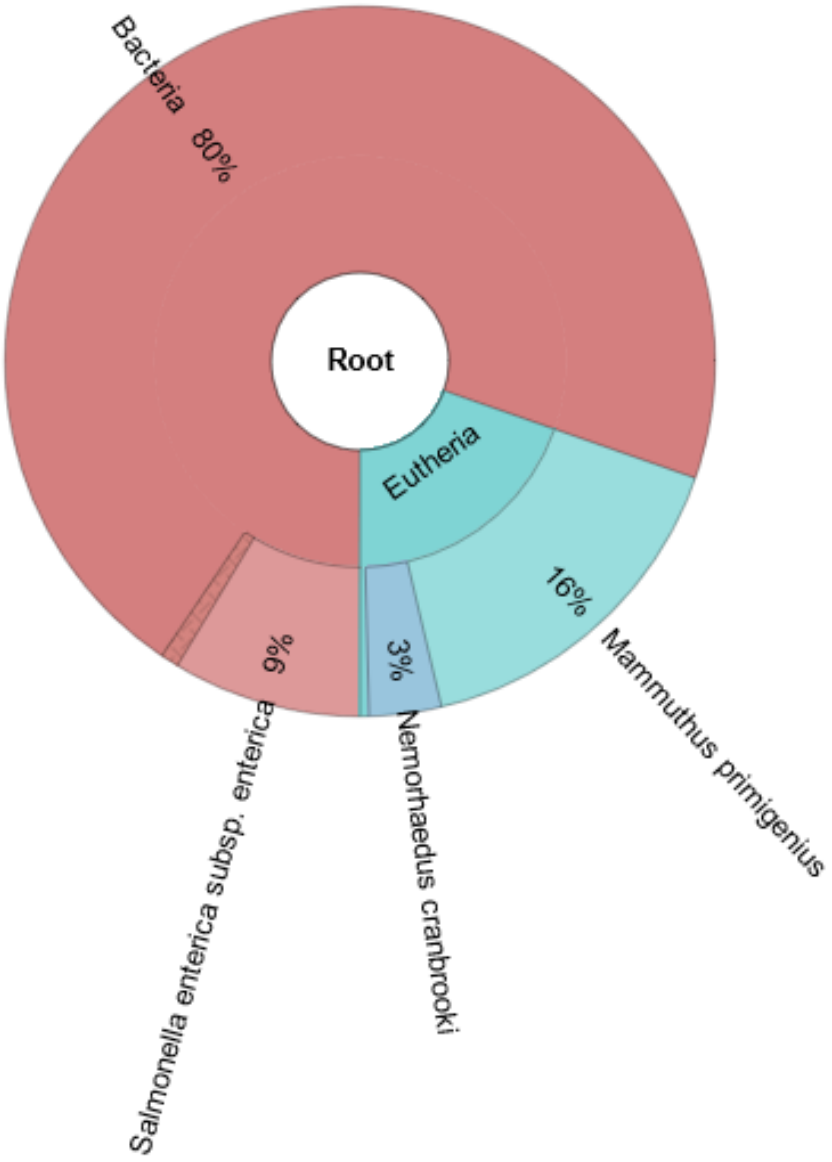
Supplementary Figure 2: MapDamage plots. Plots tracking the proportion of cytosine deamination at the 5' and 3' ends of libraries from specimens 50003/16643 (**A**) and 52002/25939 (**B**).



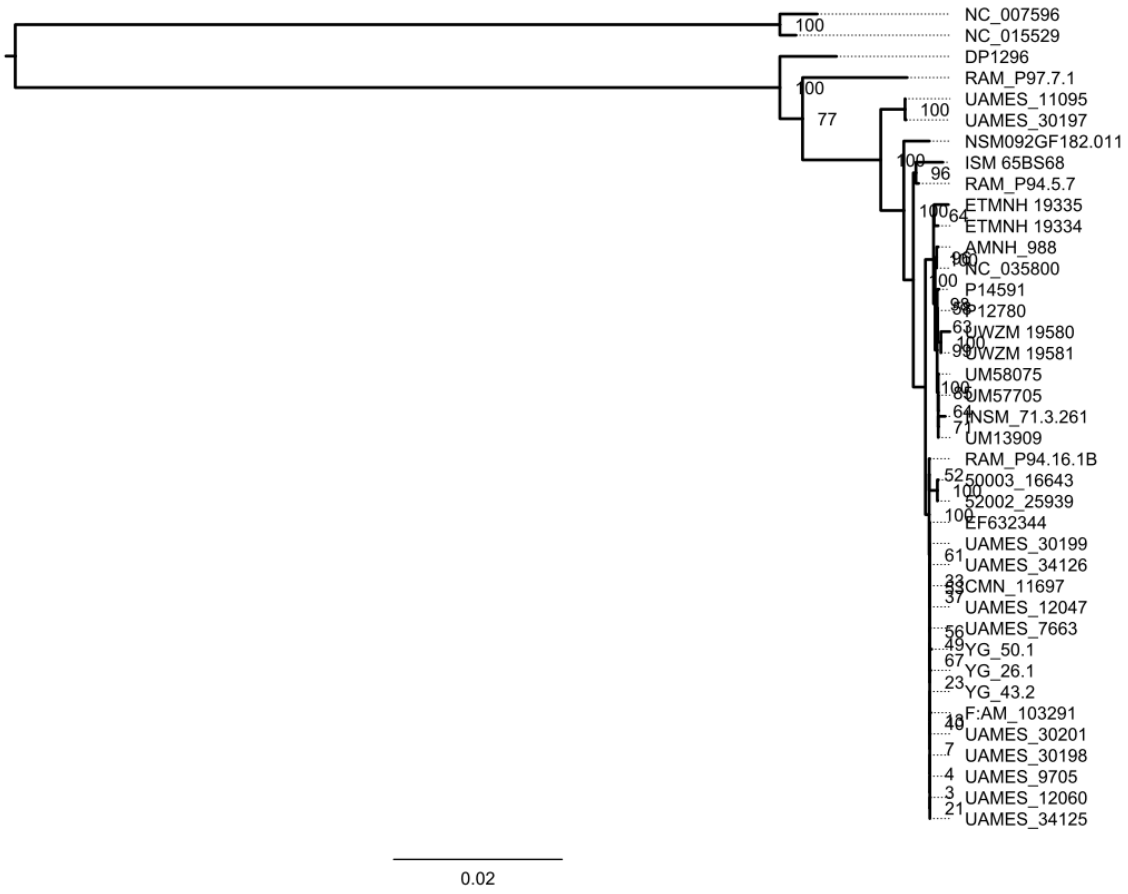
Supplementary Figure 3: Mapped fragment lengths. Fragment length distributions for specimens 50003/16643 (orange) and 52002/25939 (blue).



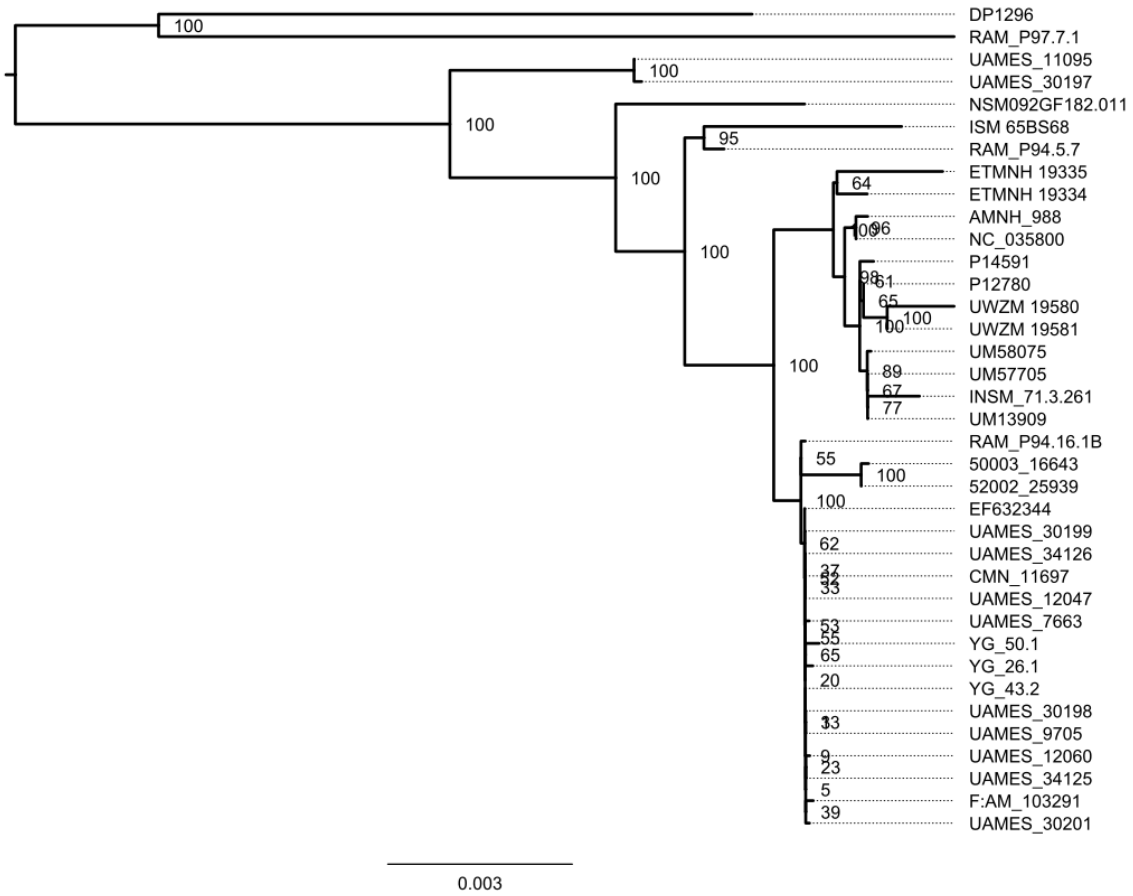
Supplementary Figure 4: Coverage plots. Coverage depth for specimens 52002/25939 (**a**) and 50003/16643 before (**b**) and after (**c**) masking of a 35 base pair fragment of the 16S rRNA region.



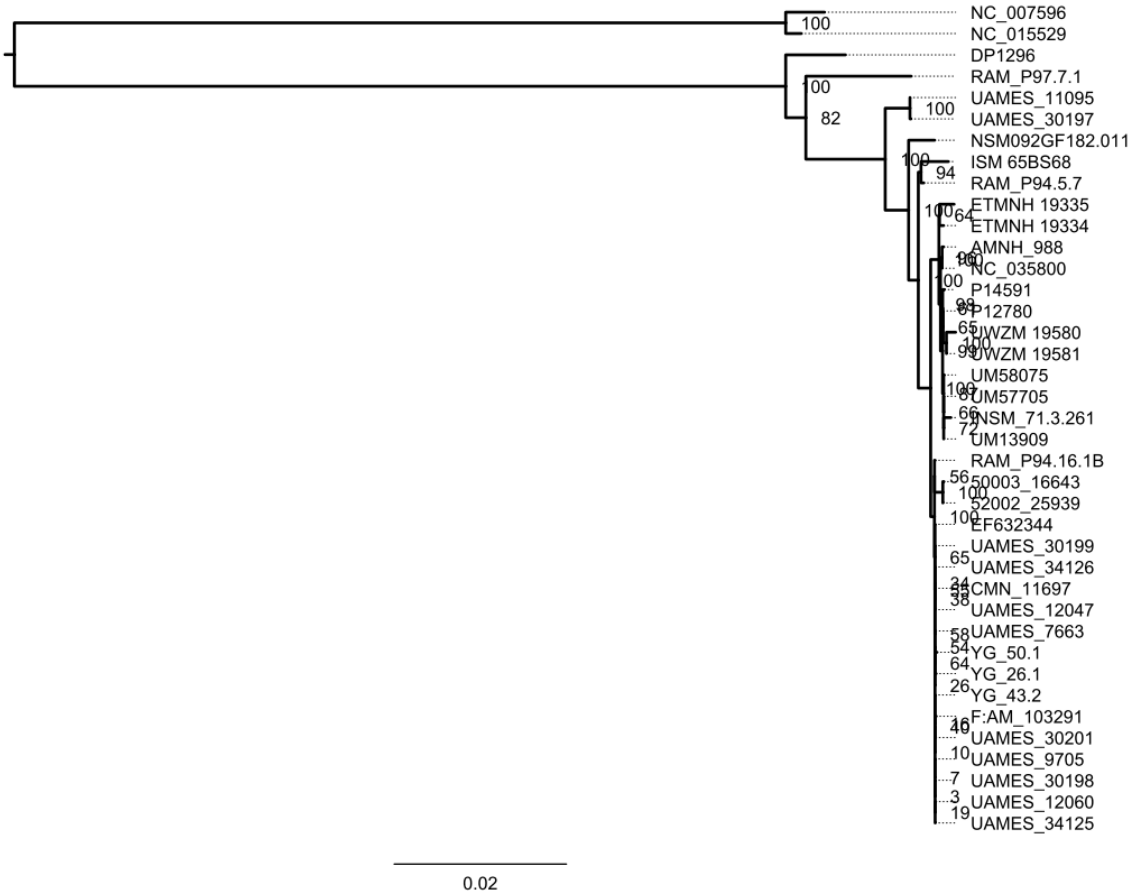
Supplementary Figure 5: BLAST analysis of reads mapping to the 16S rRNA stack in specimen 50003/16643. Reads from the >10 standard deviation stack as well as 10 bp upstream and downstream were extracted. Reads were then further filtered to a minimum fragment size of 24 bp to eliminate short reads spanning into this region and artificially truncated. The top 5 hits for each read were taken and summarized in Megan v6.21.1[2] before being summarized in KRONA[3]. Reads which returned a BLAST result (n = 244) are summarized below.



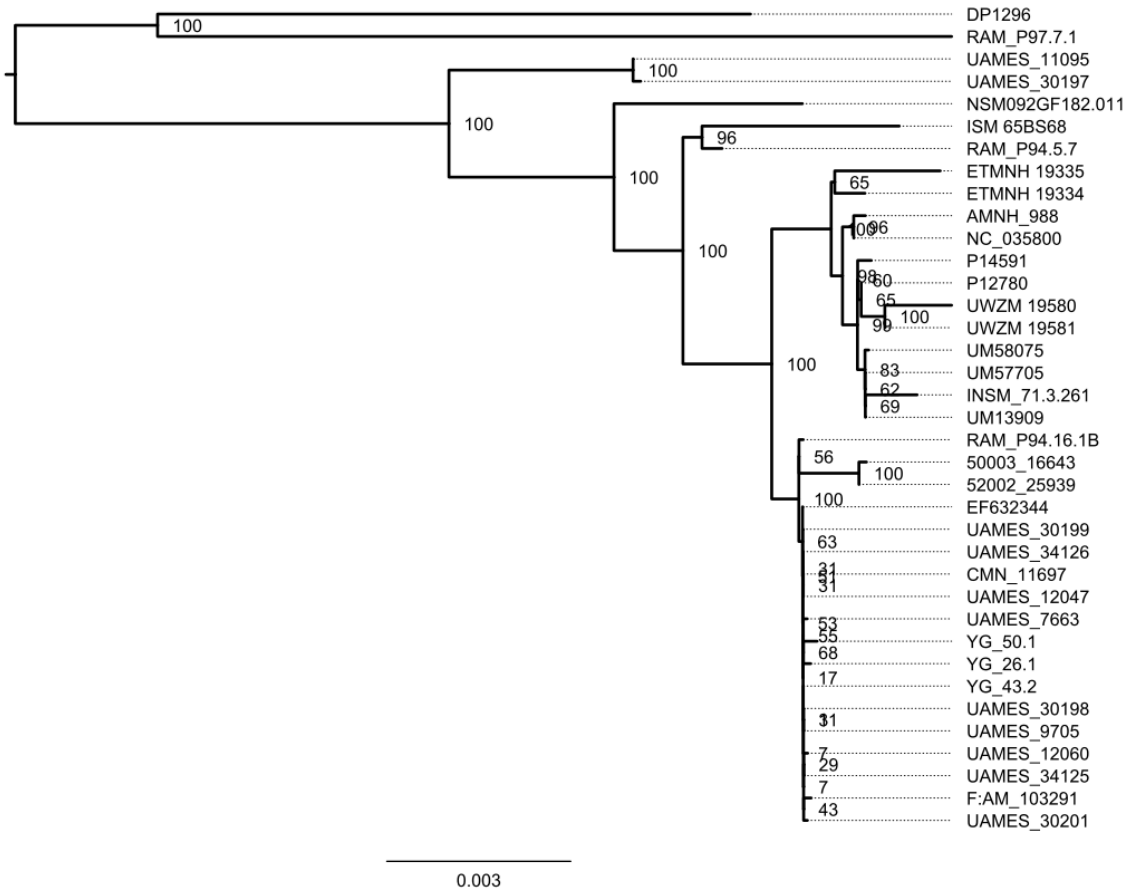
Supplementary Figure 6: ML tree with outgroup containing model. Maximum likelihood tree showing the relationships between specimens 50003/16643 and 52002/25939 and other American mastodons. Mastodons were rooted using two mammoth mitochondrial genomes – *Mammuthus primigenius* (NC_007596) and *Mammuthus columbi* (NC_015529). The phylogeny was constructed using a TIM3+F+I+G4 model with 1000 bootstrap replicates to assess node support.



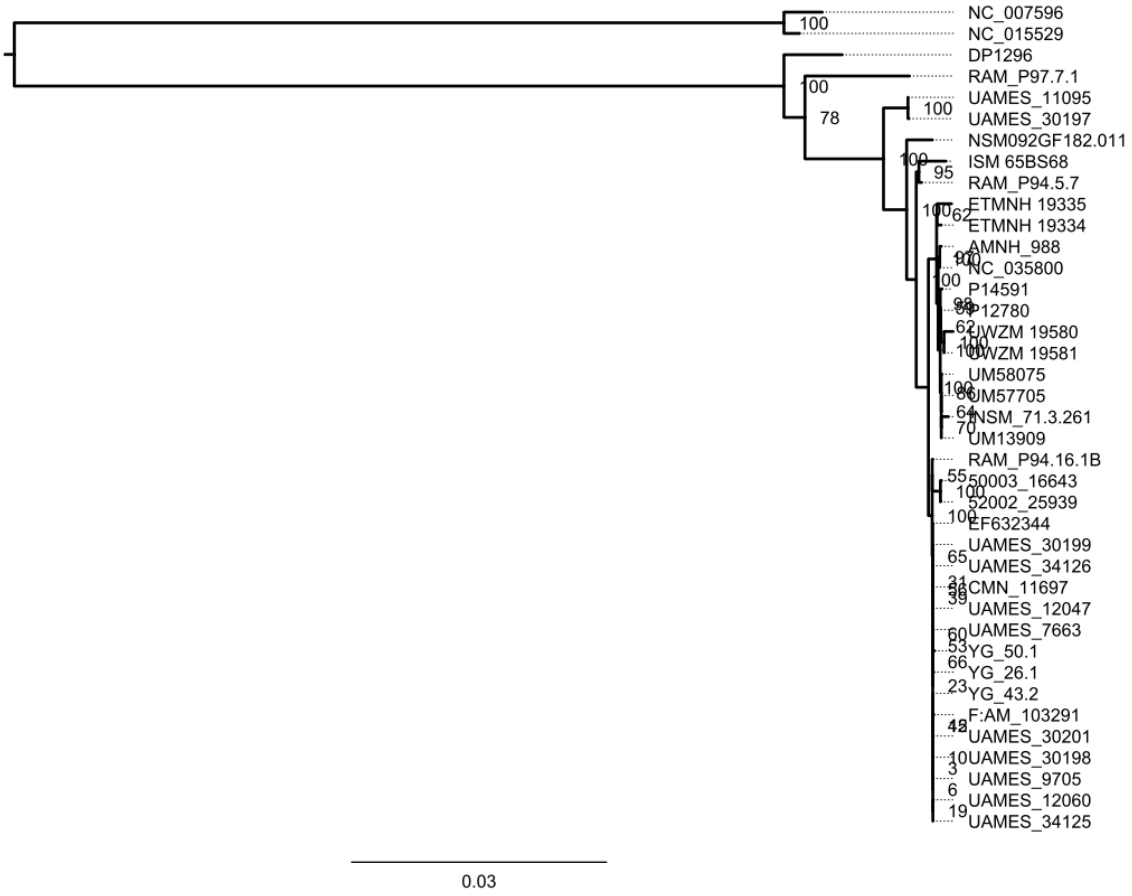
Supplementary Figure 7: ML tree with best no-outgroup model. Midpoint-rooted maximum likelihood tree showing the relationships between specimens 50003/16643 and 52002/25939 and other American mastodons. The phylogeny was constructed using a TPM3u+F+I model with 1000 bootstrap replicates to assess node support.



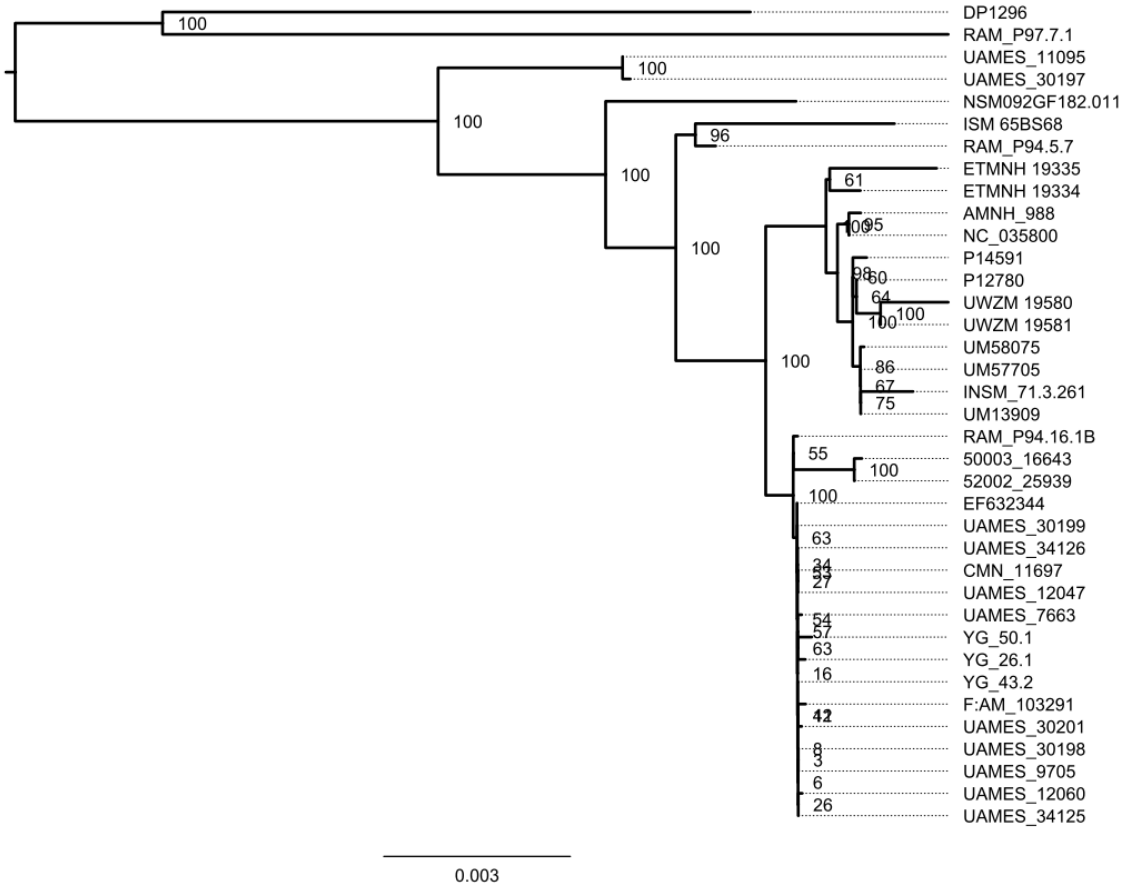
Supplementary Figure 8: Outgroup rooted tree with no-outgroup model. Maximum likelihood tree showing the relationships between specimens 50003/16643 and 52002/25939 and other American mastodons. Mastodons were rooted using two mammoth mitochondrial genomes – *Mammuthus primigenius* (NC_007596) and *Mammuthus columbi* (NC_015529). The phylogeny was constructed using the best-fitting model selected for the no outgroup dataset (TPM3u+F+I) with 1000 bootstrap replicates to assess node support.



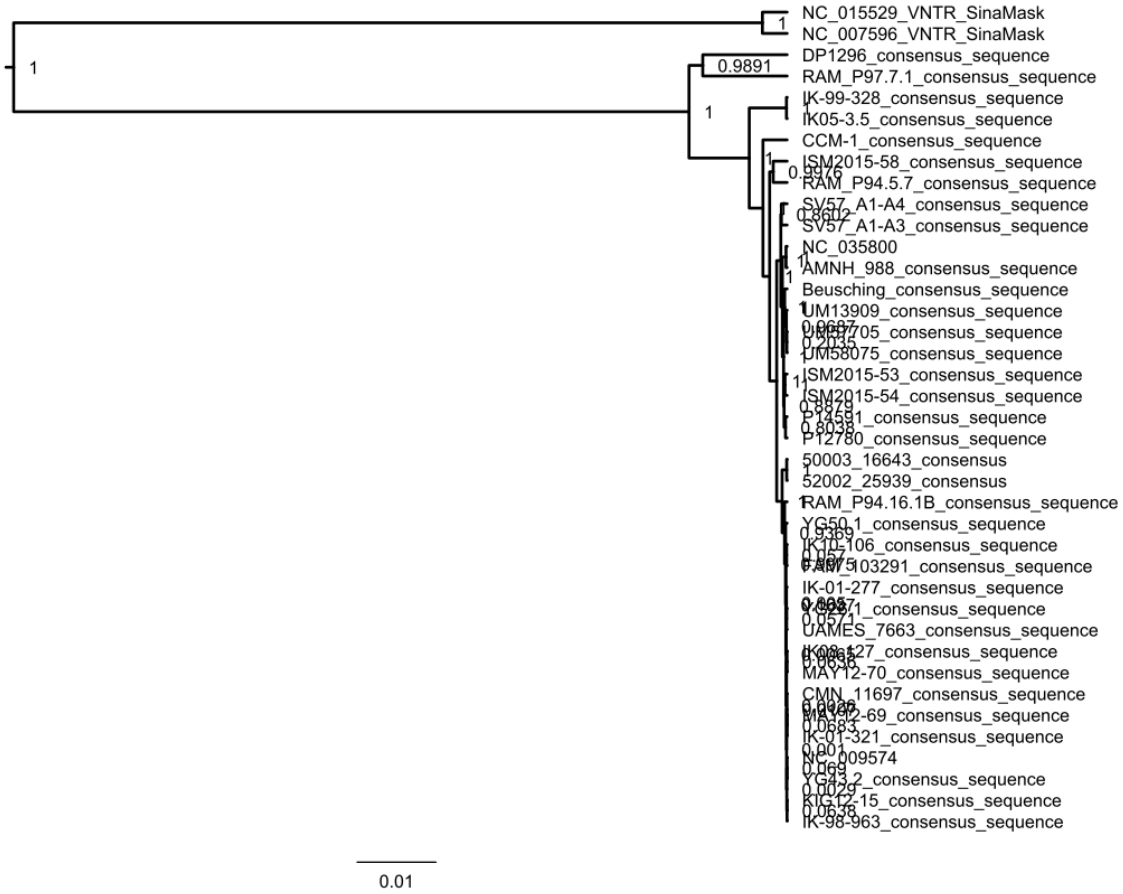
Supplementary Figure 9: Midpoint rooted tree with outgroup model. Midpoint-rooted maximum likelihood tree showing the relationships between specimens 50003/16643 and 52002/25939 and other American mastodons. The phylogeny was constructed using the best-fitting model selected for the mammoth rooted dataset (TIM3+F+I+G4) with 1000 bootstrap replicates to assess node support.



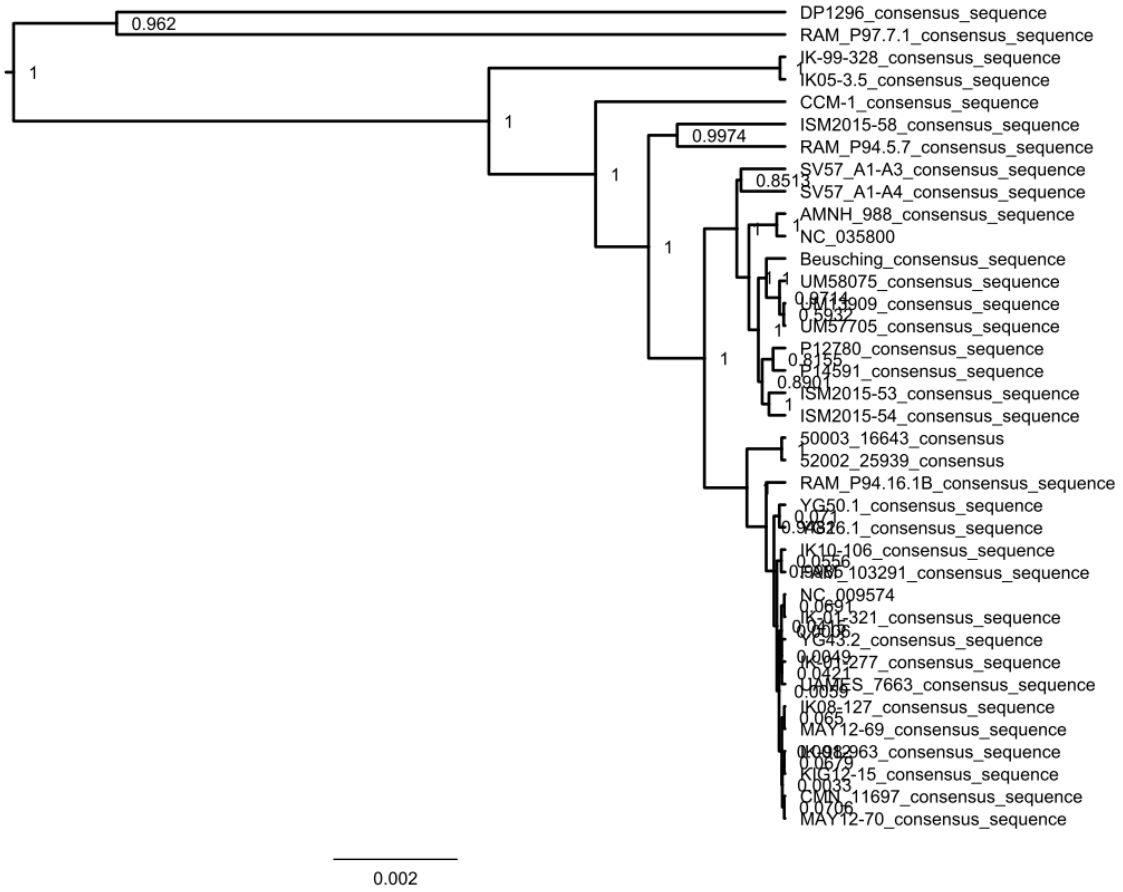
Supplementary Figure 10: Outgroup rooted tree with BEAST model. Maximum likelihood tree showing the relationships between specimens 50003/16643 and 52002/25939 and other American mastodons. Mastodons were rooted using two mammoth mitochondrial genomes – *Mammuthus primigenius* (NC_007596) and *Mammuthus columbi* (NC_015529). The phylogeny was constructed using an HKY+F+G4 model used in BEAST v1.10.5pre with 1000 bootstrap replicates to assess node support.



Supplementary Figure 11: Mid-point rooted tree with BEAST model. Midpoint-rooted maximum likelihood tree showing the relationships between specimens 50003/16643 and 52002/25939 and other American mastodons. The phylogeny was constructed using an HKY+F+G4 model used in BEAST v1.10.5pre with 1000 bootstrap replicates to assess node support.



Supplementary Figure 12: MCC tree with outgroup. Maximum clade credibility tree constructed in BEAST v1.10.5pre[4] to assess posterior probability support for nodes within the mastodon phylogeny with mammoth outgroups. The phylogeny was constructed using an HKY+F+G4 model and with no temporal calibration (*i.e.* all specimens assumed to be contemporary producing an ultrametric tree).



Supplementary Figure 13: MCC tree without outgroup. Maximum clade credibility tree constructed in BEAST v1.10.5pre[4] to assess posterior probability support for nodes within the mastodon phylogeny. The phylogeny was constructed using an HKY+F+G4 model and with no temporal calibration (*i.e.* all specimens assumed to be contemporary producing an ultrametric tree).

Supplementary Table 3: BEAST output with 52002/25939 calibration. Data table of all parameters from dating analyses when 52002/25939 is used as a calibration point. Values are after burn-in has been removed and all three independent chains combined. All age values are uncorrected and relative to the youngest specimen (UWZM 19580 – 13087 yBP).

	joint	prior	likelihood	treeModel.rootHeight	age(root)	treeLength	tmrca(Unknowns)	age(Unknowns)
Mean	-28570.5	-1139.26	-27431.3	2.96E+06	2.95E+06	1.23E+07	2.96E+06	2.95E+06
Stderr of mean	0.165	0.1695	0.0333	8000.903	8000.903	33823.59	8000.676	8000.676
Stdev	12.7586	10.6398	5.8056	6.02E+05	6.02E+05	2.54E+06	6.02E+05	6.02E+05
Variance	162.7823	113.2051	33.7055	3.62E+11	3.62E+11	6.43E+12	3.62E+11	3.62E+11
Median	-28571.1	-1140.07	-27431	2.95E+06	2.93E+06	1.23E+07	2.95E+06	2.93E+06
Value range	[-28622.7644, -28508.5414]	[-1177.0523, -1082.7562]	[-27459.5812, -27412.0594]	[1.0212E6, 6.0235E6]	[1.0081E6, 6.0104E6]	[4.194E6, 2.4454E7]	[1.0212E6, 6.0235E6]	[1.0081E6, 6.0104E6]
Geo. mean	n/a	n/a	n/a	2.90E+06	2.89E+06	1.20E+07	2.90E+06	2.89E+06
95% HPD interval	[-28594.8417, -28544.7164]	[-1159.4268, -1117.6832]	[-27442.7951, -27420.2833]	[1.8038E6, 4.1515E6]	[1.7908E6, 4.1384E6]	[7.3721E6, 1.7265E7]	[1.8038E6, 4.1515E6]	[1.7908E6, 4.1384E6]
ACT	2.26E+05	3.43E+05	44505.5	2.39E+05	2.39E+05	2.40E+05	2.39E+05	2.39E+05
ESS	5976.3	3938.7	30334	5657.8	5657.8	5621	5657.7	5657.7
# samples	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05

Supplementary Table 3. Continued.

	constant.popSize	kappa	alpha	clock.rate	meanRate	age(50003_16643)	age(AMNH_988)	age(CMN_1697)
Mean	7.66E+05	48.3341	0.054	4.70E-09	4.70E-09	14218.72	19955.74	1.20E+05
Stderr of mean	1675.485	0.0226	1.01E-04	1.58E-11	1.58E-11	40.3078	68.3887	405.8073
Stdev	1.55E+05	8.3023	0.0361	1.09E-09	1.09E-09	12264.12	18292.92	42648.86
Variance	2.41E+10	68.9282	1.30E-03	1.19E-18	1.19E-18	1.50E+08	3.35E+08	1.82E+09
Median	7.87E+05	47.4715	0.0484	4.50E-09	4.50E-09	11152.29	14707.25	1.15E+05
Value range	[1.5713E5, 1E6]	[23.7329, 114.0535]	[2.7909E-3, 0.3186]	[2.2778E-9, 1.3633E-8]	[2.2778E-9, 1.3633E-8]	[0.0922, 1.5867E5]	[0.0897, 1.874E5]	[36913.2674, 3.4299E5]
Geo. mean	7.48E+05	47.6478	0.0405	4.59E-09	4.59E-09	9073.1	11715.39	1.12E+05
95% HPD interval	[4.8263E5, 9.9999E5]	[33.4597, 65.1634]	[2.7909E-3, 0.1209]	[2.8664E-9, 6.858E-9]	[2.8664E-9, 6.858E-9]	[0.0922, 38253.1582]	[0.0897, 56553.5722]	[40894.6948, 2.0037E5]
ACT	1.57E+05	10000	10533.6	2.82E+05	2.82E+05	14583.19	18869.05	1.22E+05
ESS	8592.6	135003	128164.1	4795.9	4795.9	92574.4	71547.3	11045.2
# samples	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05

Supplementary Table 3. Continued.

	age(EF_632344)	age(FAM_103291)	age(RAM_P94.16.1B)	age(RAM_P94.5.7)	age(RAM_P97.7.1)	age(UAMES_11095)	age(UAMES_12047)	age(UAMES_12060)
Mean	1.20E+05	1.01E+05	1.93E+05	4.57E+05	1.97E+05	5.51E+05	1.20E+05	1.10E+05
Stderr of mean	416.59	297.5519	530.9651	1217.526	552.1791	3359.785	392.2957	327.9601
Stdev	42430.47	40352.7	47058.31	1.04E+05	1.30E+05	1.36E+05	42338.38	41549.84
Variance	1.80E+09	1.63E+09	2.21E+09	1.07E+10	1.68E+10	1.86E+10	1.79E+09	1.73E+09
Median	1.16E+05	95921.01	1.90E+05	4.52E+05	1.68E+05	5.61E+05	1.15E+05	1.05E+05
Value range	[36923.5585, 3.5717E5]	[36913.1338, 3.5837E5]	[37278.1609, 4.7181E5]	[1.2972E5, 7.8691E5]	[36914.8444, 7.8676E5]	[44450.5737, 7.8691E5]	[36926.0691, 3.5846E5]	[36915.0098, 3.4063E5]
Geo. mean	1.13E+05	93613.04	1.87E+05	4.45E+05	1.62E+05	5.31E+05	1.12E+05	1.02E+05
95% HPD interval	[44038.7296, 2.0261E5]	[36923.9735, 1.7585E5]	[1.0459E5, 2.8687E5]	[2.5951E5, 6.6451E5]	[36917.7534, 4.5243E5]	[3.1065E5, 7.8678E5]	[41890.5714, 2.002E5]	[36994.6389, 1.8532E5]
ACT	1.30E+05	73405.21	1.72E+05	1.87E+05	24509.22	8.21E+05	1.16E+05	84110.48
ESS	10373.7	18391.5	7854.8	7231.8	55082.5	1644.1	11647.6	16050.7
# samples	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05

Supplementary Table 3. Continued.

	age(UAMES_30197)	age(UAMES_30198)	age(UAMES_30199)	age(UAMES_30201)	age(UAMES_34125)	age(UAMES_34126)	age(UAMES_7663)	age(UAMES_9705)
Mean	5.24E+05	1.18E+05	1.20E+05	1.10E+05	1.20E+05	1.20E+05	1.09E+05	1.20E+05
Stderr of mean	3312.274	376.6707	410.8176	345.5785	394.2717	395.2017	334.5198	402.4302
Stdev	1.36E+05	42419.59	42450.22	41757.68	42583.88	42541.95	41574.85	42433.75
Variance	1.86E+10	1.80E+09	1.80E+09	1.74E+09	1.81E+09	1.81E+09	1.73E+09	1.80E+09
Median	5.33E+05	1.13E+05	1.15E+05	1.05E+05	1.15E+05	1.16E+05	1.04E+05	1.15E+05
Value range	[39096.6217, 7.8689E5]	[36914.7513, 3.5957E5]	[36916.0165, 3.4069E5]	[36913.5719, 3.8553E5]	[36913.5938, 3.568E5]	[36916.9586, 3.5972E5]	[36914.6133, 3.4437E5]	[36920.9684, 3.8426E5]
Geo. mean	5.03E+05	1.10E+05	1.13E+05	1.02E+05	1.12E+05	1.13E+05	1.01E+05	1.12E+05
95% HPD interval	[2.7248E5, 7.6487E5]	[39993.5888, 1.9723E5]	[43232.9407, 2.0167E5]	[36913.5719, 1.8665E5]	[42165.1789, 2.0089E5]	[42751.3331, 2.0148E5]	[36914.6133, 1.8551E5]	[42859.9082, 2.0143E5]
ACT	7.98E+05	1.06E+05	1.26E+05	92462.85	1.16E+05	1.17E+05	87403.49	1.21E+05
ESS	1691.1	12682.5	10677.2	14600.8	11665.3	11587.6	15446	11118.3
# samples	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05

Supplementary Table 3. Continued.

	age(UM13909)	age(YG26.1)	age(YG43.2)	age(YG50.1)	treeLikelihood	coalescent
Mean	33132.95	1.02E+05	1.20E+05	90915.33	-27431.3	-524.266
Stderr of mean	110.4195	293.4972	405.8554	244.9586	0.0333	0.1174
Stdev	24012.23	40141.01	42480.85	37640.75	5.8056	8.7589
Variance	5.77E+08	1.61E+09	1.80E+09	1.42E+09	33.7055	76.7177
Median	28898.61	96608.46	1.15E+05	84444.25	-27431	-524.958
Value range	[0.1205, 1.6679E5]	[36913.224, 3.2148E5]	[36933.8075, 3.6646E5]	[36914.0461, 3.0568E5]	[-27459.5812, -27412.0594]	[-557.3283, -475.7474]
Geo. mean	22372.32	94088.39	1.12E+05	83677.78	n/a	n/a
95% HPD interval	[0.1205, 78135.8034]	[36913.224, 1.7569E5]	[43844.3524, 2.0211E5]	[36914.0461, 1.6215E5]	[-27442.7951, -27420.2833]	[-540.8287, -506.4301]
ACT	28547.84	72173.52	1.23E+05	57176.23	44505.5	2.43E+05
ESS	47290.1	18705.3	10955.7	23611.7	30334	5564.9
# samples	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05

Supplementary Table 4: BEAST output without 52002/25939 calibration. Data table of all parameters from dating analyses when 52002/25939 is treated as a specimen of unknown, but possibly finite age. Values are after burn-in has been removed and all three independent chains combined. All age values are uncorrected and relative to the youngest specimen (UWZM 19580 – 13087 yBP).

	joint	prior	likelihood	treeModel.rootHeight	age(root)	treeLength	tmrca(Unknowns)	age(Unknowns)
Mean	-28599.1	-1167.07	-27432	3.04E+06	3.03E+06	1.26E+07	3.04E+06	3.03E+06
Stderr of mean	0.1694	0.1628	0.0361	8113.613	8113.613	34517.13	8113.439	8113.439
Stdev	12.6975	10.5526	5.8626	6.05E+05	6.05E+05	2.55E+06	6.05E+05	6.05E+05
Variance	161.2258	111.3567	34.3696	3.66E+11	3.66E+11	6.51E+12	3.66E+11	3.66E+11
Median	-28599.6	-1167.82	-27431.8	3.03E+06	3.01E+06	1.26E+07	3.03E+06	3.01E+06
Value range	[-28652.5308, -28519.6888]	[-1207.6653, -1093.3688]	[-27464.1247, -27410.7436]	[8.9833E5, 6.0732E6]	[8.8525E5, 6.0601E6]	[3.2008E6, 2.4561E7]	[8.9833E5, 6.0732E6]	[8.8525E5, 6.0601E6]
Geo. mean	n/a	n/a	n/a	2.98E+06	2.97E+06	1.24E+07	2.98E+06	2.97E+06
95% HPD interval	[-28623.8415, -28573.8803]	[-1187.0755, -1145.8241]	[-27443.7615, -27420.8694]	[1.8695E6, 4.2301E6]	[1.8564E6, 4.217E6]	[7.6923E6, 1.7643E7]	[1.8695E6, 4.2301E6]	[1.8564E6, 4.217E6]
ACT	2.40E+05	3.21E+05	51075.96	2.43E+05	2.43E+05	2.47E+05	2.43E+05	2.43E+05
ESS	5615.4	4202.7	26431.8	5554.6	5554.6	5460.9	5554.7	5554.7
# samples	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05

Supplementary Table 4. Continued.

	constant.popSize	kappa	alpha	clock.rate	meanRate	age(50003_16643)	age(52002_25939)	age(AMNH_988)
Mean	7.76E+05	48.3509	0.0542	4.57E-09	4.57E-09	40355.32	57543.9	20565.68
Stderr of mean	1666.263	0.0228	1.01E-04	1.57E-11	1.57E-11	204.0842	249.7662	70.0295
Stdev	1.52E+05	8.3349	0.0362	1.06E-09	1.06E-09	31533.12	34131.27	18870.66
Variance	2.32E+10	69.4713	1.31E-03	1.12E-18	1.12E-18	9.94E+08	1.16E+09	3.56E+08
Median	7.98E+05	47.4898	0.0488	4.39E-09	4.39E-09	33366.91	52432.27	15142.09
Value range	[1.2032E5, 1E6]	[25.3109, 103.1907]	[2.7908E-3, 0.2841]	[2.2449E-9, 1.812E-8]	[2.2449E-9, 1.812E-8]	[1.282, 2.388E5]	[9.4477, 2.6224E5]	[0.0966, 1.7257E5]
Geo. mean	7.59E+05	47.6598	0.0407	4.46E-09	4.46E-09	26414.44	45770.2	12035.08
95% HPD interval	[4.9662E5, 1E6]	[33.3526, 65.0729]	[2.7908E-3, 0.121]	[2.8438E-9, 6.6402E-9]	[2.8438E-9, 6.6402E-9]	[1.282, 1.0162E5]	[18.7796, 1.2129E5]	[0.0966, 58338.4031]
ACT	1.62E+05	10141.17	10483.11	2.97E+05	2.97E+05	56549.85	72295.12	18592.37
ESS	8340.5	133123.7	128781.5	4552.7	4552.7	23873.3	18673.9	72612
# samples	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05

Supplementary Table 4. Continued.

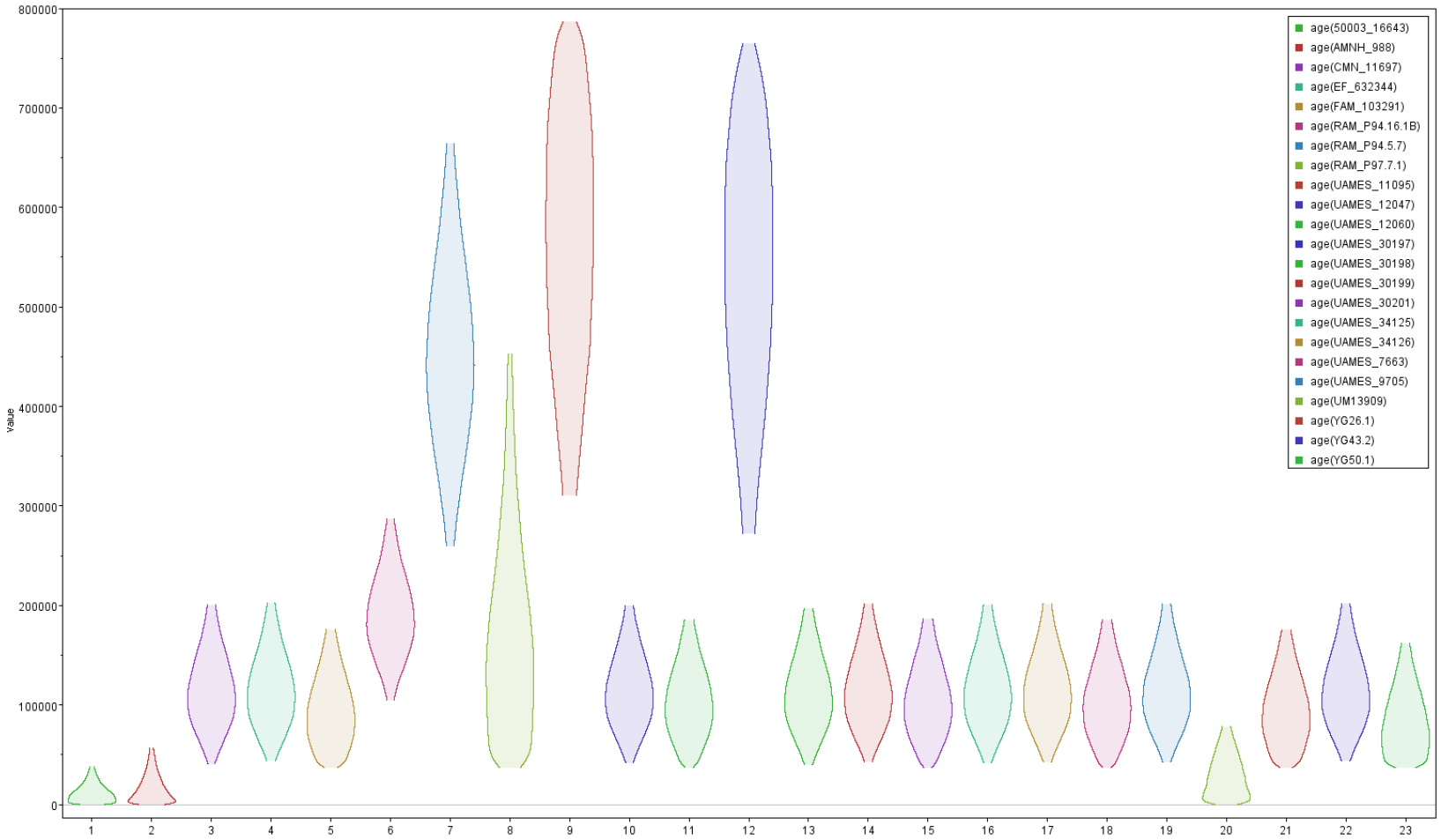
	age(CMN_11697)	age(EF_632344)	age(FAM_103291)	age(RAM_P94.16.1B)	age(RAM_P94.5.7)	age(RAM_P97.7.1)	age(UAMES_11095)	age(UAMES_12047)
Mean	1.28E+05	1.28E+05	1.09E+05	2.12E+05	4.71E+05	2.02E+05	5.57E+05	1.28E+05
Stderr of mean	461.8398	459.0891	348.1166	602.5501	1214.547	556.3233	3223.453	468.3106
Stdev	46393.52	46430.21	44182.72	50407.14	1.04E+05	1.32E+05	1.36E+05	46249.7
Variance	2.15E+09	2.16E+09	1.95E+09	2.54E+09	1.09E+10	1.75E+10	1.84E+10	2.14E+09
Median	1.23E+05	1.23E+05	1.03E+05	2.08E+05	4.66E+05	1.72E+05	5.68E+05	1.23E+05
Value range	[36913.5988, 4.0335E5]	[36914.4062, 3.9708E5]	[36915.6897, 4.2063E5]	[38881.093, 4.9782E5]	[1.2128E5, 7.8691E5]	[36913.2307, 7.8685E5]	[36981.1801, 7.869E5]	[36932.1541, 4.2502E5]
Geo. mean	1.19E+05	1.20E+05	99724.94	2.06E+05	4.59E+05	1.65E+05	5.37E+05	1.19E+05
95% HPD interval	[42031.1437, 2.1556E5]	[42619.8395, 2.1571E5]	[36915.6897, 1.899E5]	[1.1733E5, 3.1298E5]	[2.7233E5, 6.8001E5]	[36913.2307, 4.6626E5]	[3.1485E5, 7.869E5]	[43219.3609, 2.1559E5]
ACT	1.34E+05	1.32E+05	83809.14	1.93E+05	1.84E+05	23864.21	7.63E+05	1.38E+05
ESS	10090.9	10228.3	16108.4	6998.3	7355.5	56571.3	1770.3	9753.2
# samples	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05

Supplementary Table 4. Continued.

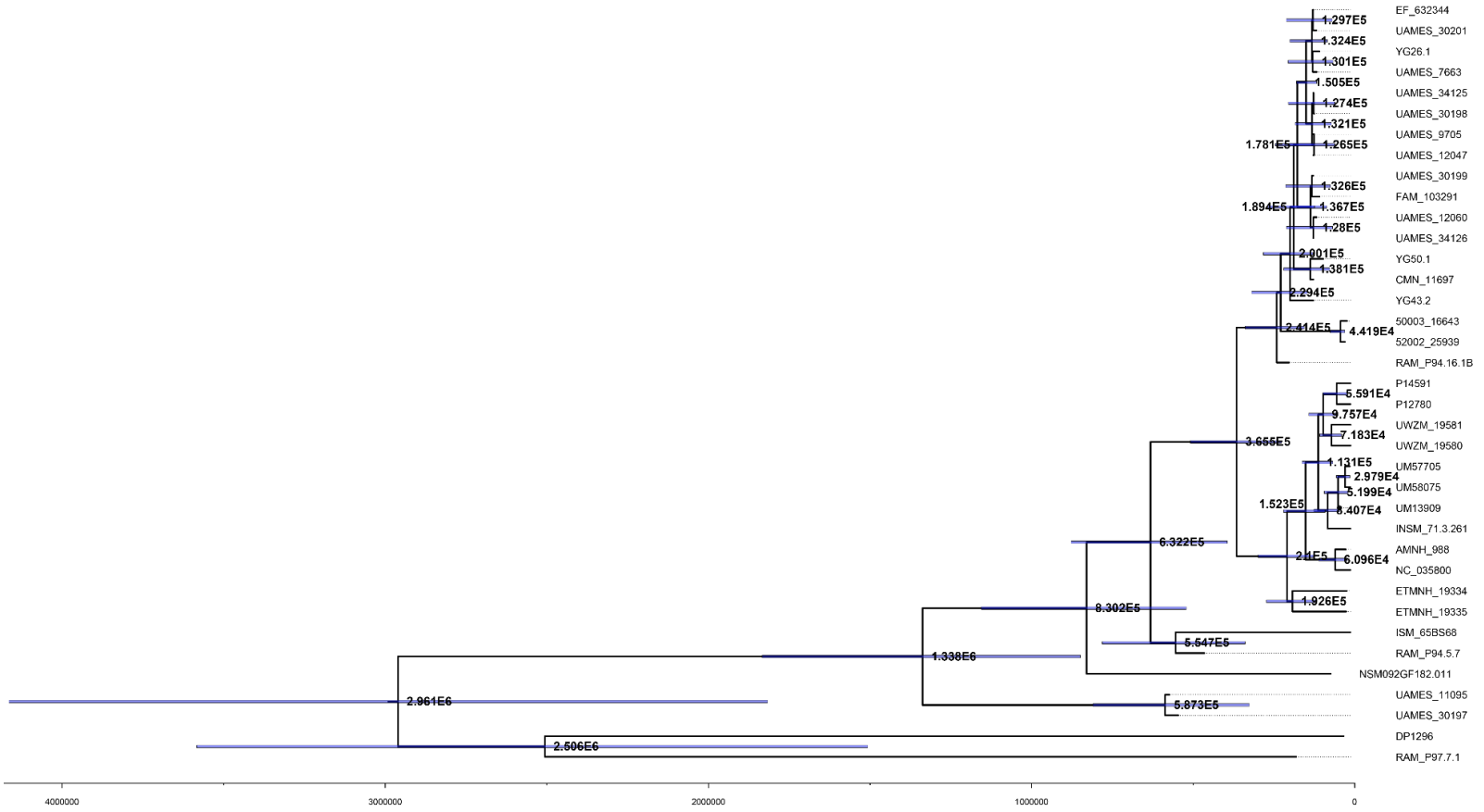
	age(UAMES_12060)	age(UAMES_30197)	age(UAMES_30198)	age(UAMES_30199)	age(UAMES_30201)	age(UAMES_34125)	age(UAMES_34126)	age(UAMES_7663)
Mean	1.18E+05	5.29E+05	1.25E+05	1.28E+05	1.18E+05	1.28E+05	1.28E+05	1.17E+05
Stderr of mean	400.1423	3193.577	447.7493	465.3575	404.3619	454.8403	467.3196	389.7794
Stdev	45857.32	1.36E+05	46178.45	46542.61	45699.65	46074.47	46306.49	45650.33
Variance	2.10E+09	1.85E+10	2.13E+09	2.17E+09	2.09E+09	2.12E+09	2.14E+09	2.08E+09
Median	1.12E+05	5.40E+05	1.20E+05	1.23E+05	1.13E+05	1.23E+05	1.23E+05	1.12E+05
Value range	[36918.7999, 4.1817E5]	[36972.7958, 7.8686E5]	[36928.9291, 4.1947E5]	[36914.5768, 3.8519E5]	[36917.9373, 4.0897E5]	[36914.4845, 4.1492E5]	[36919.9456, 3.6409E5]	[36923.2605, 3.9177E5]
Geo. mean	1.09E+05	5.08E+05	1.17E+05	1.20E+05	1.09E+05	1.19E+05	1.19E+05	1.08E+05
95% HPD interval	[36934.7583, 2.01E5]	[2.7642E5, 7.68E5]	[41788.2636, 2.1279E5]	[43362.0815, 2.1691E5]	[37712.4211, 2.0218E5]	[45459.5401, 2.176E5]	[43667.1944, 2.1653E5]	[36930.0384, 2.0053E5]
ACT	1.03E+05	7.44E+05	1.27E+05	1.35E+05	1.06E+05	1.32E+05	1.38E+05	98422.98
ESS	13133.6	1815.5	10636.7	10002.9	12772.7	10261.2	9818.7	13716.6
# samples	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05

Supplementary Table 4. Continued.

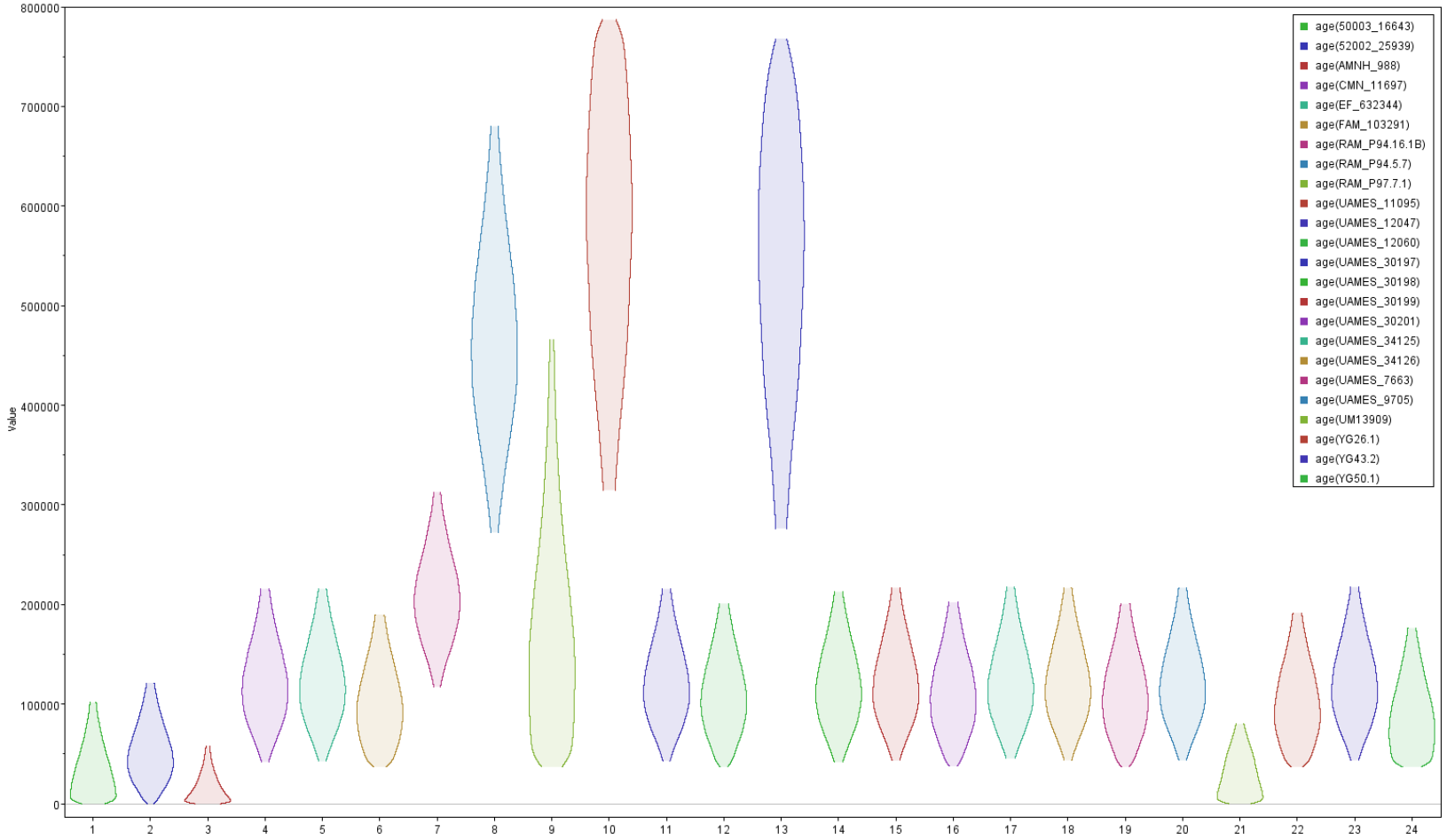
	age(UAMES_9705)	age(UM13909)	age(YG26.1)	age(YG43.2)	age(YG50.1)	treeLikelihood	coalescent
Mean	1.28E+05	33895.06	1.09E+05	1.29E+05	97423.78	-27432	-525.043
Stderr of mean	462.1487	111.1997	355.9476	474.9622	290.6172	0.0361	0.1181
Stdev	46355.91	24576.42	44471.4	46680.98	41772.57	5.8626	8.6776
Variance	2.15E+09	6.04E+08	1.98E+09	2.18E+09	1.74E+09	34.3696	75.3009
Median	1.23E+05	29524.53	1.03E+05	1.24E+05	90411.47	-27431.8	-525.643
Value range	[36919.0647, 3.9912E5]	[0.7069, 1.7043E5]	[36916.4922, 3.7278E5]	[36927.617, 3.9298E5]	[36913.59, 3.6136E5]	[-27464.1247, -27410.7436]	[-557.8955, -458.9951]
Geo. mean	1.20E+05	22871.66	1.01E+05	1.20E+05	89042.33	n/a	n/a
95% HPD interval	[43590.8164, 2.1688E5]	[0.7069, 80067.0724]	[36916.4922, 1.9155E5]	[43622.0044, 2.1748E5]	[36913.59, 1.7632E5]	[-27443.7615, -27420.8694]	[-541.3466, -507.3368]
ACT	1.34E+05	27638.63	86488.29	1.40E+05	65344.19	51075.96	2.50E+05
ESS	10061.1	48845.8	15609.4	9659.6	20660.3	26431.8	5396.3
# samples	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05	1.35E+05



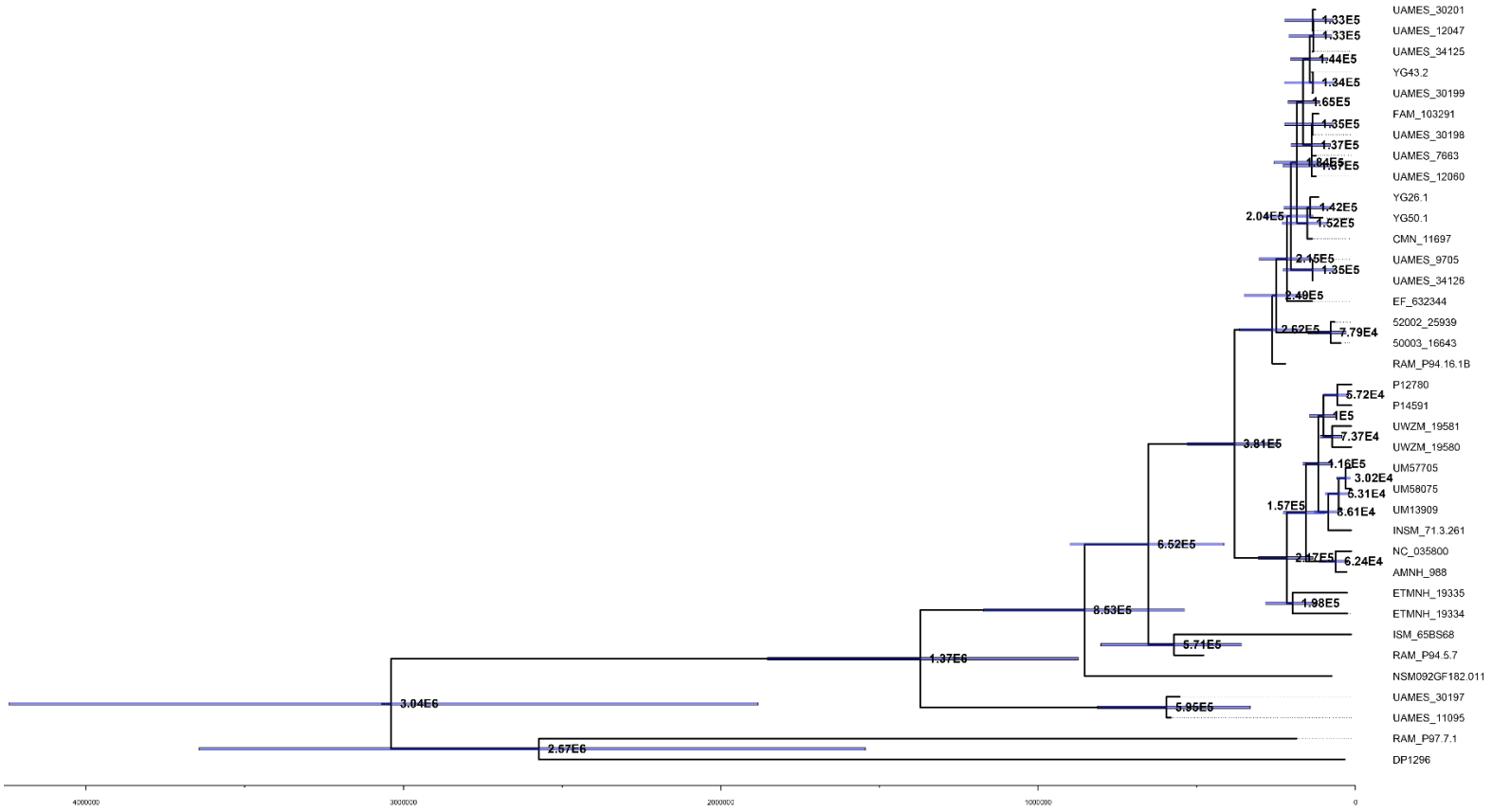
Supplementary Figure 14: Sample age estimation with 52002/25939 calibration. Violin plot with the posterior probability distributions on the estimated age of all specimens when 52002/25939 is used as a calibration point. All values are relative to the youngest specimen in the analysis (UWZM 19580 @ 13,087 years before present).



Supplementary Figure 15: MCC tree with node age estimates when the age 52002/25939 is used as a calibration point. Node bars represent 95% HPD intervals for estimated ages. Listed values are all relative to the youngest specimen in the analysis (UWZM 19580 @ 13,087 years before present).



Supplementary Figure 16: Sample age estimation without 52002/25939 calibration. Violin plot with the posterior probability distributions on the estimated age of all specimens when 52002/25939 is treated as a specimen of unknown but possibly finite age. All values are relative to the youngest specimen in the analysis (UWZM 19580 @ 13,087 years before present).



Supplementary Figure 17: MCC tree with node age estimates when the age 52002/25939 is treated as unknown, but possibly finite. Node bars represent 95% HPD intervals for estimated ages. Listed values are all relative to the youngest specimen in the analysis (UWZM 19580 @ 13,087 years before present).

Appendix A Bibliography

1. Karpinski, E. *et al.* American mastodon mitochondrial genomes suggest multiple dispersal events in response to Pleistocene climate oscillations. *Nat. Commun.* **11**, 4048 (2020).

2. Huson, D. H. *et al.* MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* **12**, 1–12 (2016).

3. Ondov, B. D., Bergman, N. H. & Phillippy, A. M. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* **12**, 1–9 (2011).

4. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, 1–5 (2018).

"But is all this true?" said Brutha.

Didactylos shrugged. "Could be. Could be. We are here and it is now. The way I see it is, after that, everything tends towards guesswork."

"You mean you don't know it's true?" said Brutha.

"I think it might be," said Didactylos. "I could be wrong. Not being certain is what being a philosopher is all about."

– Terry Pratchett, *Small Gods*