MAMMOTH PHYLOGEOGRAPHY SOUTH OF THE ICE

MAMMOTH PHYLOGEOGRAPHY SOUTH OF THE ICE: LARGE-SCALE SEQUENCING OF DEGRADED DNA FROM TEMPERATE DEPOSITS

Jacob M. Enk, B.A., M.Sc.

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

DOCTOR OF PHILSOPHY (2014) McMaster University
(Biology) Hamilton, Ontario

TITLE: Mammoth phylogeography south of the ice: large-scale sequencing of degraded DNA from temperate deposits

AUTHOR: Jacob M. Enk, B.A. (Indiana University), M.Sc. (University of Utah)

SUPERVISOR: Hendrik N. Poinar, Ph.D.

NUMBER OF PAGES: (xii), 145

ABSTRACT

Extinct Pleistocene mammoths (*Mammuthus*) have been studied extensively at the genetic level. However due to both taphonomic and technological limitations, ancient DNA has been characterized from only one of several late Pleistocene mammoth species, the woolly mammoth (*M. primigenius*). This greatly limits our impression of their evolution and population history to the steppes of the northern latitudes, just one of several environments in which mammoths lived and went extinct. It also obscures the chronology of underlying population processes of this keystone megaherbivore, which prevents putting their history in the proper climatic and biogeographic context. Fortunately recent technological advances in high-throughput sequencing and targeted enrichment promise to expand mammoth and other Pleistocene faunal population phylogeography to non-permafrost, non-cave burial contexts. However the capacity and behavior of these combined technologies for characterizing poorly-preserved ancient DNA is largely unexplored, preventing efficient and routine use for population-level studies. In this thesis I test and apply these technologies to remains of mammoth species found in the temperate areas of North America. I first demonstrate their potential in sequencing DNA from these poorly-preserved remains, and then I evaluate new methods for their efficient application to large sample sets, as well as for capturing complete nuclear genomes. I then use these technologies to sequence dozens of mitochondrial genomes from Columbian (*M. columbi*) and other non-woolly mammoths, reconstructing their matrilineal phylogeography south of the ice. The revealed patterns not only imply a deep chronology for mammoth matrilineal diversity observed to date, but also that mammoth evolution in North America was likely characterized by separate episodes of interbreeding between resident and invading populations, and between ecotypes. Overall the biological and methodological discoveries afforded by this body of work outline several tools and avenues for future research on mammoth evolution, behavior, and extinction.

ACKNOWLEDGMENTS

Hendrik Poinar is an inspiring scientist, incredibly supportive supervisor, and truly good person. I came to his laboratory intimidated by his and his group's breadth of technical and conceptual expertise, but now leave it with the confidence to embrace and master new ways of doing things. Even during my episodes of resistance and exacerbation, he provided only patience and encouragement. And occasionally frisbee during work hours.

Thanks to my committee members Brian Golding and Jim Quinn, for guidance and keen scrutiny over this past half decade of my research. Their novel and varied perspectives were invaluable in making this project address broader issues of animal behavior and evolution.

I would not be involved in this science if it were not for Rika Kaestle. It was in her ancient DNA lab that I used my first pipette, did my first PCR, ran my first gel, and first sequenced DNA. I hope to honor her decision to let a rambunctious young undergrad like my former self work with sensitive ancient remains.

Despite my short time working with Regis Debruyne, he taught me a tremendous amount and remains an inspiration. He also performed much of the work that would form the foundation of this project.

Dennis O'Rourke, my Master's supervisor, once asked "So I have this mammoth – do you want to work on it?" This is the result. Thanks especially to Dennis for that, and for encouraging me to follow wherever my curiosity led. Thanks also to Alan Rogers for always having time to bounce ideas around, and to Jack Broughton for teaching me to ask, "What is the research question?" I cannot describe how much the habit of asking that question has impacted my professional and personal frame of mind.

We were a tight group at the McMaster Ancient DNA Centre. This project grew from countless hours of discussion and troubleshooting, especially with my friends and colleagues Alison Devault, Christine King, and Melanie Kuch. I thank them and all the present and former members of our lab for always being willing to discuss even the most numbing experimental details, and teaching me how to think.

I especially thank my dear friend, Tina Li. Her unwavering companionship and simultaneous intolerance for bullshit allowed me to keep my emotional and mental sanctity through this process.

And finally, thanks to my parents. It is impossible to list all the reasons. Much love to them both.

TABLE of CONTENTS

LIST of TABLES

LIST of FIGURES

DECLARATION OF ACADEMIC ACHIEVEMENT

This thesis is comprised of four chapters, three of which have been published in *Genome Biology*, *BioTechniques*, and *Molecular Biology and Evolution*, and one of which is in preparation for submission to *Molecular Ecology*. As first author of each individual study I was the primary agent in research and experimental design, data acquisition and analysis, and manuscript composition; individual author contributions are included with each chapter. In summary, Regis Debruyne, Alison Devault, Christine King, Melanie Kuch and Yusuf Murgha assisted in laboratory work and protocol development; Steven Salzberg and Todd Treangen assisted in data analysis; John Southon and Paul Szpak performed radiocarbon dating; Regis Debruyne, Daniel Fisher, Duane Froese, Ross MacPhee, Dennis O'Rourke, Hendrik Poinar, Jean-Marie Rouillard, Jeffrey Saunders, Chris Widga and Grant Zazula contributed to research question formulation and data interpretation. All associated coauthors assisted in editing the manuscripts.

PART I

INTRODUCTION

Mammoths (*Mammuthus*, Blumenbach 1799) were one of more than 30 genera of large-bodied North American mammals that went extinct at the end of the Pleistocene [1, 2]. These charismatic elephants are not only an iconic symbol of the Ice Age in the popular conscious, but thanks to abundant, highly visible and often extraordinarily well-preserved remains [3-5], they have been studied extensively at the scientific level for over 200 years. Like extant elephants, which shape the floral and faunal communities with which they coexist [6-8], mammoths too were keystone megaherbivores that sculpted the now-extinct Pleistocene environment [9-11]; indeed a major Pleistocene biome of the northern latitudes, the "mammoth steppe," bears their name [12]. Mammoths are also one of the best-studied of the extinct megafauna at the genetic level, having served as a 'model organism' for major advances in the field of paleogenetics. Starting with DNA hybridization experiments in the early 1980s [13], they were among the first extinct animals to be characterized at the mitochondrial [14, 15] and nuclear genomic sequence levels [16, 17] and their hemoglobin has even been expressed in *E. coli* and evaluated for biochemical properties related to cold tolerance [18]. However with few exceptions, the majority of the initial mammoth genetic studies focused on resolving family-level systematic relationships between the Elephantidae (*Mammuthus, Loxodonta* and *Elephas*) [19-21], which was a long-standing issue in proboscidean paleontology. This question now stands at least tentatively resolved thanks to extensive sequencing of single individuals and by using sequence from a closely-related outgroup species, the American mastodon (*Mammut americanum*) [22]. But perhaps the most illuminating work that has been done with mammoth DNA, at least in terms of understanding the evolution and dynamism of Pleistocene ecosystems, came from genetic characterization of now hundreds of individual mammoths throughout their subarctic range and throughout radiocarbon time [23-26]. These studies, using mostly mitochondrial cytochrome *b* and hypervariable region sequences, uncovered an apparently tumultuous history in their northern populations, with deep divide between coeval matrilines and a series of expansions, lineal replacements, and population contractions during the later Pleistocene. These and other population-level studies of genetic variation and geographic structure in Pleistocene fauna [27-31] potentially offer great insight into the exact nature of that ecosystem and why it went extinct.

However, population-level phylogeographic study of Pleistocene fauna suffers from several limitations that prevent a truly global picture of past ecosystem evolution and decline. Due mainly to a combination of technological shortcomings and the taphonomy of DNA preservation, our picture of Pleistocene ecology is almost entirely limited to regions where burial conditions have been relatively thermostable since the Pleistocene; that is, northern permafrost-bearing latitudes or regions with abundant caves. This is particularly unfortunate for study of mammoths, which were incredibly widespread and inhabited not just the tundra-steppe of the north, but also the savanna-parklands and arid grasslands of more southern latitudes where they most certainly had an equally significant impact on their environment. Furthermore, an exclusively northern source of mammoth genetic information confounds placing their variation in the proper taxonomic, and thus chronological, context. Much like in human paleontology, there are both lumping [32] and splitting [33] paradigms in mammoth taxonomy, but it is mostly agreed upon that multiple species of mammoth evolved throughout the Plio-Pleistocene, and that many of these were each adapted to very different environments. In North America alone, at least two mammoth species are thought to have occupied the continent during the Late Pleistocene [34, 35], and both went extinct, despite their adaptation to different ecosystems. But a north-centric impression of mammoth genetic diversity restricts it to just one of these species, *M. primigenius* or the woolly mammoth, and for that matter to just a portion of that species' range. The lack of genetic information from a closely-related mammoth species for which we have some sense of the timing of divergence from woolly mammoths obscures the origin of the deep matrilineal divide uncovered by recent studies. Is it a consequence of very recent structure in just the later evolution of woolly mammoth, or does it instead reflect the maintenance of variation that emerged in more ancient nominal taxa, perhaps retained as a result of matriarchal philopatry common to extant elephants [36-38], or re-introduced through interspecific hybridization [39, 40]? Without a confident cladogenic chronology, the timing of underlying mammoth population processes remain mysterious and thus cannot be clearly linked to background climatic changes or the biogeography of other Pleistocene animals.

Fortunately, recent technological advances are rendering moot many taphonomic limitations to paleogenetics. Along with genetic sample sets from dozens to hundreds of individuals, the development and application of high-throughput sequencing (HTS) marks the second profound paradigm shift in paleogenetics of the last decade. Thanks to its ability rapidly sequence DNA too short for practical gene region resequencing with overlapping PCR

amplicons, HTS has been used to reconstruct long genomic sequences from ancient individuals, allowing questions of origins and phenotype to finally be addressed [41-43]. With the rapidly declining cost of HTS, and continuing advances in adapting it to ancient DNA [44-47], we are approaching a time when complete genomes of organisms that lived tens to hundreds of thousands of years ago are a standard unit of analysis for studying the evolution and physiology of extinct organisms. However, it is a combination of genomic breadth and sampling depth that stands to offer the most insight into the population dynamics and related behaviors that underpinned not only the ecosystems in which ancient organisms lived, but also what may have caused their extinction.

To achieve this, researchers must contend with the inherent inefficiency of sequencing DNA extracts with very low endogenous DNA content, which is typical for ancient remains. This is where the third major advancement in paleogenetics steps in, borrowed from years of research developing and applying nucleic acid hybridization technology [48]. Targeted enrichment allows background DNAs to be significantly depleted from complex DNA extracts, greatly increasing the efficiency of ancient extract HTS [49-52]. It has seen a significant amount of use in recent years, even for relatively large sample sets [53, 54]. But its behavior and sensitivity remain mysterious and difficult to predict, as it is driven by complex chemico-physical interactions that are undoubtedly heavily influenced by the unique chemical characteristics of any given ancient DNA extract. It has also not been applied to large sample sets of very poorly-preserved specimens like those from non-cave, non-permafrost Pleistocene burial environments. If paleogenetics is to move towards long genomic sequences from dozens to hundreds of individual specimens found in heretofore uncharacterized environments, it requires that targeted enrichment of ancient DNA be better understood and experimental conditions be honed and optimized for efficiency. Only then can phylogeographic inquiry reconstruct the multi-biome population histories and extinctions of ancient flora and fauna.

GOALS

The goals of this thesis are two-fold. The first is to test HTS and targeted enrichment on very poorly-preserved remains, with the aim of improving the efficacy and efficiency of those combined technologies for sequencing DNA from less DNA-friendly depositional contexts. The second is to use them to taxonomically and chronologically contextualize mammoth matriline evolution, and expand our understanding of their phylogeography to a previously uninvestigated area of their range. I also aim to assess the viability of a new targeted enrichment approach for sequencing complete nuclear genomes from mammoths and other extinct animals, even for specimens where the endogenous DNA proportion is very low and a reference sequence is unavailable for bait design.

Chapter 1

Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths

JACOB ENK, ALISON DEVAULT, REGIS DEBRUYNE, CHRISTINE E. KING, TODD TREANGEN, DENNIS O'ROURKE, STEVEN L. SALZBERG, DANIEL FISHER, ROSS MACPHEE, and HENDRIK POINAR
Genome Biology 2011, 12(5):R51; doi:10.1186/gb-2011-12-5-r51

1.1 PREFACE

Due mostly to taphonomic issues in DNA preservation, phylogenetic study of *Mammuthus* has been largely restricted to short mitogenome sequences from *M. primigenius*, the woolly mammoth. The following study is our first use of Illumina sequencing, which is particularly well suited to sequencing DNA from poorly-preserved remains and should help expand inquiry to less DNA-friendly depositional contexts. To scout its viability for future projects, we use it here to sequence the first complete mitogenome of *M. columbi,* the Columbian mammoth, after careful preparation of a DNA extract designed to reduce exogenous non-target DNA. We also use a multiplex PCR approach from my first major project in the Poinar lab and couple it with 454 sequencing to sequence the first complete mitogenome of a North American woolly mammoth. Our findings generate new hypotheses about mammoth evolution, some of which we test in later manuscripts.

1.2 ABSTRACT

Background: Late Pleistocene North America hosted at least two divergent and ecologically distinct species of mammoth: the periglacial woolly mammoth (*Mammuthus primigenius*) and the subglacial Columbian mammoth (*M. columbi*). To date, mammoth genetic research has been entirely restricted to woolly mammoths, rendering their genetic evolution difficult to contextualize within broader Pleistocene paleoecology and biogeography. Here we take an interspecific approach to clarifying mammoth phylogeny by targeting Columbian mammoth remains for mitogenomic sequencing.

Results: We sequenced the first complete mitochondrial genome of a classic Columbian mammoth, as well as the first complete mitochondrial genome of a North American woolly mammoth. Somewhat contrary to conventional paleontological models, which posit that the two species were highly divergent, the *M. columbi* mitogenome we obtained falls securely within a subclade of endemic North American *M. primigenius*.

Conclusions: Though limited, our data suggest that the two species interbred at some point in their evolutionary histories. One potential explanation is that woolly mammoth haplotypes entered Columbian mammoth populations via introgression at subglacial ecotones, a scenario with compelling parallels in extant elephants and consistent with certain regional paleontological observations. This highlights the need for multi-genomic data to sufficiently characterize mammoth evolutionary history. Our results demonstrate that the use of next-generation sequencing technologies holds promise in obtaining such data, even from non-cave, non-permafrost Pleistocene depositional contexts.

1.3 BACKGROUND

Conventional paleontological models [1-4] of North American mammoth evolution posit that at least two species occupied the continent during the later Pleistocene (150-10kya): *Mammuthus primigenius* (woolly mammoths, "WMs") evolved in Eurasia and immigrated to North America in the late Pleistocene, whereas *M. columbi* (Columbian mammoths, "CMs") evolved locally from an earlier Pleistocene immigrant ancestor (*M. meridionalis* [1, 2] or *M. trogontherii* [3, 4]). The species are morphologically differentiated by physical size (CMs were some 25% taller than WMs [5]), molar complexity (CMs displayed more 'primitive' crown height and lamellar configuration),

and skull morphology (CMs possessed a more downturned mandibular symphysis and more laterally oriented tusk alveoli) [1, 5]. Some of these traits are considered adaptations to their disparate habitats: WMs inhabited cold and arid periglacial regions, while CMs inhabited the temperate regions of the southern latitudes. Continental populations of both species went extinct during the Pleistocene-Holocene transition some 10kya.

Recent paleontological reconsiderations [6-8] and mitochondrial DNA (mtDNA) phylogeographic studies of predominantly Beringian mammoths [9-13] reveal a complex evolutionary history (Fig. 1, A). Their populations harbored diverse genetic lineages, two of which, haplogroups A and C, were endemic to Eurasia and North America, respectively. Certain population dynamics – including major immigration/replacement events and regional genetic introgression – have been offered as explanations for this complexity [10, 11], but its precise origins have proven difficult to define within the broader context of Pleistocene biogeography and paleoecology. This is the case for at least two reasons: 1) key coalescent dates remain difficult to measure, in large part due to lack of sequence breadth and methodological shortcomings [14, 15], and 2) almost nothing is known about the mtDNA phylogeny of *Mammuthus* beyond Beringian late Pleistocene mammoths (and thus probably exclusively *M. primigenius*). One potential solution to both problems – and means to hone conceptions of Pleistocene mammoth evolution in general – is to sequence DNA from one or more closely-related but distinct mammoth species and use it as a temporal and taxonomic calibration tool within the mammoth gene tree. Owing to their apparently separate evolutionary history (Fig. 1, A) and reasonably well-dated recent divergence from WMs about 1-2mya [16], CMs are excellent candidates for this role. To this end we targeted CM remains for mitogenomic sequencing.

We selected the Huntington mammoth [17] for this purpose on account of its secure morphological identification, direct radiocarbon date (11,220 ± 110^{14}Cya), exceptional biomolecular preservation [18] and geographic provenience (Fairview, UT, U.S.A.), far south of the Wisconsinan glaciers. Typical strategies for DNA sequencing of paleontological specimens would employ a pre-sequencing targeted enrichment approach, through the use of either labor-intensive PCR or hybridization techniques [19]. However, following serial extraction and library preparation of our specimen, a quantitative PCR-based metric projected a sufficient ratio of target to non-target DNA to warrant a shotgun-based metagenomic sequencing approach, for which we employed the Illumina platform (see Materials and Methods).

1.4 RESULTS AND DISCUSSION

Of the over 27 million <50bp (*published version erroneously says "greater than", corrigendum in progress*) reads obtained from the Huntington sample library, between 6 and 9 thousand (0.02-0.03%) mapped to a WM reference mitogenome [GenBank: NC007596.2, [20]] depending on software assembly parameters (Additional data file 2: table S3). This provided an average unique read depth of ~23X for the entire mitochondrial genome, excluding the VNTR region (positions 16157 – 16476). Roughly 2 million reads also mapped to the African elephant (*Loxodonta africana*) nuclear genome, providing ~0.03X coverage of the entire nuclear genome of the animal, and bringing the total likely mammoth DNA read count to ~7% of all sequences. Such a proportion of total endogenous DNA is consistent with taphonomic models for DNA preservation in temperate burial contexts [21], as well as experimental data from other non-permafrost remains [22, 23]. The coverage depth ratio we observe between mitochondrial and nuclear reads (~800X) also falls within the range estimated in other mammoth specimens (245–17,000X [24]). This low nuclear read coverage depth also lends evidence that potential Numts make no significant contribution to the consensus generated from the mitochondrial assembly.

To ensure the authenticity of the mitogenome sequence, we amplified, cloned and sequenced PCR products of WM haplotype-defining regions of the cytochrome *b* gene and hypervariable region (HVR) from multiple extractions of Huntington in two separate ancient DNA facilities. These all yielded consensus sequences 100% identical to the shotgun consensus where they overlapped. Furthermore, we sequenced the same loci from PCRs of another securely-identified *M. columbi* (the Union Pacific mammoth, University of Wyoming 6368, found near Rawlins, WY, U.S.A. [25, 26]), which yielded identical sequences to those acquired for Huntington. Finally, to control for ascertainment bias in assembly of the whole mitogenome, we mapped the Illumina sequencing reads to an Asiatic elephant (*Elephas maximus*) mitogenome [GenBank: DQ316068] and obtained a 99.98% identical consensus sequence where it overlapped with the to-mammoth assembly consensus. Thus we are confident that

the final Huntington mammoth mitogenome sequence derives from the genuine endogenous mitochondrial DNA of the animal.

Bayesian phylogenetic analysis demonstrates that the Huntington mammoth mitogenome is largely indiscernible from those of endemic North American WMs (Fig. 1, B). For all model and parameter variants (Additional data file 2: table S7; Additional data file 3: fig. S3-8), the sequence sorts securely within haplogroup C, a subclade additionally represented by dozens of WMs from Alaska and the Yukon [11]. To test for this relationship at the entire mitogenomic level, we also sequenced the first complete mitogenome of a woolly mammoth from this haplogroup (IK-99-70, from the Alaskan North Slope, U.S.A.), which confirmed Huntington's phylogenetic position within haplogroup C (Fig. 1, B).

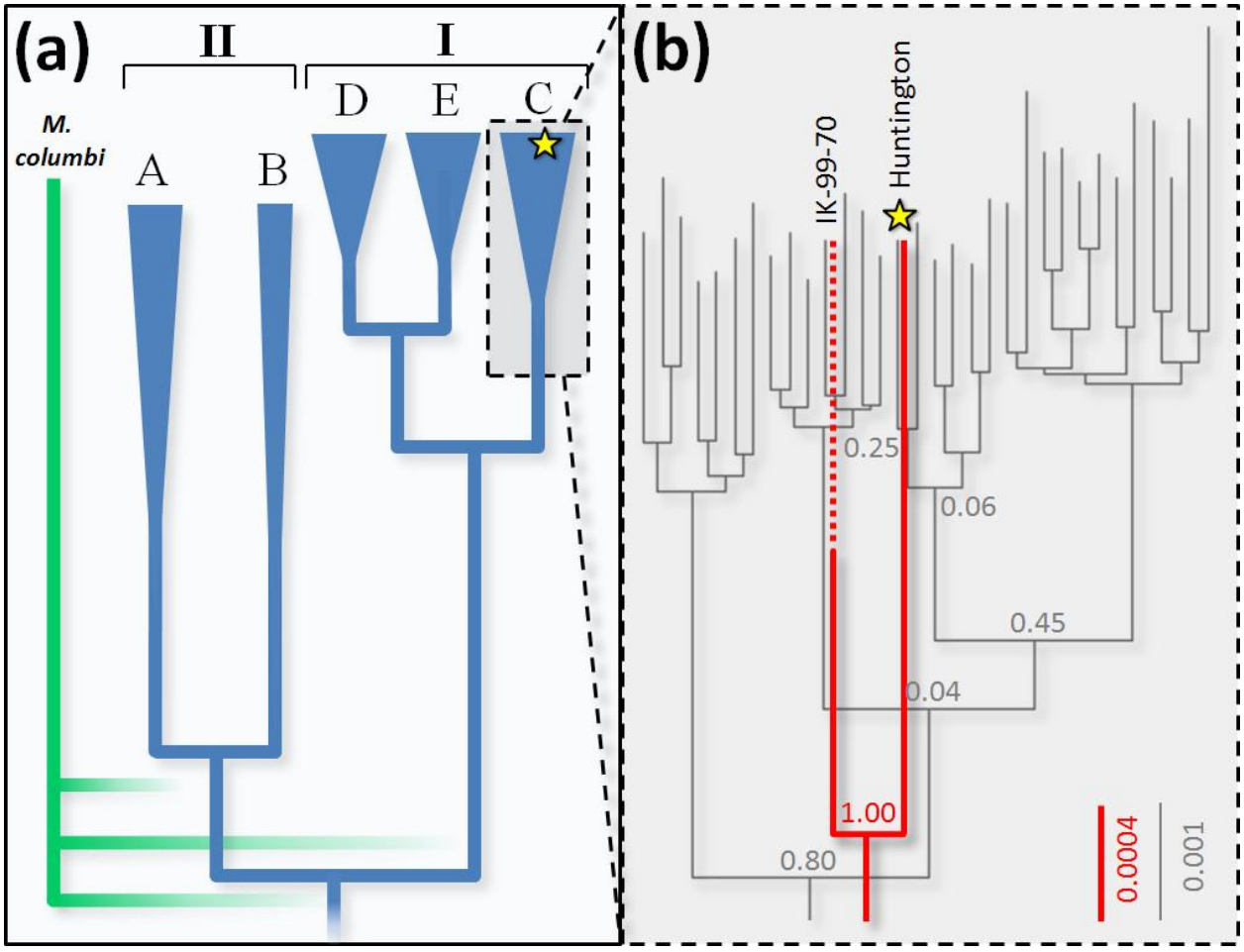


Fig. 1. Mammoth mitochondrial DNA cladograms. (a) WM lineages (blue) are summarized from previous studies [9-11] with clades indicated and haplogroups labelled at the tips. Hypothetical CM lineage positions (green) are expected positions derived from strict interpretations of paleontological models that posit the two species were separate since the early Pleistocene. The multiple node positions reflect the general uncertainty surrounding the chronology and identity of the WM lineage common ancestor. The position of WM haplogroup B is poorly resolved, exhibiting deep common ancestry with the other haplogroups. Haplogroups A and C are endemic to Eurasia and North America, respectively; haplogroups B, D, and E occur on both continents. Radiocarbon chronologies indicate that haplogroup A went extinct ~35^{14}Ckya, and clade I by ~3.2^{14}Ckya. Calculated tMRCA for all nodes yield wide confidence intervals. **(b)** Our estimated mtDNA cladograms of haplogroup C are depicted using two datasets: the black cladogram and associated scale and posterior probabilities (parameter set 1b, fig. S4) are estimated from 743bp for which several dozen mammoths have been sequenced, whereas the red cladogram and associated scale and posterior probabilities (parameter set 4b, fig. S8) are estimated from full mitochondrial

genomes, for which only one other haplogroup C mammoth has been sequenced. Each tip in the black cladogram represents a haplotype. *M. columbi* (haplotype C32) as represented by the Huntington mammoth is indicated with a yellow star. Scale units are substitutions per site.

At first glance, these results would suggest that, contrary to a strict interpretation of traditional paleontological models for their evolution, CMs and WMs did not descend from populations that were wholly separate since the early Pleistocene. One interpretation could be that mitochondrial haplogroup C corresponds to descendants of immigrant mammoth populations that ultimately gave rise to *M. columbi*. But without expansion, this interpretation would fail to explain why haplogroup C belongs to mammoths with both WM and CM morphologies. Indeed, certain paleontological interpretations have already suggested that CMs and WMs were more closely related than typically thought, even "geoclinal or chronoclinal variants" [27] descending from a very recent common ancestor. We find that our results also warrant consideration of an alternative scheme, one that operates within existing paleontological models but that accommodates incomplete reproductive barriers between WMs and CMs during some period(s) of their evolutionary history.

MtDNA phylogenies are often inconsistent with species phylogenies [28], especially for populations with sex-biased dispersion and breeding patterns. This is particularly true for extant elephants [29, 30], which exhibit male-mediated gene flow between matriarchal herds, rendering their mtDNA phylogenies incomplete representations of breeding history. For example, Asiatic elephant and WM populations both harbor(ed) at least two highly divergent mitochondrial lineages without corresponding morphological differentiation [9-11, 31]. Between WMs and CMs, we observe the opposite situation, where their morphological distinction appears to have little mitochondrial genetic correlation. One potential explanation for this is that incomplete lineage sorting (ILS) resulted in the maintenance in CM populations of what ultimately became more WM-like mitochondrial lineages. However, if this were the case, we would expect the CM-WM most recent common ancestor (MRCA) to be positioned much deeper in the cytochrome *b*/HVR phylogeny than observed. Our and previous [11] dual-calibrated estimates for the MRCA for the entirety of haplogroup C dates to the middle Pleistocene (Additional data file 2: table S7), with the CM-WM MRCA necessarily occurring much more recently, long after their purported species divergence. That said, the haplogroup C full mitochondrial dataset is too small to completely rule out ILS during CM-WM speciation as a plausible explanation.

At present, however, we suspect that hybridization between CMs and WMs may be a more parsimonious explanation for our observations. Under one conception, haplogroup C could have been a predominantly CM haplogroup that introgressed into WM populations, at such a frequency that it came to dominate the North American mitochondrial gene pool of that species. The fact that both CMs sequenced here are haplogroup C would lend some support to this hypothesis. Another possibility is that introgression occurred in the opposite direction, such that WM-typical haplogroup C introgressed into CM populations (Fig. 2, A). From a behavioral perspective, this configuration is perhaps more likely, especially in light of phenomena documented in extant African forest (*L. cyclotis*) and savanna (*L. africana*) elephants (Fig. 2, B). These living species are morphologically distinct and deeply divergent at many nuclear loci [32-35], but are known to interbreed at forest-savanna ecotones [36, 37]. The result is 'cytonuclear dissociation' [38] between genomes in hybrid individuals, such that forest-typical mitochondrial haplotypes occur at low frequency in savanna populations. Hypothetically, this is driven by savanna males reproductively out-competing physically smaller forest males [38], producing unidirectional backcrossing of hybrid females into savanna populations. Since mammoths were probably very similar to modern elephants in social and reproductive behavior [4, 27] it is conceivable that WMs and the physically larger CMs engaged in a similar dynamic when they encountered each other. Indeed, hybridization between CMs and WMs has already been suggested by others [39], and genetic exchange may explain mammoths bearing CM-WM intermediate morphologies. Such mammoths are frequently found in areas where CMs and WMs overlapped in time and space, such as the Great Lakes region [2]. Some of these apparent intermediates have been formally named (e.g., *M. jeffersonii*), but their taxonomic identity is questionable. Indeed, the large number of synonyms currently

registered for North American mammoths [40] is at least partly a function of efforts by earlier systematists to come to grips with the large amount of morphological variation expressed within *Mammuthus* (or *Elephas*). Although the Huntington mammoth exhibits no such morphological intermediacy, and was found quite distant from documented WM range, its status as a genetic hybrid would not be inconsistent with the modern analog: forest haplotype-bearing savanna elephants can be found several thousands of kilometers from modern ecotones, bearing no phenotypic indication of hybridism [38].

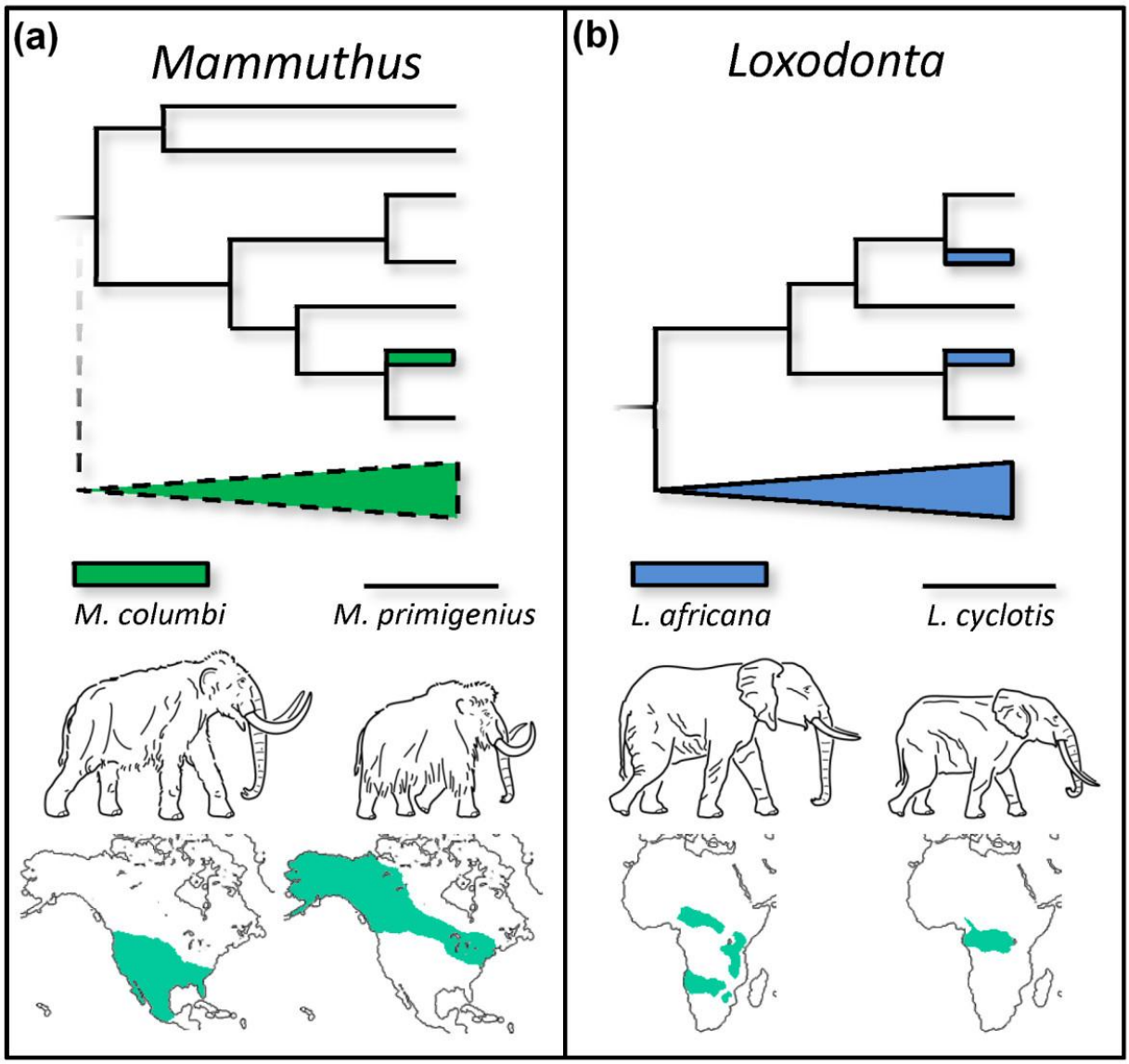


Fig. 2. Schematic representation of elephantid mtDNA phylogenies under introgression scenarios. Hypothetical mammoth (**a**, [this study]) and observed African elephant (**b**, [38]) cladograms, with male body size comparisons and predominant geographic ranges of the species indicated. Solid lines represent observed data; dashed lines represent predicted but presently unobserved lineages under an *M. primigenius* – *M. columbi* introgression hypothesis.

Both the ILS and introgression hypotheses discussed above provide straightforward testable predictions. First, under a WM-CM introgression scenario, some presently-unidentified and distinct mitochondrial haplogroup should characterize a significant percentage of CM lineages, rendering their mitogenomes polyphyletic, as they are in *L. africana* (Fig. 2). While we also observe a likely C haplotype in short sequences from one other well-identified terminal Pleistocene *M. columbi*, only a broad population-level survey of CM genetic diversity can rigorously test this prediction. Second, under the introgression hypothesis, CMs with WM-type mitogenomes should possess nuclear genes that are significantly more divergent from WMs than all haplogroup C mammoths are from each other. On the other hand, an ILS scenario would predict that CM and WM nuclear genes should show a similar degree of divergence as is detected between haplogroup C mitogenomes. Though we did recover several million nuclear sequences from the Huntington DNA library, the very low coverage depth provided by these reads is not sufficient for reliable nuclear divergence estimates between CMs and WMs. However, we anticipate that targeted enrichment techniques [41, 42] prior to high-throughput sequencing will provide the necessary coverage depth to test these hypotheses in the near future.

1.5 CONCLUSIONS

The revealed mitochondrial phylogenetic position of *M. columbi* does not immediately clarify complexities and chronological uncertainties previously observed in mammoth mtDNA phylogeny. Instead, it emphasizes that the unique reproductive behavior of elephantids necessitates a multi-genomic approach to characterizing their evolutionary history, as has been so effectively used in studies of living elephants. Their very recent mitochondrial common ancestry strongly suggests that WMs and CMs interbred at some point, most likely post-dating their morphological divergence, and in a fashion that confounds simple correlation of mtDNA phylogeny to evolutionary models derived from mammoth morphology alone. However, the precise mode and setting of genetic interchange between WMs and CMs are elusive, and therefore all hypotheses explaining our observations warrant testing. The possibility that hybridization explains our data is particularly tantalizing, since in many animals, interspecific hybridization accompanies population displacement and/or expansion resulting from habitat reconfiguration [43, 44]. Thus, interbreeding between extinct late Pleistocene taxa – especially keystone herbivores like mammoths – could serve as an indicator of major ecological events, including those surrounding the megafaunal extinctions. Our results demonstrate that the use of next-generation sequencing technologies holds promise in rigorously testing such hypotheses using full ancient genomic data, even from non-cave, non-permafrost Pleistocene depositional contexts.

1.6 MATERIALS AND METHODS

1.6.1 Samples

We included two *Mammuthus columbi* (Columbian mammoths) and one *Mammuthus* sp. in the sample set, stored at room temperature at our laboratory:

Huntington Mammoth—*Mammuthus columbi.* College of Eastern Utah Museum CEUM897 is associated with numerous radiocarbon dates, though 11220 ± 110 ^{14}C yr BP is probably most accurate [17]. It was discovered in 1988 during excavation of a stream for dam construction, at the southeast end of what is now Huntington Reservoir, just east of Fairview, Utah, USA. This 60+ year old bull is exceptionally well-preserved, and exhibits the classic character suite of his species, including low molar lamellar frequency (Additional data file 3: figure S1), broadly divergent tusk alveoli, a markedly downturned mandibular symphysis, and tremendous body size. We used tusk fragments for the shotgun sequencing, and both tusk and bone samples for PCR and Sanger sequencing.

Union Pacific Mammoth—*Mammuthus columbi*. University of Wyoming UW6368 is dated to 11280 ± 350 ^{14}C yr BP [25, 26]. It was discovered in 1960 by a gas well-drilling crew while drag-lining a spring site southwest of Rawlins, Wyoming, USA. Fragments of molar teeth were used for PCR and Sanger sequencing.

MPC IK-99-70—*Mammuthus sp.* Specimen found in the Upper Ikpikpuk River (70° 47'N, 154° 25'W) on the Alaskan North Slope of the USA. Provenience strongly suggests that it is *M. primigenius*. Radiocarbon dated to 41,510 ± 480 ^{14}C yr BP (Beta #264909, Beta Analytic Inc., Miami, FL, USA). The mitochondrial hypervariable region for this

specimen was partially sequenced previously [11] and falls within haplogroup C (haplotype C30). Its exceptional DNA preservation prompted its use in the multiplex experiments.

1.6.2 Sequence Acquisition, Assembly, and Classification

Detailed descriptions of wet laboratory procedures used for sequence acquisition, as well as laboratory procedures for data assembly, can be found in Additional data file 1. Pre-sequencing preservation evaluations were performed following [45] and [46]. We used a metagenomic high-throughput sequencing approach to characterize the whole mitochondrial genome of the Huntington mammoth, and multiplex PCR combined with high-throughput sequencing to obtain the whole mitochondrial genome of IK-99-70. We also cloned and sequenced several PCR products from the mammoths, with independent PCR amplification, cloning and sequencing of products from the Huntington mammoth performed at a separate laboratory. We assembled mitochondrial reads with AMOScmp [47] using NUCmer [48] as well as with Geneious 5.1.7 (http://www.geneious.com) and then visualized assemblies using amosvalidate [49], Hawkeye [50] and Geneious. The Huntington nuclear genome read assemblies were built using these and also classified using PhymmBL [51], comparing previously-published woolly mammoth nuclear genome sequences [52]. Sequence read files for Huntington and IK-99-70 are deposited in the NCBI Short Read Archive (SRA) as #SRP006656. Sanger trace files from Huntington, Union Pacific, and IK-99-70 are deposited in the NCBI Trace Archive as #TI2306523713-2306523816. Consensus mitochondrial sequences are deposited in GenBank as #JF912199 (Huntington) and #JF912200 (IK-99-70). Our assemblies of Huntington, IK-99-70, and Union Pacific reads and traces are available at ftp://ftp.cbcb.umd.edu/pub/data/mammoth/.

1.6.3 Phylogenetic Analyses

Detailed description of phylogenetic analyses performed can also be found in Additional data file 1. , These explored topological and chronological features mammoth mitochondrial phylogeny using a Bayesian approach, comparing hundreds of sequences from a number of studies discussed above as well as from [53]. We employed jModelTest v. 0.1.1 [54] to choose model parameters and BEAST v. 1.5.6 [55] to build trees and estimate coalescent dates, using tip calibration points corresponding to radiocarbon ages of the samples, as well as root calibration points described by [56]. These runs were analyzed in Tracer v.1.3 (http://tree.bio.ed.ac.uk/software/tracer/) and trees were visualized with FigTree v.1.3 (http://tree.bio.ed.ac.uk/software/tracer/).

COMPETING INTERESTS

The authors declare no competing interests in the publication of this data and manuscript.

AUTHOR CONTRIBUTIONS

JE, RD, DO, DF, RM and HP planned and designed the project. JE, RD, AD and CK performed wet laboratory work. JE, RD, AD, TT, and SS performed assembly and analysis of the sequence data. JE wrote the paper with assistance from all authors.

FIGURES

1.7 ADDITIONAL DATA FILES

The following additional data are available with the online version of this paper. Additional data file 1 is a detailed description of materials and methods used in this study as well as associated literature cited. Additional data file 2 is a collection of tables referred to in the text as tables S1 through S7 as well as associate literature cited. Additional data file 3 is a collection of figures referred to in the text as figures S1 to S8 as well as literature cited. Additional data file 4 is an assembly of all reads acquired from the Huntington mammoth sequencing library that were used in construction of the consensus mitogenome. Additional data files 5 and 6 are the first and second replicate assemblies, respectively, of sequence data for IK-99-70. Additional data file 7 is an assembly of Sanger sequences from Huntington and Union Pacific. Additional data file 8 is an assembly of all sequence reads from Huntington that aligned successfully to the *L. africana* nuclear genome**.** Additional data files 4-6 and 8 are available at ftp://ftp.cbcb.umd.edu/pub/data/mammoth/.

1.8 ACKNOWLEDGMENTS

We thank the College of Eastern Utah Museum, especially J. Bartlett, J. Bird, and R. Barlow for providing access to and valuable information regarding the Huntington Mammoth; the University of Wyoming Vertebrate Paleontology Collections, especially M. Clementz and B. Breithaupt, for providing access to and information regarding the Union Pacific Mammoth; and P. Matheus, G. Zazula, and D. Froese of the Yukon Department of Tourism and Culture for access to IK-99-70. Ambry Genetics provided valuable assistance in experimental design. B. Miljour drew the mammoth and elephant images in fig. 2. Funding was provided by the Natural Science and Engineering Research Council (grant #299103-2004), the Social Sciences and Humanities Research Council (grant #410-2004-0579), the Canadian Research Chairs program, Discovery Communications (a Quest grant), the National Science Foundation (EAR-0545095 to DF; GRF to JE), the Ontario Graduate Studies programme (scholarship to AD), the University of Utah, the American Museum of Natural History, McMaster University, and Eco-Anthropologie, Muséum national d'Histoire naturelle (UMR 7206). We would like to dedicate this to the late Paul S Martin, a great colleague and friend.

1.9 ADDITIONAL DATAFILE 1 – MATERIALS AND METHODS

1.9.1 Laboratories

McMaster Ancient DNA Centre ("MAC;" McMaster University, Hamilton, Ontario, Canada) http://socserv.mcmaster.ca/adna/index.htm: Sample extraction, qPCR and PCR reactions, cloning, and Sanger sequencing reactions on Huntington, Union Pacific, and MPC IK-99-70; library preparation for and sequencing on the 454 GSFLX (454 Life Sciences, Brantford, CT, USA) of MPC IK-99-70; sequence assembly and analyses.

Institute for Molecular Biology and Biotechnology Laboratory ("Mobix;" Hamilton, Ontario, Canada) http://www.science.mcmaster.ca/mobixlab/: Sanger sequencing on ABI-3730 DNA Analyzer (Applied Biosystems) of sequencing reactions generated at MAC.

Service de Systématique Moléculaire of the Muséum national d'Histoire naturelle ("MNHN;" Paris, France) http://www.mnhn.fr/mnhn/smo/: Replication experiments on Huntington, including sample extraction, PCR reactions, cloning, and Sanger sequencing on ABI-37 DNA analyzer (Applied Biosystems Inc., Foster City, CA, USA).

Ambry Genetics ("Ambry;" Aliso Viejo, California, USA) http://www.ambrygen.com/: Illumina library preparation and evaluation, library qPCR, high throughput sequencing on the Illumina GAII platform (Illumina Inc., San Diego, CA, USA).

Center for Bioinformatics & Computational Biology ("CBCB;" University of Maryland, College Park, Maryland, USA) http://www.cbcb.umd.edu/: High-throughput sequence assembly and analyses.

1.9.2 Operating procedures

Pre-PCR laboratory work was performed in dedicated cleanroom facilities (MAC and MNHN), following standard protocols for attire and handling [21]. Pre- and post-PCR work were performed in physically separate laboratories. Control reactions containing no mammoth material were used in all appropriate experiments (extraction and PCRs) in order to detect contamination by previously extracted or amplified mammoth DNA. Any experiments with positive amplification in blank reactions were wholly excluded from downstream use. All PCR products were cloned before Sanger sequencing. Consensuses of Sanger sequences were derived from at least two clones from each of at least two replicate PCR products from the same extraction.

Primers used in Materials and Methods sections 4 and 5 (Additional data file 2: table S1) have been published previously [11] or were newly designed using the Integrated DNA Technologies (IDT) SciTools OligoAnalyzer 3.1 (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/). These were provided by IDT (Coralville, IA, USA; 25nmol, standard desalting) and tested for sensitivity and PCR conditions in their working pairs using copy number standards of a straight or cloned *M. primigenius* PCR products of known sequence (hapolotype D1). Primers used

in section 6 (Additional data file 2: table S5) have been published previously [20] or were newly designed, ordered from IDT, and optimized in section 6b.

"*Taq*" and "PCR Buffer" hereafter refers to reagents supplied with the Ampli*Taq* Gold® (Applied Biosystems Inc., Foster City, CA, USA) DNA polymerase.

"SYBR" refers to SYBR Green® (Invitrogen, Carlsbad, CA, USA).

"BSA" refers to bovine serum albumin.

"Water" used in laboratory experiments refers to UV-sterilized, purified water.

For all mitogenomic base positions, we use the sequence obtained by Krause et al. [22] as a reference.

Sequence alignments are available in the Additional data files 4-8.

1.9.3 Sample Selection

Please see the main text for description of specimens.

1.9.4 Huntington Mammoth Whole Mitochondrial Genome

1.9.4a DNA Extraction

Roughly 0.98g of tusk material from Huntington was sequestered and crushed to fine particles with a hammer. These were demineralized and digested in separate steps using buffers and incubation temperatures/durations described elsewhere [11]. This was repeated in four separate rounds of demineralization+digestion (5+5mL, 5+5mL, 3+3mL, 2+3mL) on the same tusk material, though we only used solutions from the final round for all subsequent work. Demineralization and digestion supernatants from this last round were pooled and then purified using phenol:chloroform:isoamyl alcohol (25:24:1), and resultant aqueous phases were purified again with chloroform. The final aqueous phase was concentrated by ultrafiltration with Amicon Ultra 30K columns (Millipore) and eluted in 150µL 0.1X TE (pH 8.0). We hereafter refer to the purified extract from this last round of demineralization/digestion as "*HUNT1*."

1.9.4b Extraction qPCR Screen

We used a quantitative PCR of a 79bp mammoth-specific amplicon (using primers #3+4, Additional data file 2: table S1) to estimate the amount of target mammoth DNA in *HUNT1* and screen for contamination in its associated extraction blank.

Each 20µL qPCR included: 1X PCR buffer, 2.5 mM $MgCl_2$, 1mg/mL BSA, 250 µM each dNTP, 200 nM each primer, 2.5 units of *Taq* polymerase, 0.167X SYBR, 3µL of template DNA extract (straight and 0.1X dilutions), water for PCR blanks, or 0.1X TE used for sample dilution. Five mammoth DNA standards of known concentration (1 to 1,000 copies/µl) were included for each amplicon. Cycling conditions were: initial denaturation (95°C, 5m); 55 cycles of denaturation (95°C, 30s), annealing (62°C, 30s), and extension (72°C, 40s). Amplifications were executed and analyzed using the BioRad CFX96® (Biorad, Hercules, CA, USA) real-time PCR platform and associated software.

No mammoth DNA contamination was detected in the extraction blank. Estimated starting copy numbers of the 79bp fragment from *HUNT1* were ~213 (estimation from 1X concentration) and 238 (0.1X projection) copies per original microliter. Following the formula decribed by Schwarz *et al*. [24], this indicates only about 10% PCR inhibition, significantly lower than the same measures derived from another extraction from the same substrate (*HUNT2*, section 5e). Since *HUNT1* derives only from the last round of demineralization and digestion applied to the sample, and *HUNT2* from the first (and only) round from a different subsample of the same substrate, we expect that many of the inhibitory constituents associated with this specimen were removed in the first three

rounds of the extraction described in section 4a. While this is purely hypothetical, such a strategy (pre-demineralization/digestion) may therefore prove useful for highly inhibited samples of various kinds.

1.9.4c DNA Size Distribution Measurement

We performed a qPCR-based evaluation amplifiable length distribution, and thus DNA fragmentation, in *HUNT1* following the procedure designed by Deagle *et al.* [46] and Schwarz *et al.* [24]. We amplified three incrementally longer amplicons within the woolly mammoth mitochondrial 12S gene, using a single forward primer (#15) and three different reverse primers (#16-18) (Additional data file 2: table S1), in individual singleplex reactions.

Each 20µL qPCR included: 1X PCR buffer, 2.5 mM $MgCl_2$, 1mg/mL BSA, 250 µM each dNTP, 200 nM each primer, 2 units of *Taq*, 0.167X SYBR, 3µL of template DNA extract (0.1X dilution) or water for PCR blanks or 0.1X TE used for sample dilution. Four mammoth DNA standards of known concentration (1 to 1,000 copies/µl) were used for each amplicon, in replicate. Cycling conditions were: initial denaturation (95°C, 5m); 50 cycles of denaturation (95°C, 25s), annealing (62°C, 25s), and extension (72°C, 25s). The extract was amplified in triplicate. Amplifications were executed and analyzed using the BioRad CFX96® real-time PCR platform and associated software.

Estimated starting molecule counts per original microliter (averaged among triplicates) are reported in Additional data file 2 (table S2). These were log transformed and plotted against fragment length. From this, a regression line was calculated, the slope of which (λ) correlates to the rate of DNA fragmentation in the extract, and the inverse of which estimates the average amplifiable fragment size in the sample, and finally the x-intercept of which (*y=0*) estimates the maximum amplifiable fragment size. Additional data file 2 (table S2) reports the results of this evaluation, with the results from similar evaluations of other extracts used in this project and of woolly mammoth remains analyzed by Schwarz *et al.* [24] for comparison. A graphical presentation of log-transformed plots is shown in Additional data file 3 (figure S2).

These results, discussed further in Section 5e, guided size selection during library preparation.

1.9.4d Library Preparation

We sent *HUNT1* to Ambry Genetics for library preparation for the Illumina (Illumina Inc., San Diego, CA) platform, for which they followed the standard protocol with some proprietary modifications. Because our qPCR evaluation of the extract suggested a maximum amplifiable fragment length of only ~150bp (see above), Ambry selected only the size fraction that included 50-125bp original fragment length for sequencing. While selecting a narrower size fraction (e.g., 50-70bp) may have at least theoretically maximized the ratio of target:nontarget DNA, we expanded our selection window to balance our target sequencing goal with a desire to explore the metagenomic content of the sample for taphonomic purposes, as well as the capacity of the sequencing platform. Ambry then finalized the library enrichment by amplifying the mentioned size fraction, which provided a final total DNA content of roughly 9.28ng/ul in the library, as measured with an Agilent 2100 Bioanalyzer (Agilent).

1.9.4e Library qPCR Evaluation

In order to estimate the amount of target DNA in the library prepared by Ambry, they performed a quantitative PCR experiment using an identical protocol used in the size distribution measurement, targeting just the smallest amplicon in that set (primers #15+16, table 1, Section 4c). This provided an estimate of about 3750 copies per microliter of the 63bp target in the library. Assuming that the qPCR measured only template molecules that were fully adapted, this should correspond to a molecular weight of the target amplicon of at least 5.95E-7ng/µL. By dividing the length of a whole mitochondrial genome (~16,800bp) into 63bp fragments (=~267), we conservatively estimated that ~800k similar fragments, totaling 1.55E-4ng/µL, should comprise target DNA. Since this total molecular weight is roughly 1.71E-3% of the total DNA concentration, then we predicted that at least that proportion of the eventually sequencing reads would be target. If we acquired roughly 20m reads in the sequencing run, this projects that we would obtain about 342 total 63bp reads, which would amount to roughly 1.3X *duplicate* coverage of the whole mitochondrial genome. While this model and projection clearly does not account for the shorter average read length projected from the size distribution measurement (~35bp, section 4c),

we predicted that it would only underestimate the eventual read coverage. Therefore we opted to sequence the library without any further enrichment.

1.9.4f Sequencing with Illumina

The aforementioned library size fraction was sequenced using the 54bp singleton protocol on the Illumina GAII platform (Illumina Inc., San Diego, CA, USA), following standard procedures with slight modification by Ambry.

1.9.4g Data Processing & Statistics

Ambry performed preliminary data processing, including quality filtering and base calling. A total of ~28.5m final reads were generated from the experiment. Final FASTQ sequence read data files were sent to MAC and CBCB for subsequent assembly and analysis.

1.9.4h Sequence Assemblies

Three software programs were used in sequence assembly, each using slightly different protocols.

FASTQ read files were first converted to FASTA format and then assembled to the mammoth reference genome [20] using the 454 GS Reference Mapper software (454 Life Sciences, Brantford, CT, USA) under the default parameters. A total of 7784 (7614 unique) reads successfully aligned, resulting in a 22.5X average read depth per base, with at least 2X unique read depth for all bases except the seven most 5' bases and the nine most 3' bases of the reference, as well as some sections of the VNTR. This corresponds to ~1.02X duplicate read depth per base, lower than expected 1.3X (section 4e). Consensus contigs were generated following assembly using the default parameters.

We also assembled the mitochondrial reads with AMOScmp, [47] a program designed for comparative sequence assembly. The AMOScmp-shortReads pipeline was used, specifically designed to handle cases with short (< 100 bp) reads. AMOScmp first aligns the reads to a reference sequence with NUCmer [48], a widely-used alignment program for efficient pairwise DNA alignment. AMOScmp-shortReads parameters were configured as follows: MINCLUSTER = 16, MINMATCH = 16, MINLEN = 31, --MAXMATCH, MINOVL = 10, MAXTRIM = 18, MAJORITY = 50, CONSERR = 0.06, ALIGNWIGGLE = 2. These values resulted in 8048 reads mapping successfully to the reference mammoth genome. The average depth of coverage (in unique reads) of this assembly was 23X. The consensus was generated using the AMOS make-consensus program. In attempt to increase sensitivity, we also used the AMOScmp-shortReads-alignmentTrimmed pipeline, which increased the total mapped reads to 10962 and read coverage to 30X. The final assembly was manually inspected using amosvalidate [49] and the graphical assembly viewer Hawkeye [50]; visual inspection revealed 39 identical clones assembled to positions 1601 – 1628, which were identical to the reference in the first 20 bases but included an insertion and C>T transition in the latter portion (GTTGGCTTGGAAGCAGCCATTCATTTAA). Upon a BLAST search (http://blast.ncbi.nlm.nih.gov/) of this sequence, it was found to be a 100% match to several bacterial entries. This combined with the deep clonal depth lend evidence the sequence derives from non-endogenous DNA, and thus we manually removed them from the assembly. The final assembly contained two contigs gapped by the VNTR region (positions 16157 – 16476). Gaps in alignment were closed by evenly distributing aligned reads across the region, in agreement with the average depth of coverage for the assembly.

Reads from the FASTQ files were also assembled in Geneious 5.1.7 (http://www.geneious.com) to the mammoth reference and to an Asian elephant mitochondrial genome [20] using both the "low" and "medium" sensitivity levels, with no alignment fine tuning. The resultant to-elephant assemblies included several sections with no read coverage, which is consistent with the significant divergence between that genus and *Mammuthus* as assayed for woolly mammoths. From these assemblies we generated consensuses using the strict 50% threshold, with the highest quality score from all reads aligned to each base used to determine base quality, and with "N" assigned to those bases with quality scores less than 20.

A comparison of all assembly consensuses (Additional data file 2: table S3) reveals broad agreement between the to-mammoth assemblies and the medium sensitivity to-elephant assembly consensuses for those bases where they overlap. However, this to-elephant assembly consensus yields a number of disagreements with the other consensuses in a short region (positions 13707-13750), derived from a single read for that section (not included in Additional data file 2 (table S3). While these base calls are thus unlikely to be accurate, they technically bring the identity between this consensus and the to-mammoth consensuses to 99.98% excluding the VNTR. The low sensitivity to-elephant assembly, on the other hand, revealed a number of disagreements with the other assemblies, again largely corresponding to base positions with low coverage.

Broad agreement among consensuses generated from various software packages and parameters lead us to use the consensus generated from the AMOScmp assembly (Supplementary Materials: Alignment 1) as the final sequence for the Huntington mammoth. However, owing to the short read lengths, we consider the VNTR (positions 16157 – 16476) unresolvable with this data, and thus remove it from our reported consensuses, following Gilbert *et al*. [10].

1.9.5 Sanger Sequencing of Huntington and Union Pacific mtDNA

1.9.5a DNA Extraction

We subsampled 100mg of bone (Huntington) and tooth (Union Pacific) material using bleach- and heat-sterilized tools (chisel and/or rotary tool) and then crushed them to fine particles/powder with a hammer. Samples were extracted according to procedures described elsewhere [11], except that demineralization and digestion supernatants were combined prior to the PCI purification stage. Aqueous phases of PCI processing were concentrated and reconstituted in 50µL 0.1X TE (pH 8.0) by ultrafiltration with Microcon YM-30 columns (Millipore, USA). Extractions included blanks to detect contamination. These we refer to as *"HUNB1*" for the Huntington bone and "*UPT1*" for the Union Pacific molar tooth extractions.

A second round of extractions used 1g of tusk from Huntington ("*HUNT2*") and 0.68g of tooth root from Union Pacific ("*UPT2*"). We used these extracts for a preservation evaluation as well as to provide a single amplicon for sequencing (see below). This extraction protocol followed the same procedure referenced above, except for relative volume increases of EDTA (10ml), and digestion buffer (7–8ml). Final aqueous phases were concentrated to a final elution volume of 100µL.

1.9.5b PCR Amplification

HUNB1, *HUNT2*, and *UPT1* were used for PCR amplification of specific targets for cloning and Sanger sequencing.

Each 20µL PCR included: 1X PCR buffer, 2.5 mM $MgCl_2$, 1mg/mL BSA, 250 µM each dNTP, 200 nM each primer, 5 units of *Taq*, 0.167X SYBR, 3-5µL of template DNA extract (1X or 0.1X dilutions) or water for PCR blanks. At least one PCR blank was included per PCR reaction. Cycling conditions were: initial denaturation (95°C, 4-7m); 45-60 cycles of denaturation (95°C, 30s), annealing (59.5-62.5°C, 30s), and extension (72°C, 40s); and a final extension (72°C, 10m). Each extract was amplified in duplicate. Amplifications were executed and analyzed using the Stratagene Mx3000P® real-time PCR platform.

Amplification products were run on 2% (w/v in 0.5X TBE) agarose gels containing 1% ethidium bromide at a concentration of 1.5µL/50mL and visualized with UV light. No contamination was observed in any reaction blanks. When primer dimers and secondary products were apparent, 10 µL were rerun on 2% gels and the target amplicon was excised and dissolved in 50µL 1X TE. These were re-amplified using 2µL of the dissolved gel solution under the same conditions as the original PCR, with 1 unit of *Taq*. See Additional data file 2 (table S4) for a summary of amplification conditions and results.

1.9.5c Cloning and Sequencing

All PCR products were cloned using the TOPO-TA cloning kit with "One Shot" TOP-10 chemically-competent cells (Invitrogen), using one-quarter scale reactions. Prior to cloning, older PCR products were re-adenylated in

reactions containing 250µm dATP mix, 2.5mM $MgCl_2$, 1X PCR Buffer, 1 unit *Taq* , water, and 3µL of PCR product, at 94°C for 7 minutes and 72°C for 20 minutes.

Colonies were chosen using blue/white selection from culture plates (containing 50mg/L ampicillin and 20µg/mL X-Gal), lysed in 50µL of 10mM Tris-Cl at 95°C for 5 minutes. 30µL colony PCR reactions were performed containing 2µL of colony lysate supernatant with "M13" primers [following [11] with a 57°C annealing temperature], and purified over 96-well 30K Acroprep filter plates (Pall Corporation, Port Washington, NY, USA) into 0.1X TE. 7µL sequencing reactions were performed following [11] using 0.3µL BigDye terminator v1.1 (Applied Biosystems), 1µL of diluted purified colony PCR product, 0.1X TE, water and the M13 forward primer. Cycled reactions were sequenced by Mobix.

Cloned sequences were manually aligned in BioEdit v7.0.9, and consensus at each position was determined by at least two clones from at least two independent PCRs. We sequenced additional clones from the same and/or additional PCRs until consensus was achieved. Sequence alignments can be found in Supplementary Materials: Alignment 4.

1.9.5d Sequence Replication

In order to independently verify the sequences obtained at MAC, two subsamples from the Huntington mammoth (one from bone the other from tusk) were sent to MNHN, where they were analyzed in the "ancientDNA Box". Roughly 100mg of each sample were processed using the same extraction procedure described in section 5a. Only a few steps diverge from the extraction protocol used at MAC: 10% sarcosyl was replaced with 2% pre-warmed SDS, and Acroprep 30K columns (Pall, USA) were used for final concentration of the extracts in 100µL of 1X TE.

Two successful amplifications performed at MAC were attempted at MNHN: primers #3+4 and #9+12 (Additional data file 2: table S1). Straight amplifications from the extracts remained unsuccessful after 3 trials. However, after the implementation of a multiplex 15-cycle pre-amplification using a combination of the four primers, secondary PCRs using 2 µL of straight pre-amplified products yielded positive products for each amplicon twice independently. PCR conditions were identical to those used at MAC except that they were performed using the SsoFast supermix (Biorad, Hercules, CA, USA) and with only 0.4mg/mL BSA. Amplification products were blunt-ended and individually cloned using the TransformAid and CloneJET PCR cloning kits (Fermentas, Estonia). Positive colony PCR products were subsequently sequenced on the ABI-37 DNA analyzer (Applied Biosystems). Four separate clones, from 2 independent PCR for each of the two fragments were sequenced and yielded the same consensus sequence as the one obtained at MAC for both amplicons.

1.9.5e Preservation Evaluation

HUNT2 and *UPT2* were evaluated for DNA preservation in a similar fashion as was done on *HUNT1* (section 4c). However, these qPCRs used a 63bp target (primers #6+8), an 85bp target (#5+7), and a 121bp target (#5+8) (Additional data file 2: table S1).

Each 20µL PCR included: 1X PCR buffer, 2.5 mM $MgCl_2$, 1mg/mL BSA, 250 µM each dNTP, 200 nM each primer, 1-2 units of *Taq*, 0.167X SYBR, 3µL of template DNA extract (1X, 0.1X, and 0.02X dilutions) or water for PCR blanks. Five mammoth DNA standards of known concentration (1 to 10,000 copies/µl) were added as standards for each amplicon. Cycling conditions were: initial denaturation (95°C, 7m); 45 cycles of denaturation (95°C, 30s), annealing (61°C, 30s), and extension (72°C, 40s); and a final extension (72°C, 10m). Amplification properties were evaluated using the analysis software provided with the Stratagene Mx3000P® real-time PCR platform.

When combined with similar analyses performed on the extracts used in the Illumina sequencing (section 4c, Additional data file 2: table S2, Additional data file 3: figure S2), these data demonstrate that the Columbian mammoth extracts, as expected for remains from temperate contexts, exhibit significantly more highly fragmented DNA compared to permafrost-preserved woolly mammoth extracts.

1.9.6 Whole Mitogenome of IK-99-70

To obtain the whole mitogenomic sequence from this specimen, we employed a technique similar to Krause *et al.* [20], using three groups of primer pairs in multiplex amplification, followed by amplification from the multiplexes using each pair individually ("singleplex"), followed by purification, pooling, and sequencing on the Roche 454 GS-FLX (454 Life Sciences, Brantford, CT, U.S.A). To fill gaps in the 454 data, we used multiplexes, singleplexes, and/or original extract to reamplify necessary amplicons, which we then cloned and Sanger sequenced using similar procedures as in Section 5c.

1.9.6a DNA Extraction

We used the same extract generated by Debruyne *et al.* [11] at 0.5X or 0.25X concentration (diluted in 0.1X TE) for all multiplex and fill-in reactions contributing to the final sequence for the specimen.

1.9.6b Multiplex Primer Design & Optimization

Most primers used for this section of the project were taken directly from Krause *et al.* [20], while others we designed using the Integrated DNA Technologies (IDT) SciTools OligoAnalyzer 3.1 (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/), as reported in Additional data file 2 (table S5). These comprised 47 overlapping pairs, 46 of which were initially divided into two groups (A and B) such that no pairs in the same group would amplify overlapping fragments. Later, group A was divided into two separate groups (A-1 and A-2) based on optimal annealing temperature, discussed below.

In order to determine optimal annealing temperatures for the multiplex reactions, we performed a series of PCRs (including a fluorescent dye but without DNA standard) on an exceptional mammoth extract [45] using each multiplex primer pool.

Each 20µL PCR included: 1X PCR buffer, 2.5 mM $MgCl_2$, 1mg/mL BSA, 250µM each dNTP, 150nM each multiplex primer pool (each primer equilibrated), 2.5 units of *Taq*, 0.167X SYBR, 4µL of template DNA or water for PCR blanks. Cycling conditions were: initial denaturation (94°C, 9m); 30 cycles of denaturation (94°C, 30s), annealing (52-65°C, 30s), and extension (72°C, 30s); and a final extension (72°C, 10m), concluding with a melt curve from 68-95°C. Amplification properties were evaluated using the analysis software provided with the Stratagene Mx3000P® real-time PCR platform.

PCR products were also run on ethidium bromide-stained 2% agarose gels. Three properties contributed to determination of optimal annealing temperature: amplification Cq as determined by the analysis software, peak fluorescence, and visual quality/intensity of the products on agarose gels. Multiplex group A did not amplify as well as group B, therefore we split group A into two subgroups (A-1 and A-2, Additional data file 2: table S5) based on theoretical expected annealing temperatures, and determined the optimal annealing temperatures for these subgroups using identical procedures as outlined above.

We performed similar optimization experiments for each individual primer pair, amplifying from dilutions of the associated multiplexes generated beforehand.

Each 20µL PCR included: 1X PCR buffer, 2.5 mM $MgCl_2$, 1mg/mL BSA, 250µM each dNTP, 1.5µM each primer, 0.5 units of *Taq*, 0.167X SYBR, and 2µL of a 0.025x dilution (in water) of the appropriate multiplex pool (optimal temperature-amplified reaction only) or Water for PCR blanks. Cycling conditions were: initial denaturation (94°C, 9m); 35 cycles of denaturation (94°C, 30s), annealing (54-65°C gradient, 30s), and extension (72°C, 40s); concluding with a melt curve from 55-95°C. Amplifications were executed and analyzed using the BioRad CFX96® real-time PCR platform and associated software.

Amplification properties were evaluated using identical metrics as described above for the multiplexes, which allowed us to determine appropriate annealing temperatures for each primer pair, indicated in Additional data file 2 (table S5). Primer pair A8a is the only exception, as it was designed for amplification solely from extract and its optimal annealing temperature was not determined. Rather we amplified this pair using the indicated (Additional data file 2: table S5) annealing temperature, which was slightly above its calculated melting temperature.

1.9.6c Multiplex Amplification

All multiplex amplifications were performed in duplicate, using the optimal annealing temperatures determined from the experiments outlined above.

Each 20µL PCR included: 1X PCR buffer, 2.5 mM $MgCl_2$, 1mg/mL BSA, 250µM each dNTP, 150nM each multiplex primer pool with each individual primer at 1uM, 2.5 units of *Taq*, 0.167X SYBR, 3.5-4µL of template DNA extract (at 0.5X concentration) or water for PCR blanks. Cycling conditions were: initial denaturation (94°C, 9m); 28-30 cycles of denaturation (94°C, 30s), annealing [57°C, 60°C or 58.5°C (multiplex groups A-1, A-2, or B, respectively), 30s]; and extension (72°C, 40s), concluding with a melt curve from 65-93°C. Amplifications were executed and analyzed using the BioRad CFX96® real-time PCR platform and associated software.

Successful reactions were subsequently diluted to 0.04X concentration (in water) and used as template for subsequent singleplex reactions.

1.9.6d Singleplex amplifications

Using each replicate of each multiplex as template (at 0.04X concentration, diluted in water), we targeted amplicons using each primer pair in individual amplifications.

Each 20µL PCR included: 1X PCR buffer, 2.5 mM $MgCl_2$, 1mg/mL BSA, 250 µM each dNTP, 1500 nM each primer pair, 0.5 units of *Taq*, 0.167X SYBR, 2µL of a 0.04X dilution (in Water) of the appropriate multiplex, or Water for PCR blanks. Cycling conditions were: initial denaturation (94°C, 9m); 29-35 cycles of denaturation (94°C, 30s), annealing (each appropriate temperature (Additional data file 2: table S5), 30s), and extension (72°C, 40s); concluding with a melt curve from 68-91°C. Amplifications were executed and analyzed using the BioRad CFX96® real-time PCR platform and associated software.

Successful PCRs were subsequently purified over 96-well 30K Acroprep filter plates (Pall Corporation, Port Washington, NY, USA) and eluted in 40µL 0.1X TE. Since one amplicon (A3) showed two distinct melt peaks and gel bands, we ran the entire product on 2% agarose gel, plugged and kept the larger (target-sized) band, which we then purified using the QIAquick™ Gel Purification Kit (QIAGEN) eluting in 50µL EB.

In total this provided 92 total purified amplicons, 46 in duplicate from independent multiplex reactions. We attempted to quantify each purified amplicon using the quantitative plate read function of the Mx3000P® real-time PCR platform on a solution of each purified amplicon and 0.167X SYBR, with a standard of known DNA quantity (Quant-iT™ PicoGreen®, Molecular Probes Inc.). We used these quantitations to pool the products in ostensibly equimolar concentrations. However, from the sequence data, it became clear that the quantitation procedure required further optimization (see below).

1.9.6e Library Preparation

Each replicate amplicon pool was concentrated using Microcon YM-30 columns (Millipore, USA) to 50µL in 10mM Tris-HCl. We used 26µL of this with the Roche/454 GS FLX Titanium shotgun library preparation kit (454 Life Sciences, Brantford, CT, USA), replacing the standard A-adaptor with those having a 10bp multiplex identifier tag. We subsequently quantified the number of adapted molecules using a qPCR procedure described elsewhere [57], which allowed us to determine the necessary amount to use in emulsion PCR. Since these samples accompanied others on the same sequencing lanes, we only added enough, by our estimates, necessary for 20X clonal coverage of each amplicon.

1.9.6f Sequencing on 454 GS FLX & Data Processing

Pooled libraries underwent emulsion PCR, and emulsion beads bearing successful amplification were isolated for sequencing. These were packed onto a 2-region PicoTitrePlate™ and sequenced on the 454 GS-FLX sequencer. Sequencing image files were processed using the standard shotgun image processing algorithm included in the GS

software package (v2.1). This provided two individual SFF files containing those reads that passed the default quality filters.

1.9.6g Sequence Assembly & Fill-In

SFF files from the two regions, which each contained a replicate amplicon pool, were sorted by their appropriate MID (allowing 2 errors) using the "sfffile" command structure in the 454 GS software package (v2.3). This produced two SFF files with 4001 and 2345 sequences for replicates 1 and 2, respectively. We then used the GS Reference Mapper (454 Life Sciences, Brantford, CT) to assemble these reads to a woolly mammoth mitochondrial genome sequence [20]. An evaluation of the results of the mapping indicated very wide variation in coverage, ranging from 0 to ~350X, probably owing to poor equilibration among amplicons in the pools used for the library generation. Regions corresponding to amplicons A3 and B19 (Additional data file 2: table s5) had no or very low (<3X) read coverage, and so we targeted these for PCR amplification and Sanger sequencing to achieve sufficient coverage for consensus determination.

Primer pair B19 was reamplified from the original group B multiplex reaction (replicate 1) for the specimen, using identical procedures as for the original singleplexes. We reamplified primer pair A3 from the original purified post-multiplex singleplex reaction, for which we used the following PCR protocol:

Each 20µL PCR included: 1X PCR buffer, 2.5 mM $MgCl_2$, 1mg/mL BSA, 250µM each dNTP, 1.5µM each primer, 0.5 units of *Taq*, 0.167X SYBR Green, 5 µL of the purified original singleplex reaction or water for PCR blanks. Cycling conditions were: initial denaturation (94°C, 9m); 24 cycles of denaturation (94°C, 30s), annealing (each appropriate temperature (Additional data file 2: table S5), 30s), and extension (72°C, 40s); concluding with a melt curve from 68-91°C. Amplifications were executed and analyzed using the BioRad CFX96® real-time PCR platform and associated software.

Successful products were then cloned and Sanger sequenced as described in section 5c.

Sanger sequence chromatograms from the fill-in reactions, 454 reads in the MID-sorted SFF files, and all multiplex primers were viewed and reassembled to the mammoth mitogenome reference using Geneious Pro® v5.1.7 (http://www.geneious.com). The assembly parameters included a custom sensitivity level (Allowing gaps of maximum 40%; maximum gap size 100bp; word length 8; index word length 8; maximum mismatches 40%; maximum ambiguity 16) with maximum fine tuning of alignment. Reads were then vertically sorted by position, and each read was then vertically sorted manually such that reads derived from the same amplicon were aligned together. We then manually trimmed the primer sequences from each read. These we then reassembled using the highest sensitivity level and maximum fine tuning, to arrive at final assemblies.

From these trimmed assemblies we built 50% consensus sequences for each replicate and compared them. At some positions, the base calls disagreed between replicates, often in regions with low (<10X) coverage. This prompted a second round of reamplification, cloning, and Sanger sequencing to provide either more clones from each original MPX replicate (A23 and B5) which followed procedures in section 6d, or a third replicate straight from extract (B1 and B23) using the following PCR protocol:

Each 20µL PCR included: 1X PCR buffer, 2.5mM $MgCl_2$, 1mg/mL BSA, 250µM each dNTP, 200nM each primer, 2.5 units of *Taq*, 3µL of template DNA extract (at 0.25x dilution in 0.1X TE) or water for PCR blanks or 0.1X TE used to dilute template. Cycling conditions were: initial denaturation (95°C, 9m); 55 cycles of denaturation (95°C, 30s), annealing (temperature indicated in Additional data file 2 [table S5], 30s), and extension (72°C, 45s).

Upon cloning and sequencing, these data combined to provide enough coverage such that each base position in the final assembly agreed among consensuses from at least two original replicates (whether originally multiplex reactions or straight extract singleplexes), as represented by at least three clones, whether 454 reads or Sanger sequences of amplified product. Upon reassembly of all appropriate clones, both replicates yielded identical consensuses, except for 11 disagreements within the VNTR, which are not included in phylogenetic analysis. The sequence from these experiments yield a consensus 100% identical to the shorter sequence obtained for this

specimen in another study [11]. Given this coverage depth and agreement amongst replicates, we are confident that the mitogenome sequence obtained from this specimen is genuine. Sequence alignments for both replicates and all clones of IK-99-70 can be found in Supplementary Materials: Alignment 1 and 2.

1.9.7 Phylogenetic Analysis

1.9.7a Cytochrome *b*, tRNAs, HVR

The first step of phylogenetic analysis was restricted to a short region of the mitogenome (positions 15006–15748) that includes the latter portion of the cytochrome *b*, tRNA-Thr and tRNA-Pro genes, and the first portion of the D-Loop. At present 126 woolly mammoths have been sequenced for this entire 743bp region, and another 80 woolly mammoths have been sequenced for the 3' 705bp in other studies [9, 10, 12, 13, 20, 37, 45]. Sequences from these mammoths were obtained from GenBank and aligned in Geneious Pro® v5.1.3 (http://www.geneious.com) with the same section of the mitogenome sequence from Huntington, as well as with the entire 743bp from three extant elephants, representing *Loxodonta cyclotis* [37], *Loxodonta africana* and *Elephas maximus* [53].

Among all mammoths, Huntington exhibits no unique polymorphisms in this region, though it does possess a unique combination of SNPs, and thus a unique haplotype. It is assignable to clade I [9, 11] based on two definitive substitutions: a T (clade II & elephants) > C (clade I) transition at position 15044, and an A (clade II & elephants) > T (clade I) transversion at position 15059. Within clade I, they are further assignable to haplogroup C, sharing polymorphisms that are present in the majority of individuals from that haplogroup (Additional data file 2: table S6). The Union Pacific Mammoth yields an identical sequence where it overlaps with the Huntington sequence (positions 15006–15062, 15119–15156, and 15611–15748).

We used the software package BEAST v. 1.5.6 [55] to discern the phylogenetic position of Huntington among other mammoths based on the 743bp region. These analyses were restricted to single representatives of each haplotype, which, with Huntington included, totaled 91 sequences (excluding elephants [Additional data file 2: table S7, sets 1a and 1b]) and 94 sequences (including elephants, [Additional data file 2: table S7, sets 2a and 2b]).

BEAST parameters for sets 1 & 2 (Additional data file 2: table S7) were as follows: Substitution probability matrix conformed to the Hasegawa-Kishino-Yano model, with an 8-category gamma-distributed site heterogeneity scheme and an assumed proportion of invariant sites (HKY+G8+I). This model was chosen using jModelTest v. 0.1.1 [54], in which the Bayesian Information Criteria analysis indicated its appropriate application for both datasets. Two relaxed clock models were used in separate analyses, including the uncorrelated lognormal and uncorrelated exponential models. Trees were generated using the piecewise-constant Bayesian Skyline demographic model with 20 groups. Each parameter set ran for 10m generations (sampling every 1000 generations) using the default prior distributions and operator settings, except with the "skyline.Popsize" prior set to 0 to 100. Operator analyses following these runs suggested tuning modifications, which we carried out for the subsequent runs. However, previous experience suggested that expanding the integer.RandomWalk operator to a window size of 100 (rather than the suggested 2.0) is appropriate for this dataset. Following operator adjustment, each parameter set was used in 3 independent 10m generation runs (sampling every 1000 generations). Tree files were combined using LogCombiner with a 10% sample burn-in applied to each independent run. These were then combined in TreeAnnotator using the default annotation parameters, keeping target node heights. Combined log files as viewed in Tracer v1.5 (http://tree.bio.ed.ac.uk/software/tracer/) yield ESS>100 for each analysis parameter, suggesting full posterior convergence for each set. Trees were viewed in FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/) for evaluation of topology and determination of nodal posterior probabilities.

Annotated trees from all four parameter sets are presented in Additional data file 2 (figures S3 through S6), with nodal posterior probabilities indicated. As demonstrated, nodal posterior probabilities for the clade that includes haplogroups C, D, and E (clade I) are consistently 1.00 in all analyses, indicating strong support for the inclusion of Huntington within that clade. In both variants of set 1, there is also modest support for the monophyletic grouping of all C mammoths, with posterior probabilities of 0.63 (set 1a) and 0.80 (set 1b) at the MRCA node for those haplotypes. Interestingly, in sets 2a and 2b, we observe the polyphyletic relationship between two subdivisions

within haplogroup C observed in Debruyne *et al.* [11], though with only limited nodal support for monophyly of D+E and the C subgroup. While the topology within haplogroup C thus remains somewhat unresolved, we are confident that the Huntington sequence at least falls securely within clade I, and most probably within haplogroup C. Therefore we assign it an appropriate new haplotype (C32).

We also performed a set of analyses exploring temporal aspects of the tree for this mitogenomic region, using root + tip calibration and two different clock models (the uncorrelated lognormal and uncorrelated exponential models) and the Bayesian Skyline demographic model (30 groups). This was restricted to analysis of only those mammoths with finite radiocarbon dates (n = 156) and three modern elephants. For the root calibration (tMRCA for elephants + mammoths), we used 7.7my (sd=5e5), corresponding to the date calculated by Rohland et al. [56], which is also consistent with fossil evidence for the divergence of the genera. Following an initial run in order to determine appropriate operator tuning, we ran these analyses for two (exponential model) and three (lognormal model) independent 50m generation runs. The lognormal model failed to converge, yielding ESS<100 for some parameters. The exponential model did, however, converge to ESS>100 for all parameters after 100m generations. Temporal estimates on key nodes in mammoth phylogeny are reported in Additional data file 2 (table S7).

1.9.7b Whole Mitogenomes

Similar analyses as above were executed using full mitogenomic sequences excluding the VNTR, which are publically available for 20 mammoths including the ones reported here. We used sequences from only one elephant of each genus as outgroups for these analyses. For both the full data set (employed in set 4) and only those with finite radiocarbon/modern dates (set 5), analyses employed the Tamura-Nei 93 substitution model with 8 gamma categories, a Bayesian Skyline demographic model (10 groups, constant addition), and tuning, burnin, combination and annotation protocols similar to section 7a.

Consensus trees and associated nodal posterior probabilities generated in the topological analyses (set 4) are depicted in Additional data file 3 (figures S7 and S8). Temporal estimates for key nodes in mammoth phylogeny are included in Additional data file 2 (table S7). As observed elsewhere [14] temporal estimates derived from analyses of full mitogenomes are significantly more ancient than the same estimates derived from analyses of shorter mitogenomic regions.

1.9.8 Nuclear Genome Read Analysis

1.9.8a Read Classification

In order to estimate the nuclear genome divergence, we mapped all Huntington mammoth reads to *Loxodonta africana* to first determine the subset of nuclear genome reads. We used NUCmer to align reads to the reference *L. africana* assembly (http://www.broadinstitute.org/ftp/pub/assemblies/mammals/elephant/loxAfr3/) with the following parameters: MINCLUSTER = 16, MINMATCH = 16, --MAXMATCH. This yielded 2,505,001 reads classified as nuclear, or 8.7% of the total reads. Then, to estimate divergence, we downloaded 3.6 million woolly mammoth M4 reads [52] from the SRA SRX001906 genomic fragment library. Before aligning the reads we aggressively trimmed 5 nucleotides from each end in attempt to remove divergence bias caused by sequencing errors and DNA damage. We then aligned the reads using NUCmer (same parameters) and filtered the results with show-coords to exclude any reads with <80 % ID and <35 nt in length. We also only reported unique, 1-to-1 alignments. This returned a final set of 15,650 aligned nuclear reads.

To accompany this result, we also performed individual classification of a subsample of the reads using the metagenomics analysis program PhymmBL [51] and a local sequence database including *L. africana*. Instead of running Phymm on the 28 million reads we randomly selected 1000 reads with replacement from the total sample of 28+ million reads 100 times. PhymmBL assigns a species identifier to each read, and we used this to count the percentage of reads that were classified *as L. africana*, which was 6%.

Given these two results we estimate the nuclear reads in this sample to be between 6-8% of the total.

1.9.8b Nuclear Genome Divergence Estimate

In an effort to compare the Huntington and woolly mammoth nuclear genomes, we analyzed the degree of divergence between the Huntington nuclear read data and the woolly mammoth ("M4") nuclear genome [52]. In order to limit the influence of sequencing error and mitigate effects of low coverage of the M4 mammoth nuclear genome (<<1X), we aligned only those Huntington reads to regions of the M4 nuclear genome that are covered by at least 2 reads from that dataset, which reduced the alignable Huntington reads to less than 1000. Then, after carefully inspecting the alignments returned by NUCmer using show-aligns, we determined the average percent identity of these reads to be 97% to the woolly mammoth nuclear reference. We also compared the M4 mitochondrial genome to the Huntington genome, and found it more similar (99.5% identical) to Huntington than the Krause *et al*. [20] mitochondrial genome. The differences between the nuclear and mitochondrial estimates, however, are very likely to be driven largely by the significantly lower coverage of the nuclear data (for both M4 and Huntington) and therefore more susceptible to miscalls from sequencing error and DNA damage. Thus, the real nuclear divergence between the species can only be confirmed by additional, much deeper sequencing.

1.10 ADDITIONAL DATAFILE 2 – SUPPLEMENTARY TABLES

Table S1: Primers and target regions used in this study. *Ref*: * = Debruyne *et al.* [11], ◊ = This study.

					Amplicon size (bp)									
#	*F/R*	*5' position*	*5' – 3' sequence*	*Ref.*	*63*	*78*	*79*	*85*	*86*	*103*	*110*	*118*	*121*	*124*
1	F	14985	GCCATCCTACGATCTGTACCA	*						✓				
2	R	15087	GGTGTTTAGATGTATGTAGAAGTGG	◊						✓				
3	F	15097	TACTTCGACCTCTTAGCCAAGT	*			✓							
4	R	15175	GGTTGACTGCCAATTCATG	*			✓							
5	F	15337	CTATTTTCTAAGGGTATTCAGGGAAGAG	◊				✓					✓	
6	F	15395	CTGAAATTCTTCTTAAACTATTCCCTGC	◊	✓									
7	R	15421	CAGGGAATAGTTTAAGAAGAATTTCAGT	◊				✓						
8	R	15457	TTAATGCACGATGTACATAGCGG	◊	✓								✓	
9	F	15587	CAAGTCATATTCGTGTAGATTCAC	◊							✓			
10	F	15647	GATAAACCATAGTCTTACATAGCAC	◊										✓
11	F	15668	GCACATTAAAGCTCTTGATCGTACAT	◊		✓								
12	R	15696	GCTATGTACGATCAAGAGCTT	◊							✓			
13	R	15745	CAACCGTTGGAGGTGATATGC	◊		✓								
14	R	15770	TTCTCGGAGGTAGGTAGTTAAG	*										✓
15	F	134	GCCAGTGAATACGCCTTCTAA	◊	✓									
16	R	196	GGTGTGTGTGCTTGATGC	◊	✓									
17	R	219	GCGAGACGTCATGAGCTACA	◊					✓					
18	R	251	ACTACTGCTGTTTCCCGTGG	◊								✓		

Table S2. Copies/mg of substrate and resultant regression statistics, following the model proposed by Deagle *et al.* [46]. *M. columbi:* includes samples of various dilutions. *(λ)* (lambda): the linear slope of the regression line calculated from the relationship between log-transformed molecular count and amplicon size. *1/λ* (inverse Lambda): estimated average amplifiable fragment length in the sample. *y=0* (x-intercept): estimated maximum amplifiable fragment size in the sample. *Since *HUNT1* comes from 5 of 31 (~16%) total mL of solution generated from the 0.98g tusk sample, we project these metrics assuming it derives from 158mg (0.16*980mg). Red = metrics derived from only 2, rather than 3 or more, data points.

	Amplicon Length (bp)								
M. primigenius	*84*	*151*	*279*	*490*	*677*	*921*	*Λ*	*1/λ*	*y=0*
Ber12	1,932.5	807.7	42.9	1.3	0.0	0.0	-0.0080	125.4	500.3
173	1,536.5	773.9	69.9	5.1	1.1	0.3	-0.0047	214.7	722.8
917	19,200.5	13,125.9	1,508.3	138.7	26.9	2.7	-0.0047	212.3	985.0
473	864.3	420.5	54.1	5.1	0.8	0.0	-0.0052	192.7	640.8
472	7,261.3	3,249.1	253.9	16.8	4.3	0.5	-0.0050	200.3	813.0
915	6,095.5	3,870.4	534.1	57.9	15.7	1.9	-0.0043	235.1	956.1
M. columbi	*63*	*85*	*86*	*118*	*121*	-			
HUNT1 0.1X*	487.6	-	92.8	12.6	-	-	-0.0287	34.8	155.7
HUNT2 1X	20.8	10.3	-	-	0.5	-	-0.0289	34.6	113.0
HUNT2 0.1X	53.6	17.8	-	-	1.0	-	-0.0299	33.4	123.1
HUNT2 0.02X	99.7	3.0	-	-	0.0	-	-0.0692	14.5	91.9
UPT2 1X	2.5	0.9	-	-	0.0	-	-0.0214	46.8	81.9
UPT2 0.1X	0.9	1.0	-	-	0.0	-	-	-	-

Table S3. Assembly and consensus comparisons. *Con.*: Assembly/Consensus ID. *Program*: Software program used in assembly and consensus generation. *Reference*: reference sequence for assembly. *Reads aligned*: number of reads aligning to the reference in the assembly. *Consensus Length*: length of the consensus sequence after gaps removed, including any within the VNTR. Red = disagreements from other consensuses at indicated base positions, not including gaps and anomalous sections in the *Elephas* assemblies as discussed in section 4h.

Con.	*Program*	*Reference*	*Sens.*	*Reads Aligned*	*Consensus Length*	2882	3541	4774	4776	5665	7298	9710	10650	10792	10952	11514	11628	13769	15043	15257
1	454 RefMap	Mammoth	-	7784	16738	C	T	T	T	C	C	T	C	T	C	C	T	C	C	G
2	AMOScmp	Mammoth	-	8048	16770	C	T	T	T	C	C	T	C	C	C	C	T	C	C	G
3	Geneious	Mammoth	Low	6664	16689	C	T	T	T	C	C	T	C	C	C	C	T	C	C	G
4	Geneious	Mammoth	Med	7484	16689	C	T	T	T	C	Y*	T	C	C	C	C	T	C	C	G
5	Geneious	*Elephas*	Low	5137	16355	Y	G	N	C	Y	C	-	T	C	T	S	N	Y	S	A
6	Geneious	*Elephas*	Med	6682	16814	C	T	T	T	C	C	T	C	C	C	C	T	C	C	G

* Six reads cover this position; of the five unique, three call “C” while two call “T.”

Table S4. PCR summary. Primer pairs are indicated, referencing table S1.

Amplicon	*Sample*	*PCRs.*	*Ext. vol (μl)*	*PCR product*	*Reamp?*	*Sequenced?*	*Clones Seq.*	*Init.*	*Cycles*	*T_a (°C)*
78bp (#11+13)	*HUNT2*	1	3	+	N	Y	3	4 min60		60
		2	3	+	N	Y	3			
79bp (#3+4)	*HUNB1*	1	3	+	N	Y	4	7 min50		62.5
		2	3	+	N	N				
		3	3	+	N	Y	5	7 min51		62
	UPT1	1	5	+	N	Y	2	7 min50		62.5
		2	5	+	N	Y	3			
103bp (#1+2)	*HUNB1*	1	3	(+)?	N	N		7 min50		62.5
		2	3	-						
		3	3	(+)	Y	Y	3	7 min51		62
		4	3	(+)	Y	Y	1			
	UPT1	1	5	(+)?	N	N		7 min50		62.5
		2	5	(+)?	N	N				
		3	3 (1in10)	(+)	Y	N		7 min51		62
		4	3 (1in10)	-						
		5	5	+	N	Y	5	4 min60		62.5
		6	5	+	N	Y	6			
110bp (#9+12)	*HUNB1*	1	5	+	N	Y	3	7 min45		59.5
		2	5	+	N	Y	3			
	UPT1	1	5	+	N	Y	2	7 min45		59.5
		2	5	+	N	Y	2			
124bp (#10+14)	*HUNB1*	1	5	(+)?	N	N		7 min45		59.5
		2	5	-						
		3	5	-				7 min51		59.5
		4	5	+	N	Y	3			
		5	5	(+)	Y	Y	6	4 min60		59.5
		6	5	-						
		7	5	-				4 min60		59.5
		8	5	-						
	UPT1	1	5	+	N	Y	1	7 min45		59.5
		2	5	+	N	Y	2			
		3	3	+	N	Y	3	7 min51		59.5
		4	3	+	N	N				

Table S5. Primers used in the multiplex PCR experiments (section 6). *Grp*: multiplex group name. *For/Rev 5'*: The 5' position each primer. *Rf*: * = Krause *et al.* [20], ◊ = This study. *Bp*: amplicon length in nucleotides. *Ta*: annealing temperature used.

Grp	*Pair*	*For 5'*	*Forward*	*Rf*	*Rev 5'*	*Reverse*	*Rf*	*Bp*	*Ta*
A-2	A1	84	CCGGCCTTCTTATTGGTTAC	*	570	GTAGTTCTCTGGCGGATAGC	*	487	63.5
A-2	A2	1023	TGGGTAACTCAAAGTGTAGCTT	*	1458	TCTGGACAACCAGCTATCATC	*	436	63.2
A-2	A3	1876	AGGGAAAGATTAAAAGAAGGA	*	2361	GGTAACTTGTTCCGTTGATCA	*	486	62.0
A-2	A4	2585	CTTACCAAGACGCCTTCAGC	*	3021	GGATATGGTATTGGAAGAGG	*	437	63.2
A-1	A5	3371	TTCAACGTCGAATATTCAGC	*	3755	CAGGGTTTAGACCTCTATAATTT	*	385	61.0
A-1	A6	4092	CAAGCCACAGCATCCATAAT	*	4461	GGGAGATTGAAGAGTAGGC	*	370	61.0
A-1	A7	4837	TGAAAATAACCTGACAATTTA	*	5224	AAGCAGCTTCAATTCTGCC	*	388	56.2
A-2	A8	5547	ATTATAATTGGAGGCTTTGG	*	6021	ATAGAATTGGGTCTCCTCCT	*	475	62.0
-	A8a	5504	CACAGCACACGCCTTTGTAA	◊	6017	AATTGGGTCTCCTCCTCCTG	◊	514	63.5
A-1	A9	6397	TTGTTCTTGCCAATTCTTCACT	*	6855	CTGGTTCTTCGAATGTATGA	*	459	61.0
A-2	A10	7115	TCCTTATTAGCTCCTTAGTCTTG	*	7545	AATTGCATCTGTTTTTAGACC	*	431	63.5
A-2	A11	7724	GCACTAACCTTTTAAGTTAGAGTAT	*	8075	TTTGACTAGTCATTGTTGGA	*	352	57.5
A-2	A12	8290	TCTCACTAGCCCATCTTCTC	*	8670	GATAATGCTCCGGTAAGAGGTC	*	381	63.2
A-1	A13	9052	CTTAATAGAAGGAAATCGTAAA	*	9436	CAAAGCCTACTAATTGGAAGTT	*	385	56.2
A-1	A14	9778	CAAAAAGGCCTTGAATGAAC	*	10202	GGGAATAAGTATGATTGTTGGTA	*	425	61.0
A-1	A15	10461	ATCCTCTTACAAGTATCCCTAAT	*	10834	GGTGCTTCTACATGAGCTTT	*	374	63.2
A-2	A16	11169	CGCATTCATAGCCGAACTAT	*	11736	TTTTATCTGGAGTTGCACCA	*	568	63.5
A-1	A17	12522	CCGCTTTTATCCATTAATAGAAA	*	12815	ATTTTGCGGATGTCTTGTTCG	*	294	63.2
A-1	A18	11996	CCCTAACATTCATGCCAATTG	*	12421	CAAATTGGGCTGATTTTCCT	*	426	63.2
A-2	A19	13150	AATACTATTCCGCATACAACAC	*	13627	GGTTTTGATTTTTGGCTATG	*	478	61.0
A-2	A20	14032	ACCACATAAAGCACACTCAT	*	14577	CCCCTCAGAATGATATTTGT	*	546	61.0
A-1	A21	14802	ATCCCCTTTCACCCGTACTA	*	15230	GGAGAAATATAGAATTGAGGCTA	*	429	63.2
A-1	A22	15395	CTGAAATTCTTCTTAAACTATTCC	*	15736	GAGGTGATATGCATGATGA	*	342	61.0
A-2	A23	16086	GCACGGTATATATGGGGTAT	*	16534	TGAGCCAAGGGTAACTAAGG	*	449	57.5
B	B1	522	CCCTAAACTTTGATAGCTACC	◊	1071	CTAGGTGTAAGCCAGATGCT	*	550	63.0
B	B2	1405	CTTACAGATAGAGGTGAAATACCA	*	1919	GTTTATGTTTGCCGAGTT	*	515	60.3
B	B3	2269	AAAACCTCCGAACGATATTA	*	2664	AATCCTGTTCTTGGATTGG	*	396	60.5
B	B4	2973	CCCTAACCTTAGCCCTAACT	*	3437	TGATAATGTTAGCGTATTCG	◊	465	60.0
B	B5	3683	CCACCACAAGCATAGAAATA	*	4157	TCATTGTCCTGAGTATATTAGATT	*	475	60.5
B	B6	4412	CCTAAATCAAACACAACTACGA	*	4898	GGTGATTAGAGTCGGTAATAT	◊	487	58.7
B	B7	5149	TGGCTTCAATCTACTTCTCC	*	5593	CCGATTATAAGTGGAATTAATCA	*	445	60.0
B	B8	5963	GGACCGCAACCTCAATACTAC	*	6441	AGGTATCATGTAGGACAATG	◊	479	63.0
B	B9	6762	TCTAAGCGCGAAGTTTCT	*	7228	CTGGGAGAATGGTTCAGAT	*	467	60.5
B	B10	7502	ATTCATGGGCTGTCCCAT	*	7941	CGTTCACTTCTTCTTTCAAGG	*	440	63.0
B	B11	8025	AATCGCCTAATTACCAACC	*	8330	GAATGTAGGTGTTCCTTGTGG	◊	306	63.0
B	B12	8612	CAAACACATGCCTATCACAT	*	9102	ATAGTAATAAGGAGGGCTTG	◊	491	60.0
B	B13	9297	ACTTTGGCTTTGAAGCAG	*	9837	TCGAAATCATTTGTTTTGTT	*	541	58.7
B	B14	10124	CCTACGGACTAGACTACGTACAA	*	10541	AAGTGTTGTTTCAAATATAATG	◊	418	58.7
B	B15	10767	GCTTTCATAGTAAAAATACCTCTA	◊	11230	GCTATTAGTGGGAGAAGGGTTT	*	464	63.0
B	B16	11696	TGGTCTTAGGCACCAAAA	*	12051	TGAAAATTCTATGATTGATCAGG	◊	356	58.7
B	B17	12323	CCTGAGAATTTCAACAAATCTT	◊	12648	TGCGATAATCTTTTTGATGT	*	326	58.7
B	B18	12769	GGCTCTATCATCCACAACCT	*	13192	GTGGAGGTATGGTTATTTGG	*	424	63.0
B	B19	13565	CCTCAATAGCAATAAAAATACTAA	*	14120	CGATGGTTTTTCAGATCATT	*	556	60.3
B	B20	14527	CCGCCTTCATAGGATATGT	*	14926	GCTGGTATGTAGTTGTCGGG	*	400	63.0
B	B21	15062	ACCACTTCTACATACATCTA	*	15543	AATGTGATGCACGATTATACA	*	482	58.7
B	B22	15677	AGCTCTTGATCGTACATAGC	*	16128	GTCCTCCGAGCATTGACT	*	452	63.0
B	B23	16478	CCACTATGTAACTATCTCTTCAAA	*	196	GGAGTGTGTGCTTGATGC	*	489	63.0

Table S6. Clade- and haplogroup-defining polymorphisms assayed for the 743bp region of the mammoth mitogenome. Base positions (sites) refer to the reference sequence [20]. Coded nucleotide bases: Uppercase = nucleotide base found in >95% of all haplotypes of the indicated group; lowercase = base found in most haplotypes (>70%) of that group; blue = experimental data from this study, with Huntington-only data indicated in standard face, and data observed in both Huntington and Union Pacific in underlined bold face. *Haplogroup B is represented by two individuals, hence the ambiguous consensus at several positions.

			Clade-specific sites		*Haplogroup-specific sites*							
Clade	*Taxon*	*Hap*	*15044*	*15059*	*15144*	*15183*	*15425*	*15568*	*15612*	*15621*	*15622*	*15625*
II	*M. primigenius*	*A*	T	A	t	C	G	C	G	C	T	T
	M. primigenius	*B**	T	A	T	T	A/G	T/C	G	T/C	T/C	T
I	*M. primigenius*	*C*	C	T	T	T	G	T	G	T	t	T
	M. columbi		**C**	**T**	**T**	T	G	T	**G**	**T**	**T**	**T**
	M. primigenius	*D*	C	T	C	T	G	t	A	T	C	C
	M. primigenius	*E*	C	T	C	T	A	t	A	T	C	C

Table S7. BEAST analyses. *Samples:* "Elephants" = two (sets 1-3) or one (sets 4-5) *Loxodonta* sequence(s) and one *Elephas* sequence. "Mammoths" = all woolly mammoths and Huntington. "1ea hap" = one representative of each haplotype. "Dated" = remains with associated finite radiocarbon dates. *Clock:* Clock model used for each analysis, "Exp" = the uncorrelated exponential clock model. "Log" = the uncorrelated lognormal clock model; *Root Cal:* indicates whether root age prior was used, "N" = prior omitted, "Y"=normal distribution prior with mean 7.7my and standard deviation 500ky, derived from the findings of Rohland et al. [56]. *Tip Cal:* Indicates whether or not tip dates were used. For those mammoths with associated radiocarbon dates, the raw date was used (not calendar-calibrated). For elephants, an age of 0rcya was used. Red = combined runs did not reach ESS>100 for each parameter.

											tMRCA mammoths			*tMRCA clade I*			*tMRCA haplogroup C*		
Set	*# Seqs*	*Bp*	*Samples*	*Clock*	*Root Cal*	*Tip Cal*	*Runs*	*Gens*	*Mean Posterior Probability*	*Mean Tree Likelihood*	*95%HPD lower*	*Median (years)*	*95%HPD upper*	*95%HPD lower*	*Median (years)*	*95%HPD upper*	*95%HPD lower*	*Median (years)*	*95%HPD upper*
1a	91	743	Mammoths, 1ea hap	Exp	N	N	3	10m	-1814.01	-2073.50	-	-	-	-	-	-	-	-	-
1b	91	743	Mammoths, 1ea hap	Log	N	N	3	10m	-1820.96	-2092.03	-	-	-	-	-	-	-	-	-
2a	94	743	Elephants, mammoths, 1ea hap	Exp	N	N	3	10m	-2207.29	-2434.60	-	-	-	-	-	-	-	-	-
2b	94	743	Elephants, mammoths, 1ea hap	Log	N	N	3	10m	-2208.34	-2454.13	-	-	-	-	-	-	-	-	-
3a	159	743	Elephants, mammoths, dated	Exp	Y	Y	2	50m	-4848.56	-2254.51	1.64e5	**4.53e5**	1.12e6	9.06e4	**2.04e5**	3.97e5	7.10e4	**1.57e5**	3.26e5
3b	159	743	Elephants, mammoths, dated	Log	Y	Y	3	50m	-4926.59	-2257.58	2.07e5	**6.65e5**	1.54e6	1.10e5	**2.81e5**	6.45e5	7.86e4	**2.16e5**	5.12e5
4a	22	16.5k	Elephants, mammoths	Exp	N	N	3	20m	-30323.57	30417.87	-	-	-	-	-	-	-	-	-
4b	22	16.5k	Elephants, mammoths	Log	N	N	3	20m	-30322.35	30419.93	-	-	-	-	-	-	-	-	-
5a	18	16.5k	Elephants, mammoths, dated	Exp	Y	Y	3	20m	-30491.91	30051.40	2.67e5	**1.58e6**	4.92e6	1.35e5	**6.25e5**	1.65e6	6.15e4	**2.54e5**	7.77e5
5b	18	16.5k	Elephants, mammoths, dated	Log	Y	Y	1	20m	-30505.31	30054.62	1.02e6	**1.65e6**	2.46e6	3.85e5	**6.12e5**	8.87e5	2.18e5	**4.09e5**	6.25e5

1.11 ADDITIONAL DATAFILE 3 – SUPPLEMENTARY FIGURES

Fig. S1. The Huntington mammoth lower left third molar. Reproduced with permission from [17].

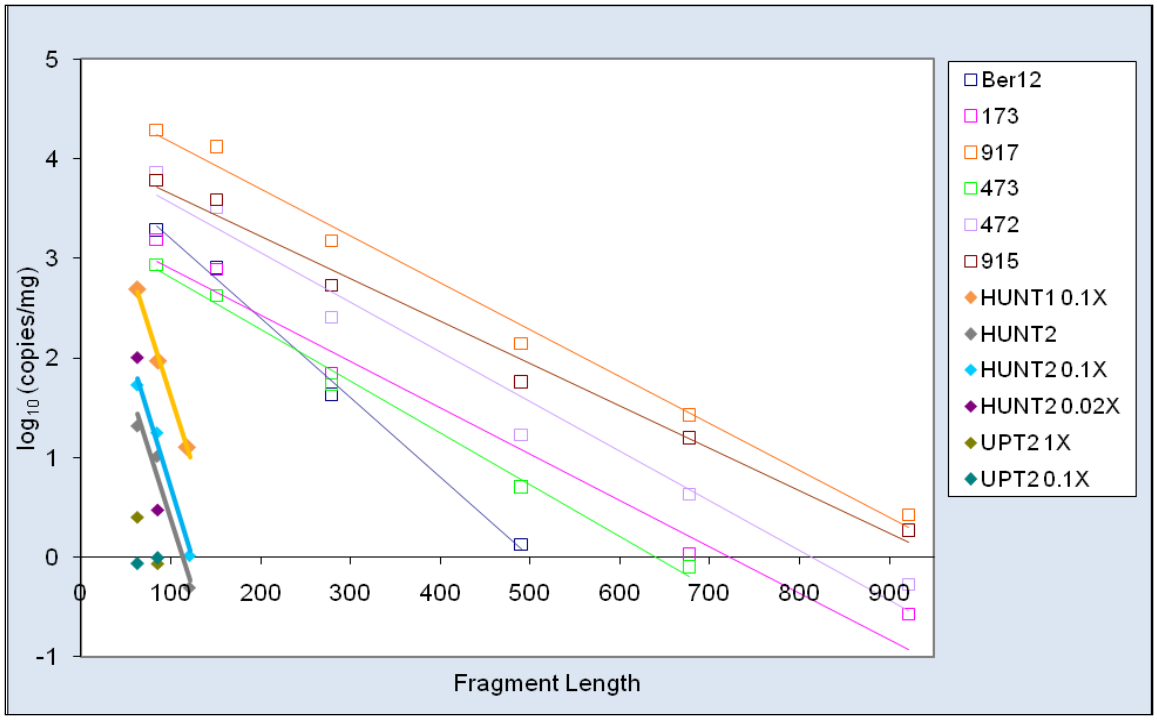


Fig. S2. Log-transformed amplicon copy number per milligram of substrate vs. fragment length. Open squares are woolly mammoth samples assayed by Schwarz et al. [24], with regression lines indicated.

Fig. S3. Maximum clade credibility tree from phylogenetic analysis set 1a, with nodal posterior probabilities indicated. Tip names are preceded by their haplotype as determined by [11] and [12].

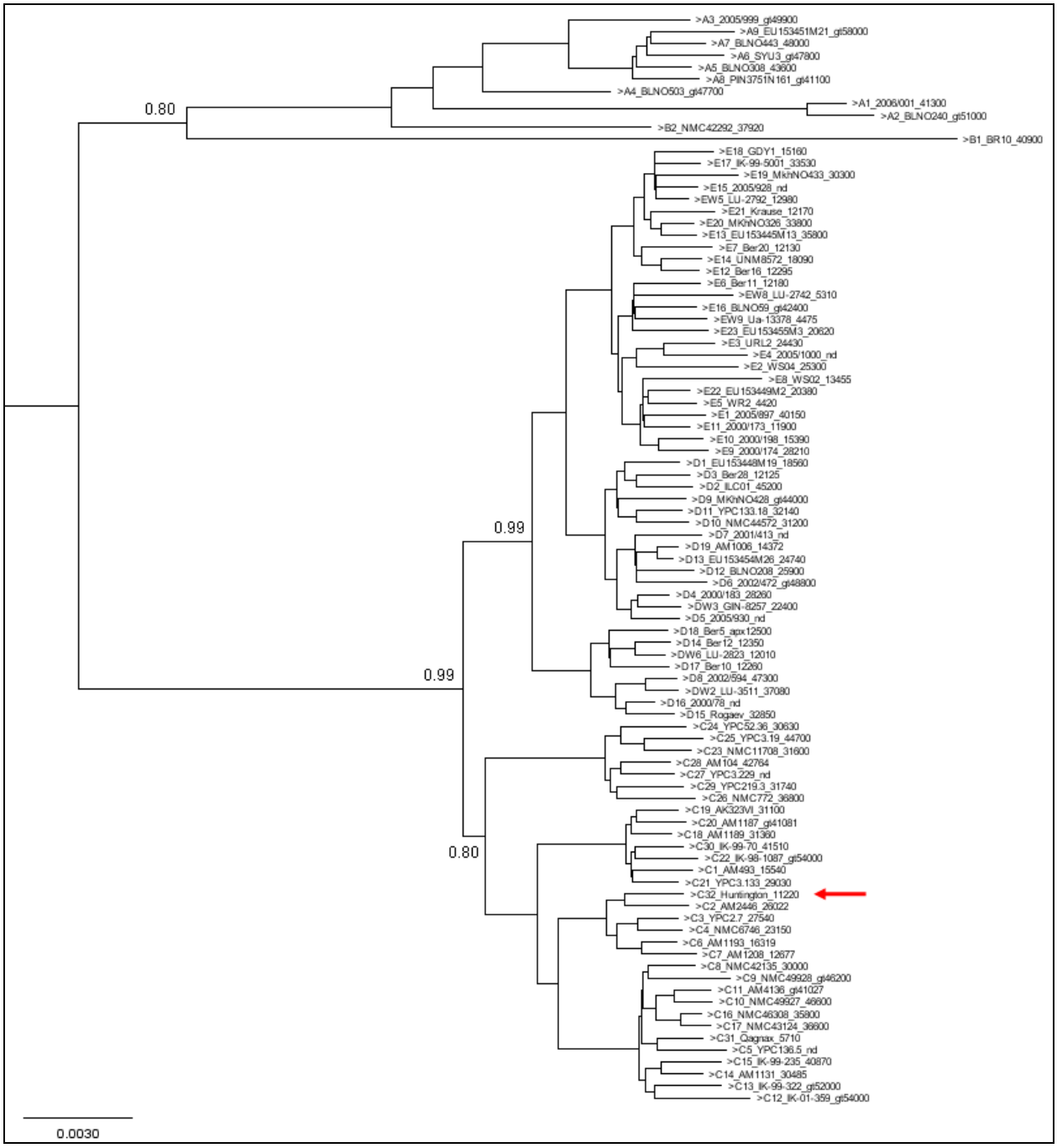


Fig. S4. Maximum clade credibility tree from phylogenetic analysis set 1b, with nodal posterior probabilities indicated. Tip names are preceded by their haplotype as determined by [11] and [12].

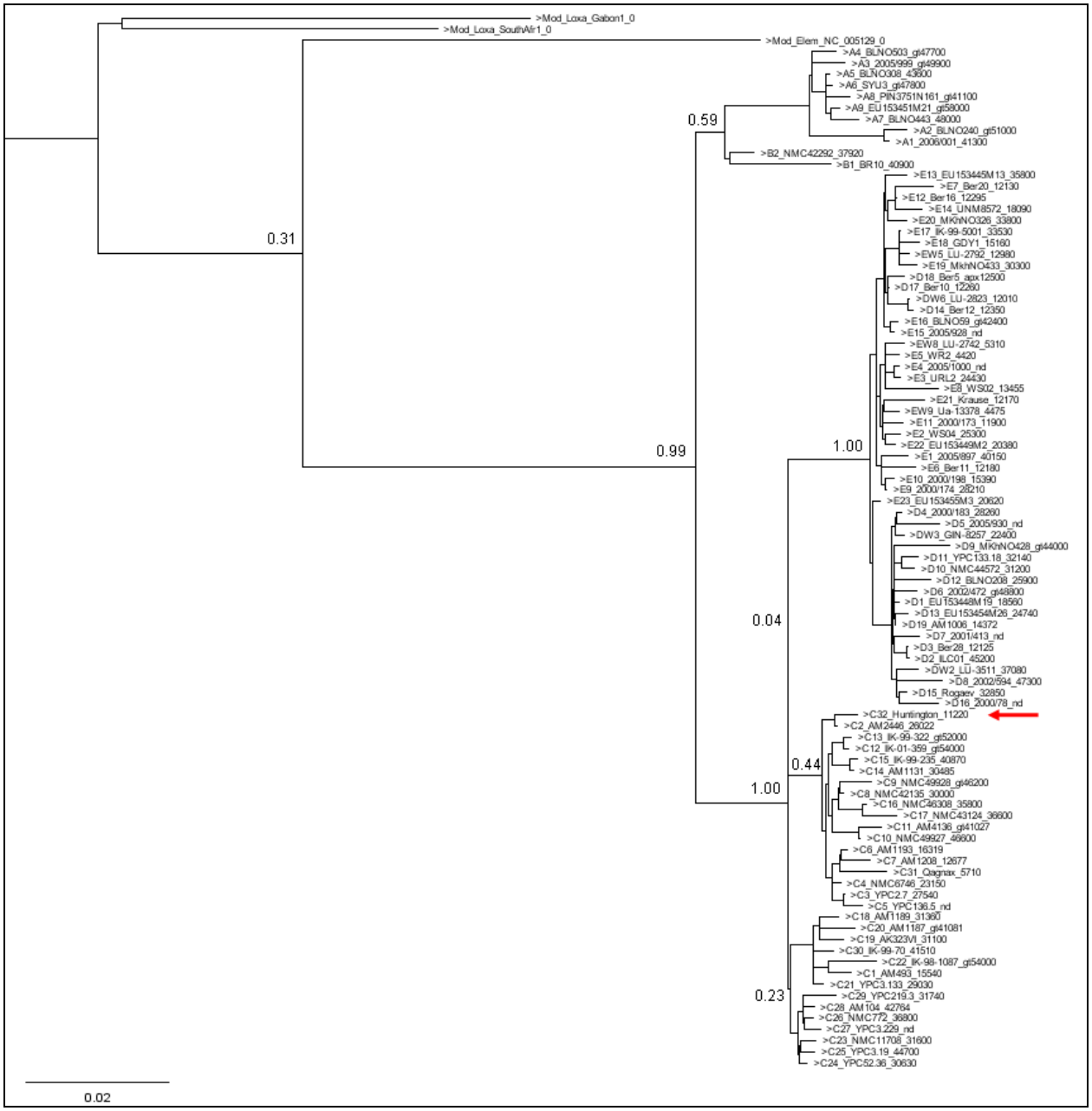


Fig. S5. Maximum clade credibility tree from phylogenetic analysis set 2a, with nodal posterior probabilities indicated. Tip names are preceded by their haplotype as determined by [11] and [12].

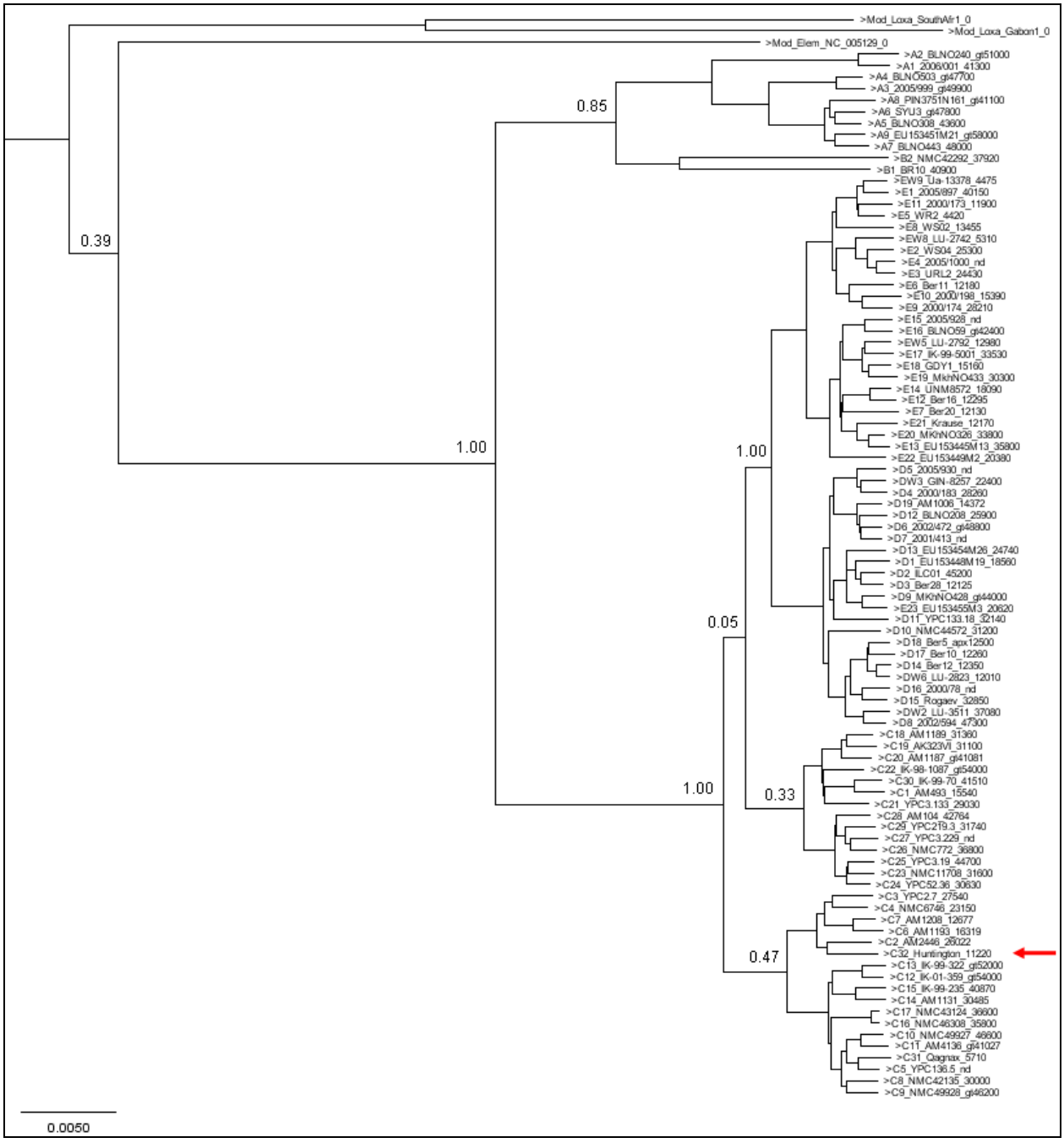


Fig. S6. Maximum clade credibility tree from phylogenetic analysis set 2b, with nodal posterior probabilities indicated. Tip names are preceded by their haplotype as determined by [11] and [12].

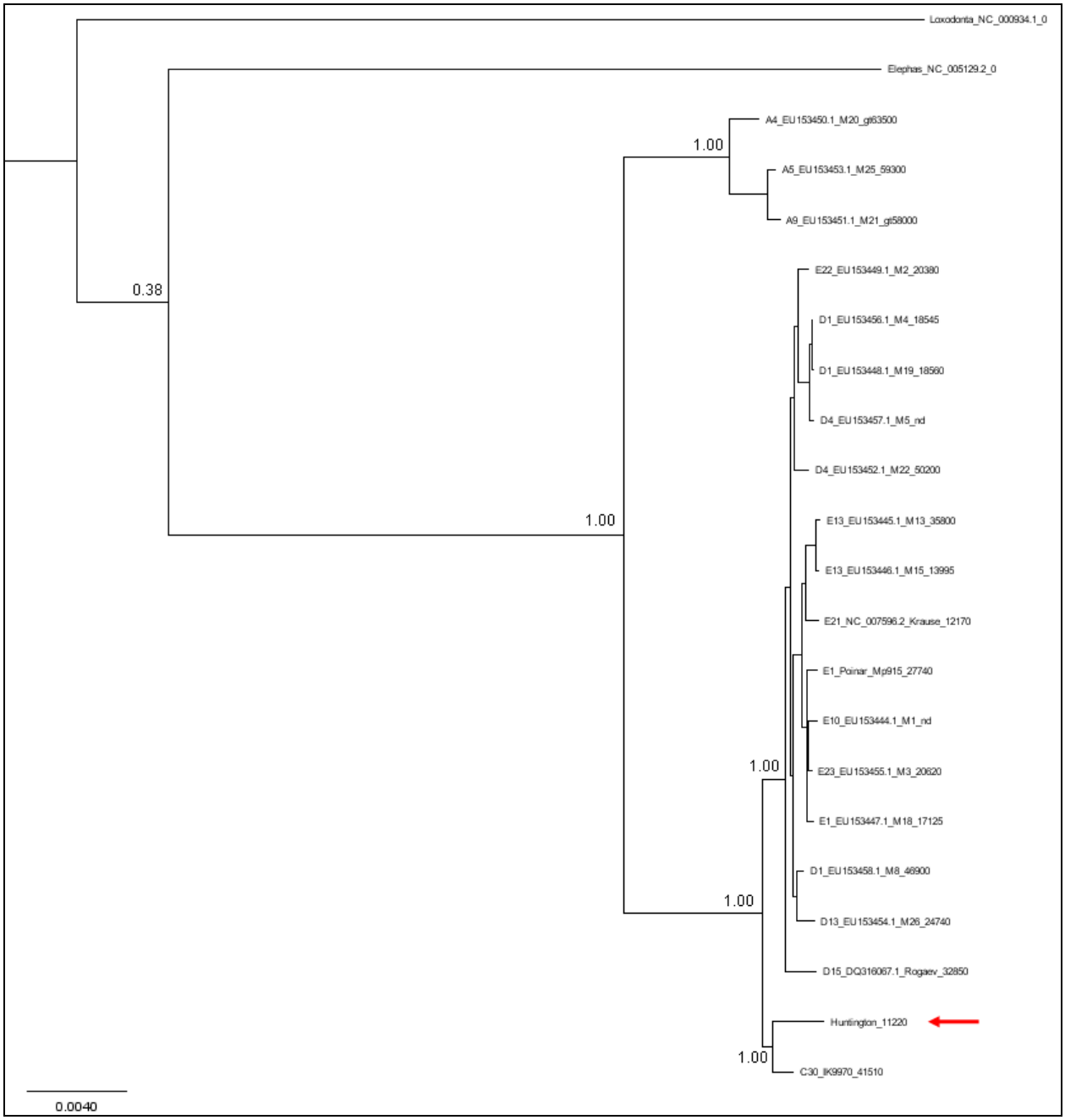


Fig. S7. Maximum clade credibility tree from phylogenetic analysis set 4a, with nodal posterior probabilities indicated. Tip names are preceded by their haplotype as determined by [11] and [12].

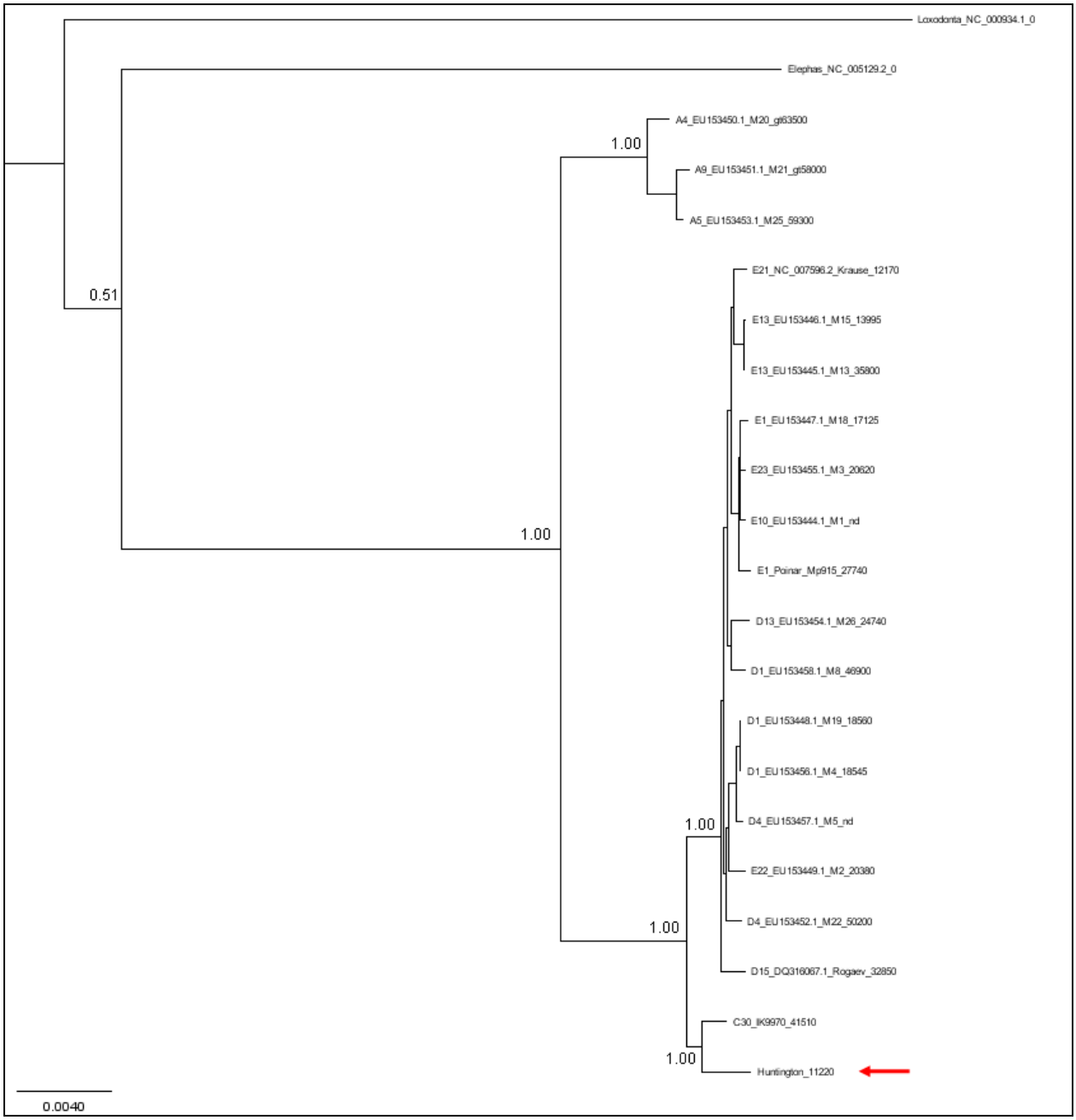


Fig. S8. Maximum clade credibility tree from phylogenetic analysis set 4b, with nodal posterior probabilities indicated. Tip names are preceded by their haplotype as determined by [11] and [12].

1.12 REFERENCES

1. Maglio VJ: **Origin and evolution of the Elephantidae.** *Transactions of the American Philosophical Society* 1973**:**1-149.
2. Agenbroad LD: **North American Proboscideans: Mammoths: The state of Knowledge, 2003.** *Quaternary International* 2005, **126:**73-92.
3. Harington C: **Mammoths, bison and time in North America.** *Developments in Palaeontology and Stratigraphy* 1984, **7:**299-309.
4. Lister A, Bahn PG: *Mammoths : giants of the ice age.* Rev. edn. Berkeley, Calif.: University of California Press; 2007.
5. Kurtén Br, Anderson E: *Pleistocene mammals of North America.* New York: Columbia University Press; 1980.
6. Lister AM, Sher AV: **The origin and evolution of the woolly mammoth.** *Science* 2001, **294:**1094-1097.
7. Lister AM, Sher AV, van Essen H, Wei GB: **The pattern and process of mammoth evolution in Eurasia.** *Quaternary International* 2005, **126:**49-64.
8. Shoshani J, Tassy P: **Advances in proboscidean taxonomy & classification, anatomy & physiology, and ecology & behavior.** *Quaternary International* 2005, **126:**5-20.
9. Barnes I, Shapiro B, Lister A, Kuznetsova T, Sher A, Guthrie D, Thomas MG: **Genetic structure and extinction of the woolly mammoth, Mammuthus primigenius.** *Curr Biol* 2007, **17:**1072-1075.
10. Gilbert MTP, Drautz DI, Lesk AM, Ho SYW, Qi J, Ratan A, Hsu CH, Sher A, Dalen L, Gotherstrom A, et al: **Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes.** *Proc Natl Acad Sci U S A* 2008, **105:**8327-8332.
11. Debruyne R, Chu G, King CE, Bos K, Kuch M, Schwarz C, Szpak P, Grocke DR, Matheus P, Zazula G, et al: **Out of America: Ancient DNA evidence for a new world origin of late quaternary woolly mammoths.** *Current Biology* 2008, **18:**1320-1326.
12. Nystrom V, Dalen L, Vartanyan S, Liden K, Ryman N, Angerbjorn A: **Temporal genetic change in the last remaining population of woolly mammoth.** *Proceedings of the Royal Society B-Biological Sciences* 2010, **277:**2331-2337.
13. Enk JM, Yesner DR, Crossen KJ, Veltre DW, O'Rourke DH: **Phylogeographic analysis of the mid-Holocene Mammoth from Qagnax Cave, St. Paul Island, Alaska.** *Palaeogeography Palaeoclimatology Palaeoecology* 2009, **273:**184-190.
14. Debruyne R, Poinar HN: **Time Dependency of Molecular Rates in Ancient DNA Data Sets, A Sampling Artifact?** *Syst Biol* 2009, **58:**348-359.
15. Ho SY, Phillips MJ, Cooper A, Drummond AJ: **Time dependency of molecular rate estimates and systematic overestimation of recent divergence times.** *Molecular Biology and Evolution* 2005, **22:**1561-1568.
16. MacFadden BJ, Hulbert RC: **Calibration of mammoth (Mammuthus) dispersal into North America using rare earth elements of Plio-Pleistocene mammals from Florida.** *Quaternary Research* 2009, **71:**41-48.
17. Gillette DD, Madsen DB: **The Columbian Mammoth, Mammuthus-Columbi, from the Wasatch Mountains of Central Utah.** *Journal of Paleontology* 1993, **67:**669-680.
18. Schaedler JM, Krook L, Wootton JAM, Hover B, Brodsky B, Naresh MD, Gillette DD, Madsen DB, Horne RH, Minor RR: **Studies of Collagen in Bone and Dentin Matrix of a Columbian Mammoth (Late Pleistocene) of Central Utah.** *Matrix* 1992, **12:**297-307.
19. Knapp M, Hofreiter M: **Next Generation Sequencing of Ancient DNA: Requirements, Strategies, and Perspectives.** *Genes* 2010, **1:**227-243.
20. Krause J, Dear PH, Pollack JL, Slatkin M, Spriggs H, Barnes I, Lister AM, Ebersberger I, Paabo S, Hofreiter M: **Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae.** *Nature* 2006, **439:**724-727.
21. Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M: **Genetic analyses from ancient DNA.** *Annual Review of Genetics, Vol 46* 2004, **38:**645-679.
22. Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M, Paabo S: **Analysis of one million base pairs of Neanderthal DNA.** *Nature* 2006, **444:**330-336.

23. Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, Rabeder G, Krause J, Detter JC, Paabo S, Rubin EM: **Genomic sequencing of Pleistocene cave bears.** *Science* 2005, **309:**597-599.
24. Schwarz C, Debruyne R, Kuch M, McNally E, Schwarcz H, Aubrey AD, Bada J, Poinar H: **New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains.** *Nucleic Acids Res* 2009, **37:**3215-3229.
25. Irwin C, Irwin H, Agogino G: **Wyoming muck tells of battle: Ice Age man vs. Mammoth.** *National Geographic Society Magazine* 1962, **121:**828-837.
26. McGrew PO: **The Rawlins Mammoth.** *Wyoming Geological Society Guidebook* 1961**:**315-317.
27. Haynes G: *Mammoths, mastodonts, and elephants : biology, behavior, and the fossil record.* Cambridge New York: Cambridge University Press; 1991.
28. Moore WS: **Inferring Phylogenies from Mtdna Variation - Mitochondrial-Gene Trees Versus Nuclear-Gene Trees.** *Evolution* 1995, **49:**718-726.
29. Nyakaana S, Arctander P: **Population genetic structure of the African elephant in Uganda based on variation at mitochondrial and nuclear loci: evidence for male-biased gene flow.** *Molecular Ecology* 1999, **8:**1105-1115.
30. Vidya TNC, Sukumar R: **Social organization of the Asian elephant (Elephas maximus) in southern India inferred from microsatellite DNA.** *Journal of Ethology* 2005, **23:**205-210.
31. Vidya TNC, Sukumar R, Melnick DJ: **Range-wide mtDNA phylogeography yields insights into the origins of Asian elephants.** *Proceedings of the Royal Society B-Biological Sciences* 2009, **276:**893-902.
32. Rohland N, Reich D, Mallick S, Meyer M, Green RE, Georgiadis NJ, Roca AL, Hofreiter M: **Genomic DNA Sequences from Mastodon and Woolly Mammoth Reveal Deep Speciation of Forest and Savanna Elephants.** *Plos Biology* 2010, **8**.
33. Roca AL, Georgiadis N, Pecon-Slattery J, O'Brien SJ: **Genetic evidence for two species of elephant in Africa.** *Science* 2001, **293:**1473-1477.
34. Comstock KE, Georgiadis N, Pecon-Slattery J, Roca AL, Ostrander EA, O'Brien SJ, Wasser SK: **Patterns of molecular genetic variation among African elephant populations.** *Molecular Ecology* 2002, **11:**2489-2498.
35. Grubb P, Groves CP, Dudley JP, Shoshani J: **Living African Elephants Belong to Two Species: Loxodonta africana (Blumenbach, 1797) and Loxodonta cyclotis (Matschie, 1900).** *Elephant* 2000, **2:**1-4.
36. Groves CP, Grubb P: **Do *Loxodonta cyclotis* and *L. africana* interbreed?** *Elephant* 2000, **2:**4-7.
37. Debruyne R: **A case study of apparent controversy between molecular phylogenies: the interrelationships of African elephants.** *Cladistics* 2005, **21:**31-50.
38. Roca AL, Georgiadis N, O'Brien SJ: **Cytonuclear genomic dissociation in African elephant species.** *Nat Genet* 2005, **37:**96-100.
39. Fisher DC: **Season of death, growth rates, and life history of the North American mammoths.** In *Mammoth Site Studies Proceedings of the First International Conference on Mammoth Site Studies, Lawrence, KS, March 11-13, 1998*. Edited by West D. 2001: 121-135.
40. Shoshani J, Tassy P: *The Proboscidea : evolution and palaeoecology of elephants and their relatives.* Oxford ; New York: Oxford University Press; 1996.
41. Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, Meyer M, Good JM, Maricic T, Johnson PL, Xuan Z, et al: **Targeted investigation of the Neandertal genome by array-based sequence capture.** *Science* 2010, **328:**723-725.
42. Maricic T, Whitten M, Paabo S: **Multiplexed DNA sequence capture of mitochondrial genomes using PCR products.** *PLoS One* 2010, **5:**e14004.
43. Allendorf FW, Leary RF, Spruell P, Wenburg JK: **The problems with hybrids: setting conservation guidelines.** *Trends in Ecology & Evolution* 2001, **16:**613-622.
44. Rhymer JM, Simberloff D: **Extinction by hybridization and introgression.** *Annual Review of Ecology and Systematics* 1996, **27:**83-109.
45. Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, et al: **Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA.** *Science* 2006, **311:**392-394.
46. Deagle BE, Eveson JP, Jarman SN: **Quantification of damage in DNA recovered from highly degraded samples--a case study on DNA in faeces.** *Front Zool* 2006, **3:**11.

47. Pop M, Phillippy A, Delcher AL, Salzberg SL: **Comparative genome assembly.** *Brief Bioinform* 2004, **5:**237-248.
48. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: **Versatile and open software for comparing large genomes.** *Genome Biol* 2004, **5**.
49. Phillippy AM, Schatz MC, Pop M: **Genome assembly forensics: finding the elusive mis-assembly.** *Genome Biol* 2008, **9**.
50. Schatz MC, Phillippy AM, Shneiderman B, Salzberg SL: **Hawkeye: an interactive visual analytics tool for genome assemblies.** *Genome Biol* 2007, **8**.
51. Brady A, Salzberg SL: **Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models.** *Nat Methods* 2009, **6:**673-U668.
52. Miller W, Drautz DI, Ratan A, Pusey B, Qi J, Lesk AM, Tomsho LP, Packard MD, Zhao FQ, Sher A, et al: **Sequencing the nuclear genome of the extinct woolly mammoth.** *Nature* 2008, **456:**387-U351.
53. Rogaev EI, Moliaka YK, Malyarchuk BA, Kondrashov FA, Derenko MV, Chumakov I, Grigorenko AP: **Complete mitochondrial genome and phylogeny of Pleistocene mammoth Mammuthus primigenius.** *Plos Biology* 2006, **4:**e73.
54. Posada D: **jModelTest: Phylogenetic model averaging.** *Molecular Biology and Evolution* 2008, **25:**1253-1256.
55. Drummond AJ, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees.** *BMC Evol Biol* 2007, **7**.
56. Rohland N, Malaspinas AS, Pollack JL, Slatkin M, Matheus P, Hofreiter M: **Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup.** *Plos Biology* 2007, **5:**1663-1671.
57. Meyer M, Briggs AW, Maricic T, Hober B, Hoffner B, Krause J, Weihmann A, Paabo S, Hofreiter M: **From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing.** *Nucleic Acids Res* 2008, **36:**e5.

Chapter 2

Quantitative PCR as a predictor of aligned ancient DNA read counts following targeted enrichment

JACOB ENK, JEAN-MARIE ROUILLARD, and HENDRIK POINAR
BioTechniques 2013, 55(6):300-309; doi 10.2144/000114114

2.1 PREFACE

Based on experimental results from our projects and others in the field, sequencing multiple mitogenomes from very poorly-preserved remains like those of Columbian mammoths would require some kind of targeted enrichment prior to economical amounts of high-throughput sequencing. After considering multiple potential enrichment approaches (PCR, array-based capture, synthetic amplicon-based capture, or multi-primer extension capture), we settled on using in-solution biotinylated RNAs, which provide the best balance of economy and scalability available on the market. However ancient DNA enrichment is not well understood, and thus can be quite unpredictable. In this study we test whether standard sample quality assays (namely qPCR) can be used to assess the potential success of enrichment, and as such whether qPCR can serve as a guide to improving the efficiency of pooled multiplex sequencing. We not only confirmed qPCR as a useful decision-making tool in aDNA targeted enrichment projects, but also uncovered features of enrichment behavior that can further guide experimental design of future, larger-scale projects.

2.2 ABSTRACT

Targeted DNA enrichment through hybridization capture (EHC) is rapidly replacing PCR as the method of choice for enrichment prior to genomic resequencing. This is especially true in the case of ancient DNA (aDNA) from long-dead organisms, where targets tend to be highly fragmented and outnumbered by contaminant DNA. However, the behavior of EHC using aDNA has been quite variable, making success difficult to predict and preventing efficient sample equilibration during multiplexed sequencing runs. Here, we evaluate whether quantitative PCR (qPCR) measurements of aDNA samples correlate with on-target read counts before and after EHC. Our data indicate that not only do simple target qPCRs correlate strongly with high-throughput sequencing (HTS) data but that certain sample characteristics, such as overall target abundance as well as experimental parameters (e.g., bait concentration and secondary structure propensity), consistently influenced enrichment of our diverse set of aDNA samples. Taken together, our results should help guide experimental design, screening strategies, and multiplexed sample equilibration, increasing yield and reducing the expected and actual cost of aDNA EHC high-throughput sequencing projects in the future.

2.3 INTRODUCTION

Recently, high-throughput sequencing (HTS) has enabled reconstruction of large genomic regions from heavily degraded paleontological and archival specimens (1-3). These results promise to expand ancient DNA (aDNA)-based genomic projects to greater breadth and rigor. However, aDNA extracts are typically dominated by DNA from microorganisms, with genomic regions of interest often representing only a tiny fraction of the total genomic content of the target organism. Therefore, even with the significant amount of raw sequence data generated by HTS, obtaining multi-fold coverage of a target genomic region can be prohibitively expensive. As such, most aDNA extracts require targeted enrichment prior to HTS, especially when the goal is to pool and sequence multiple individuals to sufficient coverage depth.

Target aDNAs are typically well under the practical length for resequencing with overlapping PCR amplicons. Thus, enrichment through hybridization capture (EHC) coupled with HTS is rapidly replacing targeted PCR as the method of choice for reconstructing several and/ or long aDNA genomic targets (4-18). However, the reported on-target read proportions obtained using EHC with aDNA vary considerably (Table 1). These results hinder efficient sample

equilibration in multiplexed sequencing runs, and are a primary barrier to EHC's routine use by many groups working with aDNA.

EHC variation is unsurprising since the process is driven by mostly unevaluated interactions between experimental parameters, target choice, and DNA sample characteristics (19,20). Of these, the absolute and/or relative abundance of targets are likely particularly relevant, especially given the highly complex non-target DNA content typical of aDNA extracts. Fortunately, quantitative PCR (qPCR) can cheaply, albeit indirectly, estimate target abundance and has been used extensively to evaluate aDNA sample quality (2,21-25). Recently, Wales and colleagues (25) demonstrated that qPCR measures of target and non-target molecules in aDNA samples correspond ordinally to HTS data from non-enriched sequencing libraries. What remains unanswered, however, is whether qPCR-based metrics from aDNA samples correspond to reads-on-target after enrichment. Here we examine this question experimentally to determine if qPCR-based predictions could make EHC-HTS projects more predictable and efficient. In addition, we discuss experimental parameters that impact enrichment results and may thereby inform future experimental design strategies.

Method summary

We quantified a 49 bp locus of the mammoth *12S* rRNA mitochondrial gene on 11 different mammoth DNA extracts indexed libraries, and whole mitochondrial DNA-enriched libraries. The same libraries were high-throughput sequenced, and reads were aligned to the target. qPCR values, bait concentration, and secondary structure propensity were compared against on-target read counts, enrichment rates, and mitogenome coverage patterns.

Table 1. Summary of aDNA enrichment statistics

Ref. #	Target organism	Target genome	Target length (kbp)	Enrich. method	Mapped before enrichment	Unique before enrichment	Mapped after enrichment	Unique after enrichment
(6)	*H. sapiens* (Neandertal)	Mitochondrial	16.6	bDNAs	0.0005 - 0.005%	-	18.2 - 40.2%	1.1 - 7.1%
(7)	*H. sapiens* (Neandertal)	Nuclear + mitochondrial	1400 + 16.6	Array	~0.0002%	-	~37%	-
(10)	*H. sapiens* (AMHS)	Mitochondrial	16.6	bDNAs	-	-	23%	0.3%
(11)	*H. sapiens* (Denisovan)	Mitochondrial	16.6	bDNAs	-	0.047% [1]	12.4%	2.6%
(4)	*Zea mays*	Nuclear + chloroplast	560 + 140	bRNAs	0.2 - 3.5%	0.002 - 0.130%	6.6 - 74.8%	0.006% - 1.86%
(14)	*H. sapiens + Yersinia pestis*	Mitochondrial + pPCP1 plasmid	16.6 + 9.6	bDNAs	-	-	0.3 - 49.0%	0.017 – 6.22% [2]
(5)	*Yersinia pestis*	Genome + plasmids	4400	Array	-	-	64.3% [3]	13.7%
(12)	*Dermoptera*	Mitochondrial	16.6	bDNAs	-	-	1.6 - 95.4%	-
(15)	Mammalia	Mitochondrial	16- 17	bDNAs	-	-	<0.01 - 52.6%	0.002 - 26.2% [1]
(9)	*Castor fiber*	Mitochondrial control region	0.65	bDNAs	-	-	0.1 - 62.7%	0.05 - 0.78% [1]
(16)	*Mycobacterium tuberculosis*	Genome	45	bRNAs	-	-	12.2%	0.04 - 4.07% [4]
(8)	*H. sapiens* (AMHS)	Chromosome 21	29800	bDNAs	-	-	46.8%	8.4%
Current [5]	*Mammuthus* sp.	Mitochondrial	16.5	bRNAs	0.0003 - 0.15%	0.0003 - 0.09%	0.35 - 46.0%	0.05 - 2.76%

[1] Data provided by corresponding author. [2] Mitogenome only. [3] CO92 chromosome only. [4] Per-locus values. [5] Analyzed at normalized sequencing depth. bDNAs refers to use of biotinylated DNA baits for capture in suspension, according to either Reference 47 or Reference 48. bRNAs refers to biotinylated RNA baits, such as SureSelect (Agilent) or MYbaits (MYcroarray). Array refers to use of oligonucleotide microarray for capture. AMHS refers to anatomically modern humans from the Late Pleistocene age (~125,000 to 12,000 years ago).

2.4 MATERIALS AND METHODS

As depicted in Figure 1, we selected 11 DNA extracts from various Pleistocene-age *Mammuthus* specimens of a wide range of depositional contexts and preservation levels, as well as 2 associated non-mammoth controls. We then used a highly sensitive proboscidean-specific qPCR assay to measure the amplifiable molecules at a 49 bp locus (including primers) of the mammoth mitochondrial *12S rRNA* gene using 2 template concentrations (1× and 0.1×). After converting the extracts to Illumina sequencing libraries (26,27) , we quantified the *12S* locus again, along with the total number of adapted molecules using a global qPCR assay (26). We then enriched the libraries in duplicate using an in-solution EHC technique with two different concentrations of biotinylated RNA baits matching several known mammoth mitochondrial genome haplogroups (28-30). Following enrichment, we again quantified both the *12S* locus and total adapted molecules. Finally, we sequenced the non-enriched and enriched libraries on

an Illumina HiSeq platform and measured on-target read proportions, coverage patterns, and enrichment rates, evaluating how these corresponded to qPCR metrics, bait concentration, and bait properties.

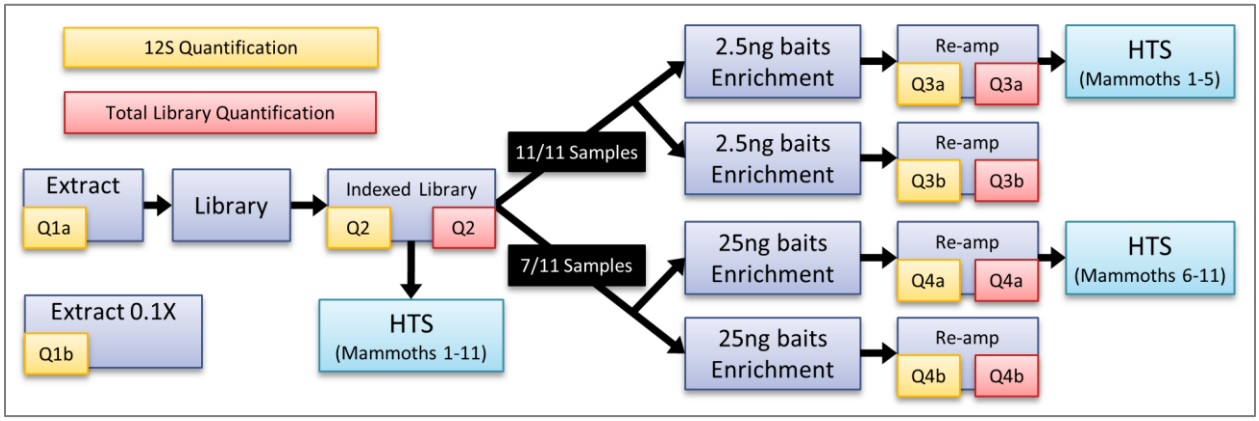


Figure 1. Flowchart depicting main experimental stages and points of quantitative PCR measurement and high-throughput sequencing. 12S quantification refers to qPCR of a 49 bp section of the *12S* mitochondrial gene. Total quantification refers to qPCR of entire libraries using primers matching the sequencing adapters. Samples were extracted and quantified both undiluted and at 0.1× concentration, then converted to indexed sequencing libraries and sequenced. Portions of those indexed libraries were subsequently subjected to replicate enrichments, with 2.5 ng and/or 25 ng of baits, and were sequenced after re-amplification.

2.4.1 Quantitative PCR assays

Quantitative PCRs were performed using a CFX96 real-time PCR instrument (BioRad, Hercules, CA). For the mammoth 49 bp *12S* assay, we paired 2 previously-published (31) primers (B1_Forward: CCCTAAACTTTGATAGCTACC, A1_Reverse: GTAGTTCTCTGGCGGATAGC). The target has GC content of 42.9%, compared with the mammoth mitogenome average of roughly 38%. As a standard, we used a previously generated amplicon encompassing the mammoth *12S rRNA* gene quantified with a GeneQuant Pro UV spectrophotometer (Amersham, now GE Healthcare; Freiburg, Germany). An optimal primer annealing temperature (54°C) was determined by a temperature gradient qPCR experiment. The assay exhibits near single-copy sensitivity and is accurate to at least 10 copies. Test amplifications from human, xenarthran, coprolite, soil, and chicken DNAgenerated sporadic and off-size products. For library target and total quantifications, the standards were amplicons of the target that were converted to a non-indexed Illumina sequencing library (26). Total library quantifications used primers matching the universal internal adapter sequence (IS7_short_amp. P5: ACACTCTTTCCCTACACGAC; IS8_short_amp.P7: GTGACTGGAGTTCAGACGTGT) (26).

Each *12S* qPCR reaction used 3 µL (extracts) or 1 µL (libraries) template and followed the recipes and thermal programs in Supplementary Table S6, with plate reads after annealing steps. In *12S* assays, dissociation ramps from at least 65°C to at least 89°C were used to confirm the positive product (74.5 ± 0.5°C). At least three different concentrations of standard were included in duplicate in each assay and values were estimated using the automatic threshold calculation by the CFX Manager software version 3.0. Extracts were screened using both 1× and 0.1× versions diluted in EBT (QIAGEN Buffer EB brought to 0.05% Tween-20) (QIAGEN, Dusseldorf, Germany). For indexed libraries, *12S* quantifications used 0.1× versions and total quantifications used either 0.001× or 0.0001× versions. Duplicate quantifications were averaged, with values projected to be in the undiluted sample reported even when less than 10 copies (Supplementary Table S1).

2.4.2 DNA extraction and library preparation

We extracted DNA from roughly 100 mg subsamples of 11 *Mammuthus* specimens and a Pleistocene *Mylodon darwinii* specimen (Supplementary Table S1) in aDNA-dedicated facilities using an organic protocol with filter concentration (2). Libraries used 50 µL of each extract diluted to various concentrations; the 0.1× concentration

was used when inhibition was estimated to be >90% (Supplementary Table 1). Library preparation followed Meyer and Kircher (26), with each purification step performed with the MinElute PCR purification kit (QIAGEN) eluting in 20 µL EBT.

Libraries were diluted to between 1× and 0.1× concentration in EBT based on available volume (Supplementary Table S1) and then index-amplified with unique P5 and P7 indexing primers (27) using 5 µL of template according to the recipe and thermal program in Supplementary Table S6. Reactions were removed from the cycler and placed on ice upon reaching similar amplification points (11–23 cycles), estimated to be near PCR plateau. After review of the trace data without baseline correction, it was apparent that Mammoth 7 reached plateau by cycle 9. Reactions were purified with the MinElute kit to 20 µL EBT.

2.4.3 Bait design

For bait references, we used the light strand of several published *Mammuthus mitochondrial* genomes (30,32) from four different haplogroups (NCBI accession no.s EU153453, JF912199, EU153456, EU153447) and a complete mitogenome sequence from M*ammut americanum* (American Mastodon, NCBI accession no. EF632344)(33). We also included a 734 bp region of the cytochrome b, tRNA-Pro, tRNA-Thr from a fifth mammoth haplogroup (NCBI accession no. FJ015496) (29) The VNTRs of the control regions were removed prior to bait design. We also selected a number of nuclear loci, which we extracted from the *Loxodonta africana* genome sequence (NCBI RefSeq Assembly ID: GCF_000001905.1) and a number of entries in the NCBI nucleotide database and did not repeat-mask (Supplementary Datafile 1). With these, we designed 100 bp baits tiled every 6 bp along the targets and then collapsed identical baits, providing 8530 unique mitochondrial and 7996 unique nuclear baits, which MYcroarray manufactured in 2 separate batches as part of a MYbaits enrichment kit. Bait dimer and hairpin scores were calculated with the Primer3 plug-in (34) in Geneious Pro version 5.6.5 (Biomatters, Auckland, New Zealand) using a salt concentration setting matching capture conditions (0.82 M).

2.4.4 Enrichment and re-amplification

We used the MYbaits targeted enrichment kit protocol with some adjustments:

- Enrichments used 3.4 µL of indexed library (an estimated 1.5–18.8 ng of template based on total qPCR quantification) rather than the recommended 100–500 ng, as we have successfully enriched concentrations this low previously (data not shown).
- Given the short length of the genomic regions targeted, we scaled down the input bait concentration from the typical 500 ng, which is appropriate for up to 2 Mb targets. All 13 libraries as well as 3.4 µL of EBT were enriched in replicate using 2.5 ng total input of each bait set, and then 9 libraries (Mammoths 4, 6–11 and associated blanks) and EBT were enriched in replicate using 25 ng.
- We used blocking oligonucleotides matching both strands rather than one strand of double-indexed library (26). These blocks were also used as bait diluents, maintaining 500 ng oligonucleotides per reaction.
- The second 2.5 ng bait experiment and both 25 ng bait experiments were erroneously performed using 95% and 105% of the recommended amounts of Block 1 and Block 2, respectively.
- Reactions hybridized for between 37 and 39 h at 45°C with 45°C wash steps, rather than 65°C. Previous experiments indicated this lower temperature would improve the capture of short, rare inserts typical to our samples (data not shown).
- We used 20 µL of hydrophobic Dynabeads MyOne Streptavidin T1 beads (Invitrogen, Life Technologies, Carlsbad, CA) for enrichment cleanup rather than hydrophilic C1 beads. A previous version of the kit called for hydrophobic M280 beads.

Each enrichment elution was purified to 30 µL EBT with MinElute. An 8 µL aliquot was then amplified using primers that match the 5' flanks of each index (IS5_reamp. P5: AATGATACGGCGACCACCGA; IS6_reamp.P7: CAAGCAGAAGACGGCATACGA) (26) following the recipe and program in Supplementary Table S6. Reactions were put on ice and purified to 21 µL EBT with MinElute.

2.4.5 Sequencing, demultiplexing

Non-enriched and enriched libraries were sequenced on two separate lanes of a HiSeq 1000 Illumina flowcell. cBot cluster generation and sequencing employed the version 3 chemistry and a 2 × 101 bp dual 8 bp indexing protocol, using the alternative primer mixes from the TruSeq Dual Index Sequencing Primer Kit (paired-end). Raw data were processed with HCS version 1.5.15.1 and RTA version 1.13.48.0. File conversion and demultiplexing using each 7 bp reverse index (requiring a 100% match) were performed using CASAVA version 1.8.2.

2.4.6 Curation

Reads were trimmed of the 3' universal adapter sequence (AGATCGGAAGAGC) using cutadapt version 1.2 (35), requiring a 1 bp overlap (-O 1) and tolerating 16% sequence divergence (-e 0.16). Paired reads were then merged with FLASH version 1.0.3 (36) , requiring an 11 bp overlap (-m 11) and tolerating 15% overlap divergence (-x 0.15). These merged reads were then combined with the non-mergeable Read 1s to generate a final data set for downstream analysis. We used random 3 million read subsets from each library to analyze on-target raw and unique read proportions, while coverage patterns were analyzed using these and full read sets.

2.4.7 Alignment

Using Burrows-Wheeler Aligner (BWA) version 0.6.1-r104 (37) we aligned reads 24 bp and longer to the light strand of a mammoth mitochondrial genome with the VNTR masked with 10 Ns (GenBank Accession No: JF912200.1, a haplogroup C/Clade I mammoth from Alaska). We used default *aln* parameters except disabled seeding (-l 9999) and allowed 2% of assemblies to be missed assuming a 2% error rate (-n 0.02). We also mapped reads to the nuclear targets (42,127 bp), but given the dominance of repeat-region reads in these alignments, we did not analyze this data further, though alignment rates to hard-masked nuclear targets are shown in Supplementary Table S5. Resulting SAM files were converted to BED with BEDOPS version 2.2.0 (38), and using these we generated complexity curves by randomly sampling raw mapped reads in increments corresponding to 100,000 total library reads up to the available total reads per sample. Unique reads were identified by collapsing reads with unique 5' and 3' coordinates and direction. Fragment misincorporation plots, which exhibit typical aDNA deamination patterns, were generated with mapDamage version 2.0 (39) (Supplementary Figures 1a–k).

2.5 DATA AVAILABILITY

HTS sequence data for this project can be found on the NCBI SRA at Study Accession #SRP026317.

2.6 RESULTS AND DISCUSSION

The qPCR results from each mammoth sample, as measured using (*i*) original DNA extracts, (*ii*) non-enriched sequencing libraries, and (*iii*) enriched sequencing libraries, are listed in Table 2 and Supplementary Table S1 along with the HTS read counts and proportions of reads that map to the mitochondrial genome out of random 3 million read subsets. The nature of qPCR assays requires their values be carefully compared, since they estimate two different parameters. Our mammoth *12S* assay measures locus frequency among all fragments ~49 bp and greater, and therefore correlation of this result to on-target HTS reads will vary depending on the endogenous insert length distribution as well as any intra-target abundance biases. The total assay estimates copy numbers of a heterogeneous length distribution, so this accuracy is also dependent on the overall library length distribution and amplification biases.

Table 2: qPCR and HTS data

Sample	Extract		Indexed library		Non-enriched hits data (3 million read subset)			2.5 ng enrichment replicate 1		2.5 ng enrichment replicate 2		25 ng enrichment replicate 1		25 ng enrichment replicate 2		Enriched HTS data (3 million read subset)		
	12S, 1× Q1a	*12S*, 0.1× Q1b	Total Q2	*12S* Q2	Mitogenome Reads		Unique mito.	Total Q3a	*12S* Q3a	Total Q3b	*12S* Q3b	Total Q4a	*12S* Q4a	Total Q4b	*12S* Q4b	Mitogenome reads		Unique mito.
Mammoth 1	246	899	6.7E+09	10423	103	0.0034%	0.0032%	2.6E+09	906800	2.3E+09	714100	-	-	-	-	221,776	7.4%	0.06%
Mammoth 2	2333	3569	1.9E+10	199500	3788	0.1263%	0.0729%	1.4E+10	6211000	1.8E+10	1993000	-	-	-	-	1,379,489	46.0%	0.18%
Mammoth 3	478	691	1.3E+10	37945	472	0.0157%	0.0129%	3.2E+09	1797000	1.1E+10	2047500	-	-	-	-	587,951	19.6%	0.06%
Mammoth 4	1550	2238	6.6E+09	5102	1538	0.0513%	0.0509%	6.2E+09	104965	4.3E+09	138900	7.9E+09	487150	1.0E+10	246850	782,089	26.1%	2.76%
Mammoth 5	199	23081	6.5E+09	5105	249	0.0083%	0.0082%	1.2E+10	395950	5.7E+09	288600	-	-	-	-	498,381	16.6%	1.32%
Mammoth 6	21	32	7.5E+09	161	85	0.0028%	0.0028%	3.5E+09	15735	3.4E+09	4073	4.7E+09	43935	5.0E+09	30810	82,523	2.8%	0.05%
Mammoth 7	27	241	3.4E+09	44	11	0.0004%	0.0004%	7.0E+08	19760	6.6E+08	4	9.3E+08	21455	8.3E+08	8502	11,321	0.37%	0.30%
Mammoth 8	87	360	8.7E+09	107	9	0.0003%	0.0003%	3.3E+09	19095	5.3E+09	8096	6.6E+09	36110	1.3E+10	89045	10,542	0.35%	0.22%
Mammoth 9	135	2504	7.4E+09	123	432	0.0144%	0.0140%	5.5E+09	5742	5.2E+09	6797	8.1E+09	15805	8.1E+09	7487	158,052	5.3%	0.27%
Mammoth 10	323	2721	1.3E+10	16505	2560	0.0853%	0.0808%	4.9E+09	340200	1.2E+10	319500	1.2E+10	869000	8.8E+09	558250	499,317	16.6%	1.06%
Mammoth 11	6	13	2.2E+09	19	7	0.0002%	0.0002%	5.1E+08	9	5.6E+08	26	3.5E+09	48730	2.4E+09	24300	19,510	0.65%	0.14%
Extr. Blank	<1	0	3.0E+09	0	4	0.0001%	0.0001%	8.0E+08	3	5.4E+08	8	1.9E+09	116	3.4E+09	194	104	0.00%	0.00%
Mylodon	0	0	2.7E+10	0	8	0.0003%	0.0003%	1.4E+10	5	1.1E+10	18	1.2E+10	149	1.7E+10	75	732	0.02%	0.01%

Q# labels refer to qPCR measurement points depicted in Figure 1. *12S* refers to the 49 bp mitochondrial *12S rRNA* gene qPCR assay. Total refers to the overall library qPCR assay that uses adapter-specific primers. All qPCR values are corrected for template dilution factor and therefore reflect inferred quantities in the undiluted template source. Mitogenomic read counts are restricted to mapped reads of at least 14 bp in length. Percentages are of 3 million random sequencing reads.

Figure 2 plots qPCR measures against mitogenome reads both before and after enrichment. All *12S* qPCR quantifications show significant positive power correlations with raw reads on-target before and after enrichment, with undiluted extract and indexed library quantifications correlating most strongly both before (R^2 = 0.69, P = 0.002 and R^2 = 0.66, P = 0.003, respectively) and after enrichment (R^2 = 0.70, P = 0.001 and R^2 = 0.80, P < 0.001, respectively). Interestingly, the ratios between the target values and total qPCR values (referred to hereafter as *12S*:Total) are weaker predictors of HTS read counts, likely due at least in part to the well-documented length bias inherent to the polymerase used for our qPCR assays (AmpliTaq Gold; Invitrogen) (40).

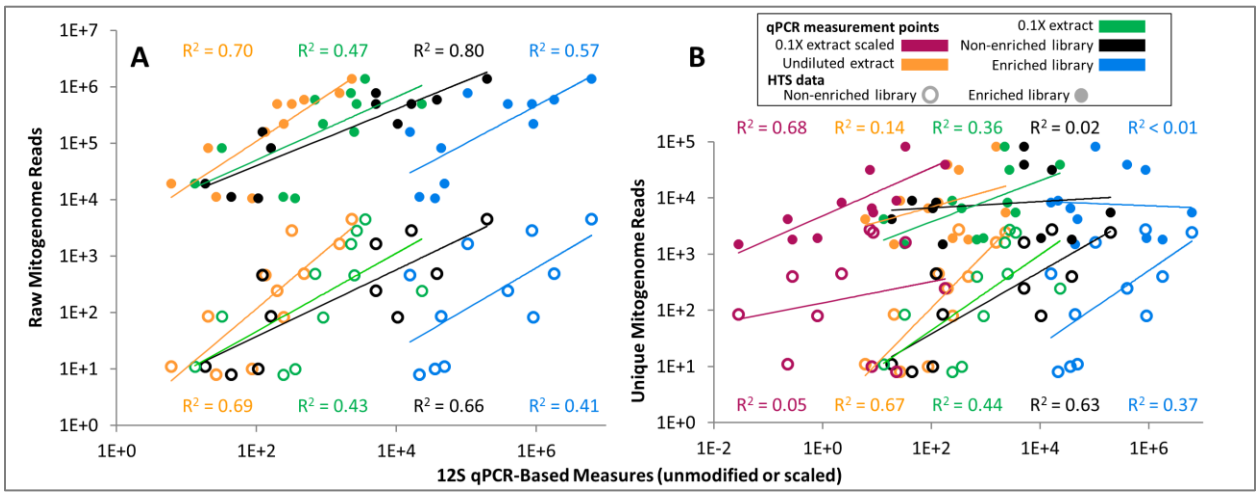


Figure 2. Relationships between 12S qPCR-based measures and on-target read counts. Data are depicted for both raw (A) and unique (B) read counts analyzed at 3 million total reads per sample. Open circles correspond to data points from non-enriched libraries, closed circles to enriched libraries. All qPCR values are dilution-corrected. The 0.1× extract scaled markers represent the *12S* qPCR values scaled by concentration used for indexing and indexing cycles [the result of the function (0.1× qPCR value) × (library concentration used for indexing) ÷ (1.3 ^ indexing cycles)].

This strong correspondence between the single-locus target qPCR measurements and mitogenome raw reads after enrichment is encouraging. However, raw on-target read proportions are often irrelevant in many sequencing projects, with the reliability of consensus calls more appropriately gauged by unique read coverage. Prior to enrichment, qPCR values correlate comparably in strength between unique and raw on-target reads, but this correlation is insignificant following enrichment (best correlation $R^2 = 0.36$, $P = 0.0527$). This makes sense since the enriched libraries show a range of complexities (Figure 3). Unsurprisingly, the number of indexing cycles correlates strongly with measures of post-enrichment complexity, such as overall duplication rate (positive power correlation $R^2 = 0.83$, $P = 0.0001$) and relative increases in unique on-target reads between 100,000 and 3 million reads sequencing depth (negative power correlation $R^2 = 0.90$, $P < 0.0001$). Indexing cycles also predict raw reads on-target after enrichment with comparable strength as undiluted extract qPCRs ($R^2 = 0.74$, $P < 0.0007$). As such, scaling the original 0.1× extract target qPCR values by indexing cycles greatly improves their correlation with post-enrichment unique read count, especially when also scaled by the dilution factor of the library used in the indexing amplification reaction (Figure 2, "0.1× extract scaled").

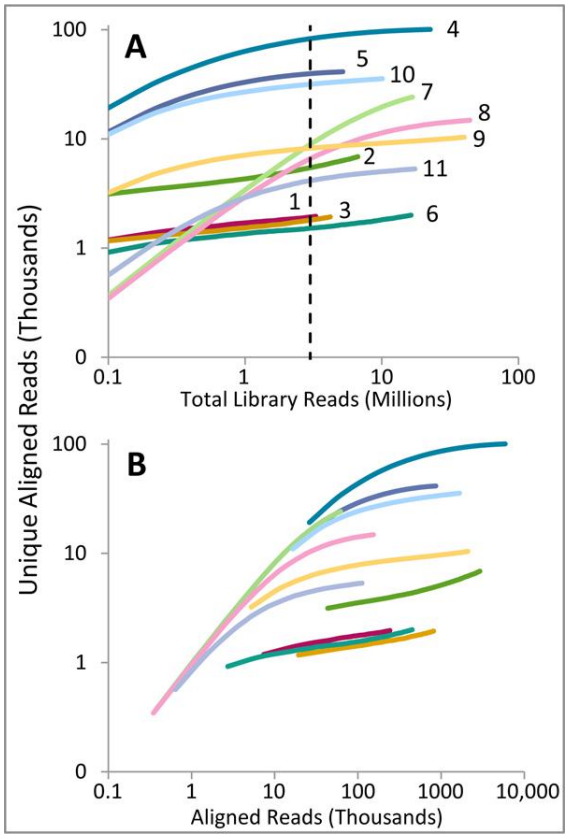


Figure 3. Library complexity curves. Complexity curves in terms of total reads sequenced (A) and raw on-target reads sequenced (B), with mammoth numbers adjacent to each line. Curves were generated by randomly sampling raw mapped reads in increments corresponding to 100 thousand total library reads up to the available total reads per sample (3.3 to 44.1 million). Raw mapped reads with identical 5' and 3' end coordinates and strand origin (direction) were collapsed to determine unique counts. The dotted line in (A) marks 3 million total read depth, at which on-target read counts and enrichment rates were analyzed.

These observations demonstrate that simple qPCR metrics can indeed predict on-target read counts after enrichment, and therefore could form the basis of equilibration schemes for efficient sequencing of enriched aDNA libraries. However, the relationship between qPCR and reads-on-target is nonlinear, and for any given combination of samples, targets, enrichment techniques, and data analysis strategies, pilot screening of a subset of enriched samples would be beneficial for accurate equilibration.

In addition to qPCR read count correlations, we also examined variables that potentially influence enrichment

rate—the ratio between on-target read proportions after and before enrichment (Table 3)—as well as correlations between bait properties and the intra-target variation in enrichment rate and pre- and post-enrichment coverage patterns (Supplementary Figure S2). Correlations between these properties and per-base mitogenome coverage are included in Supplementary Tables S3 and S4.

Table 3: Enrichment rates (proportion after enrichment:proportion before enrichment) in 12S: Total qPCR values and mitogenome-alignable HTS reads.

Sample	*12S*:Total, 2.5 ng baits			*12S*:Total, 25 ng baits			HTS raw	HTS unique
	Rep. 1	Rep. 2	Average	Rep 1.	Rep 2.	Average		
Mammoth 1	**227**	201	214	-	-	-	2053	19
Mammoth 2	**41**	11	26	-	-	-	332	2
Mammoth 3	**200**	69	135	-	-	-	1200	4
Mammoth 4	**22**	42	32	79	31	55	473	50
Mammoth 5	**44**	64	54	-	-	-	1894	151
Mammoth 6	210	55	133	**435**	287	361	893	16
Mammoth 7	2217	0	1109	**1820**	803	1311	995	788
Mammoth 8	463	124	294	**446**	568	507	1100	686
Mammoth 9	62	78	70	**116**	55	86	350	19
Mammoth 10	57	21	39	**57**	52	54	180	12
Mammoth 11	2	5	4	**1601**	1155	1378	2534	532
Extr. blank	-	-	-	-	-	-	24	31
Mylodon	-	-	-	-	-	-	37	20

Enrichment rates depicted bold and underlined derive from enriched libraries that were sequenced.

Whether measured by increases in the qPCR-derived *12S*:Total or on-target HTS reads, enrichment rates correlate negatively with starting abundances, a relationship observed in other aDNA enrichment data sets (4,7). This suggests a pattern of diminishing returns wherein the degree to which samples could benefit from additional rounds of enrichment might decrease as the endogenous portion approaches a certain threshold. In addition to abundances, the number of indexing cycles (related to starting input DNA) is again a strong predictive variable, correlating negatively with unique enrichment rate (R^2 = 0.84, P < 0.0001).

Bait concentration was influential as well. As measured by increases in the *12S*:Total ratio, enrichment rates ranged from 22- to 2,217-fold (mean = 221) when enriched with 2.5 ng of baits, with the 2 lowest-copy libraries (Mammoths 7 and 11) failing to enrich in 1 and both replicates, respectively. With 25 ng of baits, enrichment rates improved 15- to 1374-fold over their 2.5 ng counterparts, with Mammoth 7 and 11 successfully enriching in both replicates. However, inter-replicate enrichment consistency did not improve significantly when bait concentration was increased (Student's *t*-test on inter-replicate coefficients of variation P = 0.29).

Bait coverage, which we designed to be deeper (and more diverse) in mitogenomic regions known to be polymorphic in extinct proboscideans, may have also affected enrichment. For all mammoth samples, whether analyzed at 3 million reads or at total available depth, bait and read coverage depth almost universally increased in positive correlation with aligned read coverage after enrichment. Bait coverage also correlates positively with raw enrichment rates. While not as strong an association as seen in other enrichment studies (e.g., Reference (41), this suggests that relatively minor bait coverage variation (mean coverage = 51×, standard deviation = 12.6×) impacts post-enrichment coverage.

The propensity for individual baits to form secondary structures also appeared to impact enrichment. As with bait coverage, correlations between coverage or enrichment rates and per-base average bait hairpin and dimer scores became universally stronger following enrichment (in this case, negatively). However, it is difficult to rule out inter-locus amplification bias as the origin of this pattern; only two samples (Mammoths 9 and 10) show low correlations between regional duplication rate and unique coverage depth, and for these samples unique enrichment rates do not correspond to bait properties. That amplification biases appear to dictate coverage patterns in this data set clearly encourages the use of amplification-minimal techniques (15,42,43) with less biased DNA polymerases (40,44) or emulsion PCR (45).

Both qPCR measures and experimental design features correlate with post-enrichment read counts, enrichment rates, and coverage metrics. More sophisticated qPCR strategies, such as multi-locus and/or techniques that simultaneously estimate the target insert length distribution (22,23,46) might improve predictive power. Predicting unique target reads from qPCR metrics is a more complicated task, requiring complexity-based modification to be accurate. Therefore, normalizing complexity between samples as much as possible is obviously recommended, such as by equalizing starting library molarity prior to indexing as well as indexing amplification cycles. Since higher bait concentration and bait coverage corresponded to higher enrichment, maintaining even target coverage depth in bait sets and/or modifying bait sequences to reduce self-dimer/hairpin propensity might also improve enrichment consistency. However, it is unclear how these correspondences depend on hybridization and washing conditions, especially temperature and salt concentration, among other things. Given the demonstrated power of EHC for aDNA, such knowledge gaps encourage continued systematic evaluation of how enrichment is affected by experimental parameters and sample characteristics, beginning with variables that are cheaply assessed for large sample sets or easily controlled during experimental design.

2.7 AUTHOR CONTRIBUTIONS

All authors designed the experiments. JE executed the experiments, performed analyses, and prepared the manuscript. All authors edited the manuscript.

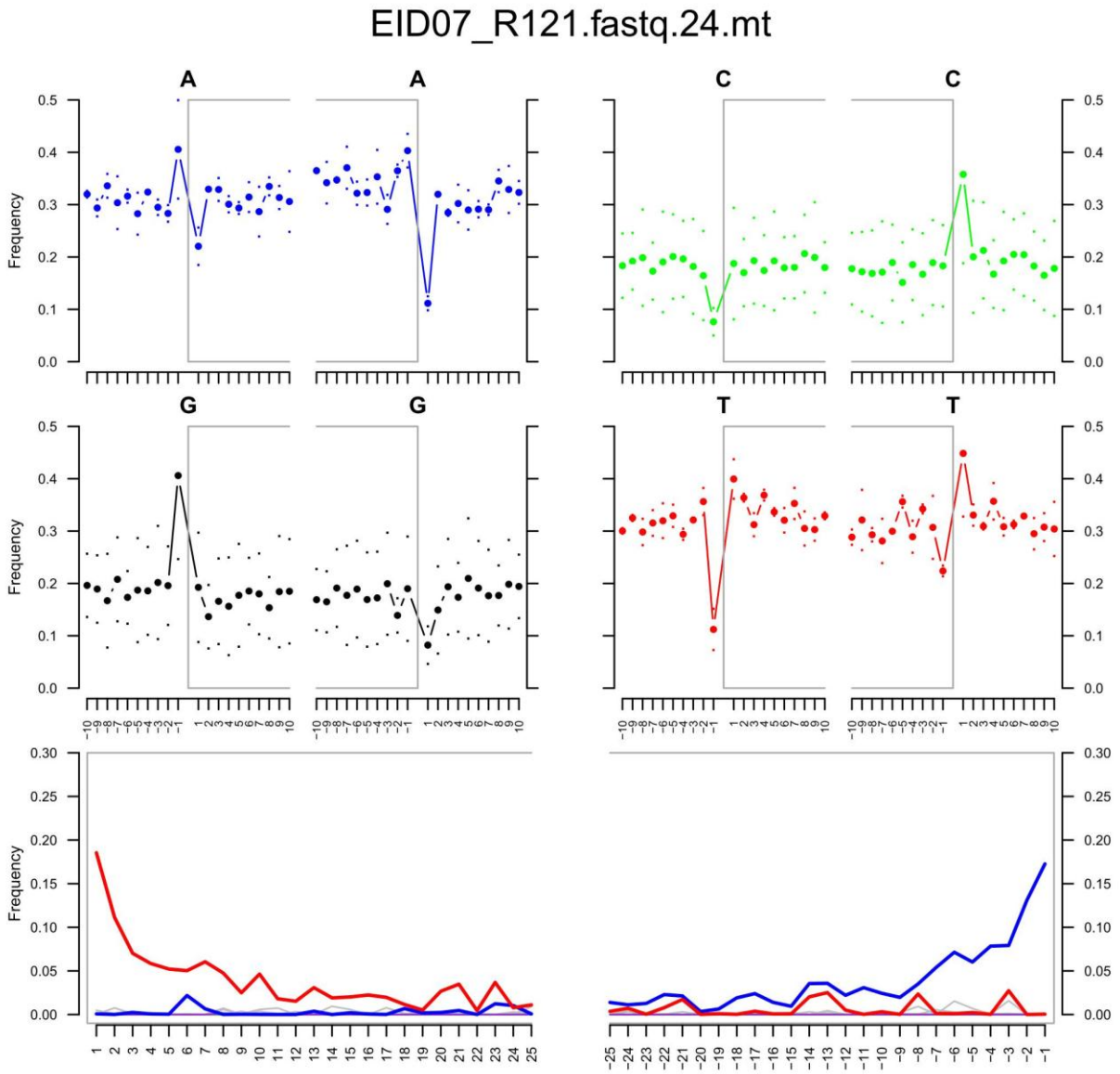
2.8 ACKNOWLEDGMENTS

We thank the entire McMaster Ancient DNA Centre for guidance and discussion during the experiments and manuscript preparation. S. Horn, J. Krause, and S. Sawyer graciously provided accessory data to their published works referenced here. M. Clementz, D. Fisher, C.R. Harington, P. Matheus, and G. Zazula and their respective institutions provided mammoth samples. We also thank the McMaster Farncombe Family Digestive Health Research Institute, especially C.E. King, for performing the sequencing experiments and data pre-processing. Comments from two anonymous reviewers greatly improved the manuscript. This research was funded by NSERC and CRC grants to HP and contributions from MYcroarray.

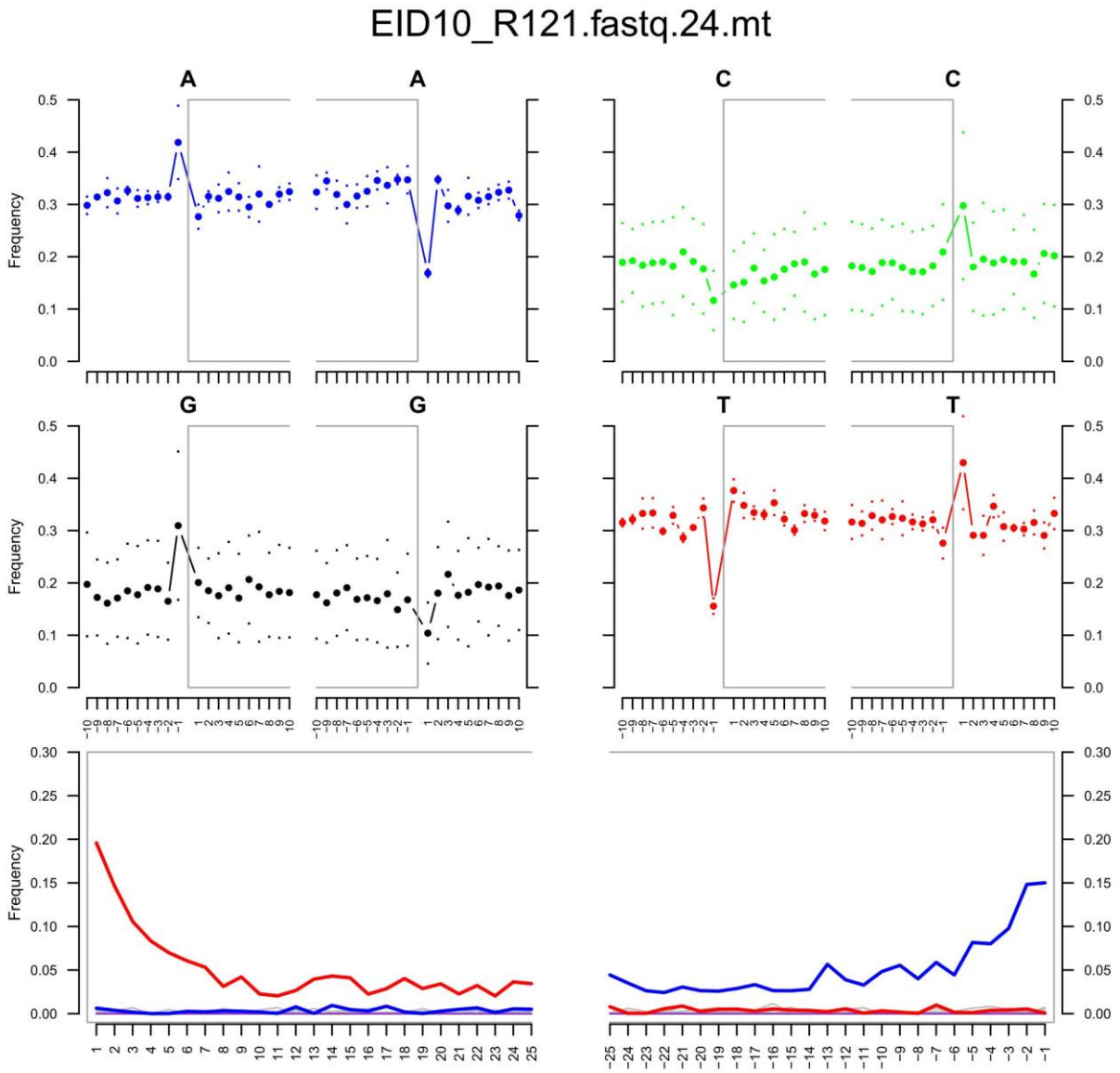
2.9 COMPETING INTERESTS

JMR has financial interest in MYcroarray, the company providing the enrichment kit for this study.

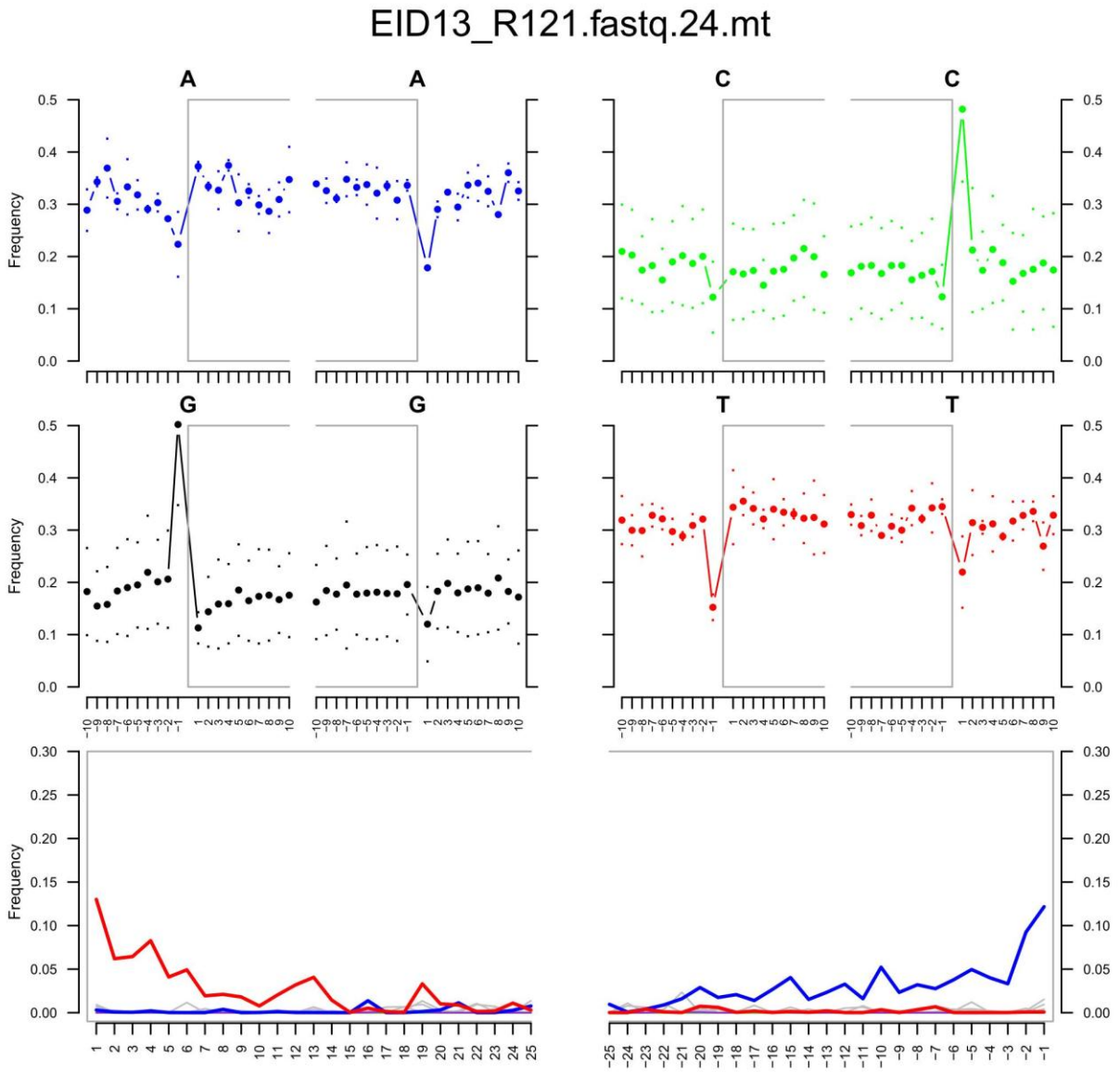
2.10 SUPPLEMENTARY FIGURES



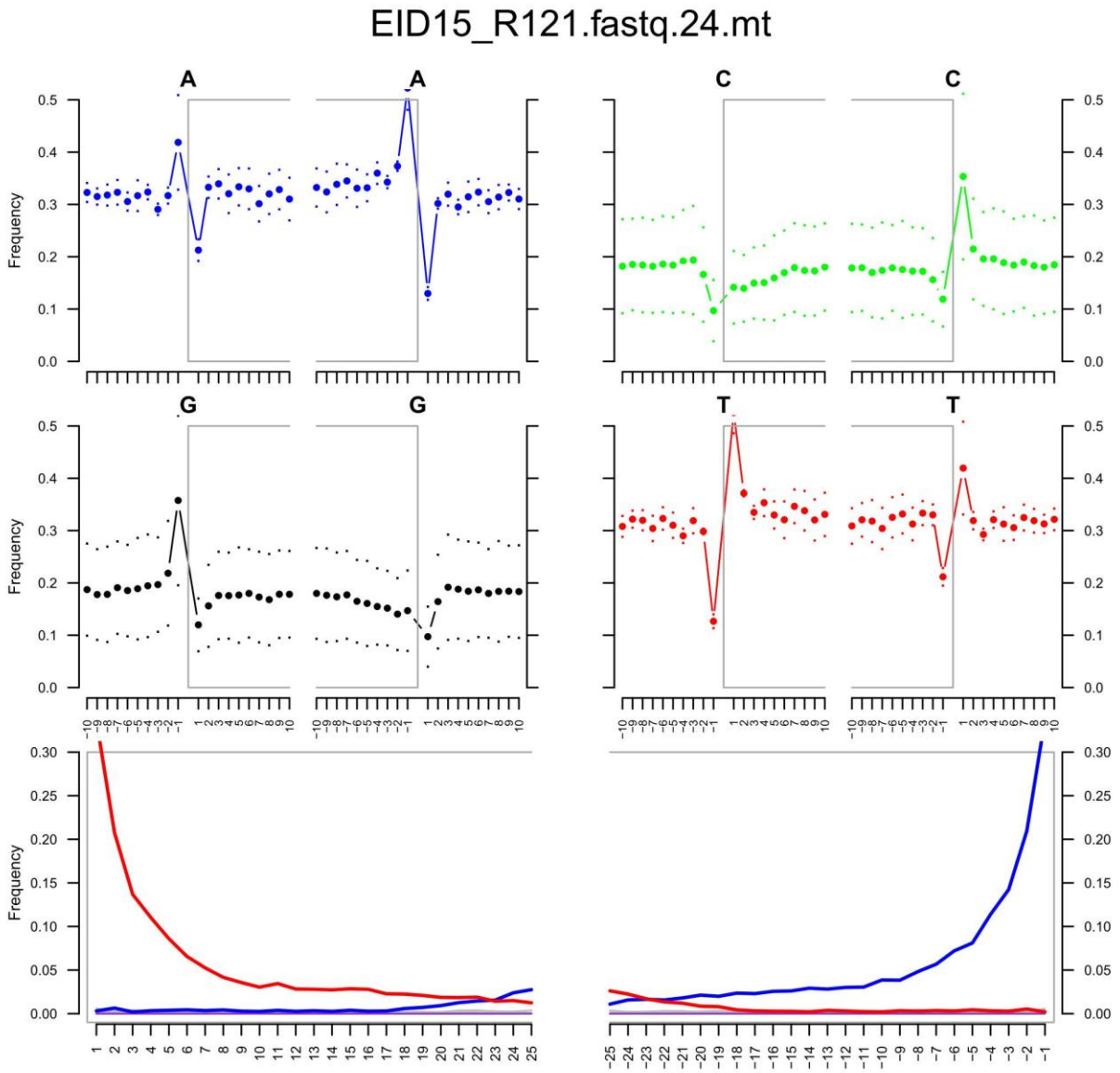
Supplementary Figure 1a. Mammoth 1 (YPC 3.0229 Yukon Permafrost) fragment misincorporation plot for enriched non-collapsed mapped reads. Generated with mapDamage version 2.0 (39).



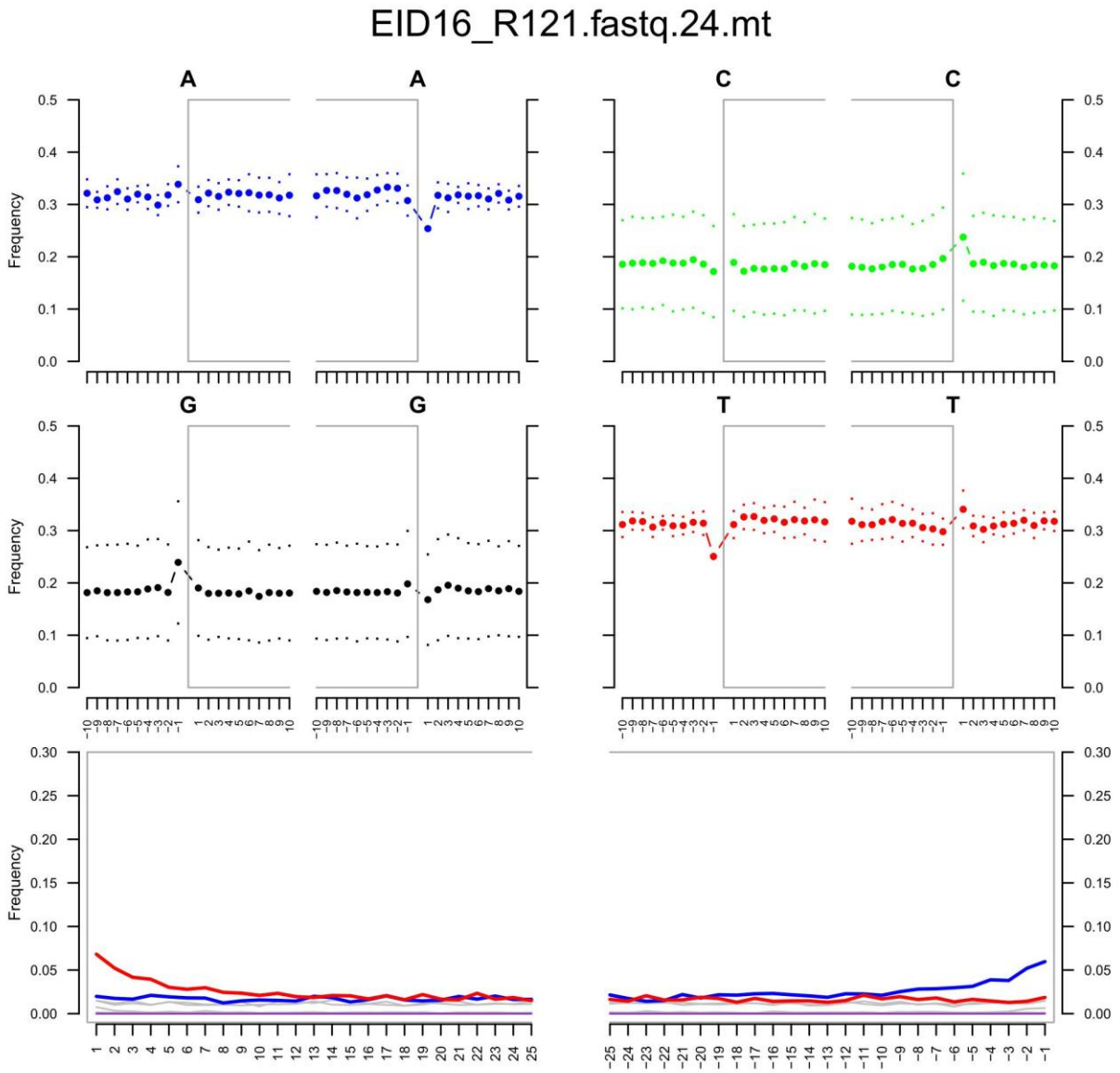
Supplementary Figure 1b. Mammoth 2 (YPC5.0046 Yukon Permafrost) fragment misincorporation plot for enriched non-collapsed mapped reads. Generated with mapDamage version 2.0 (39).



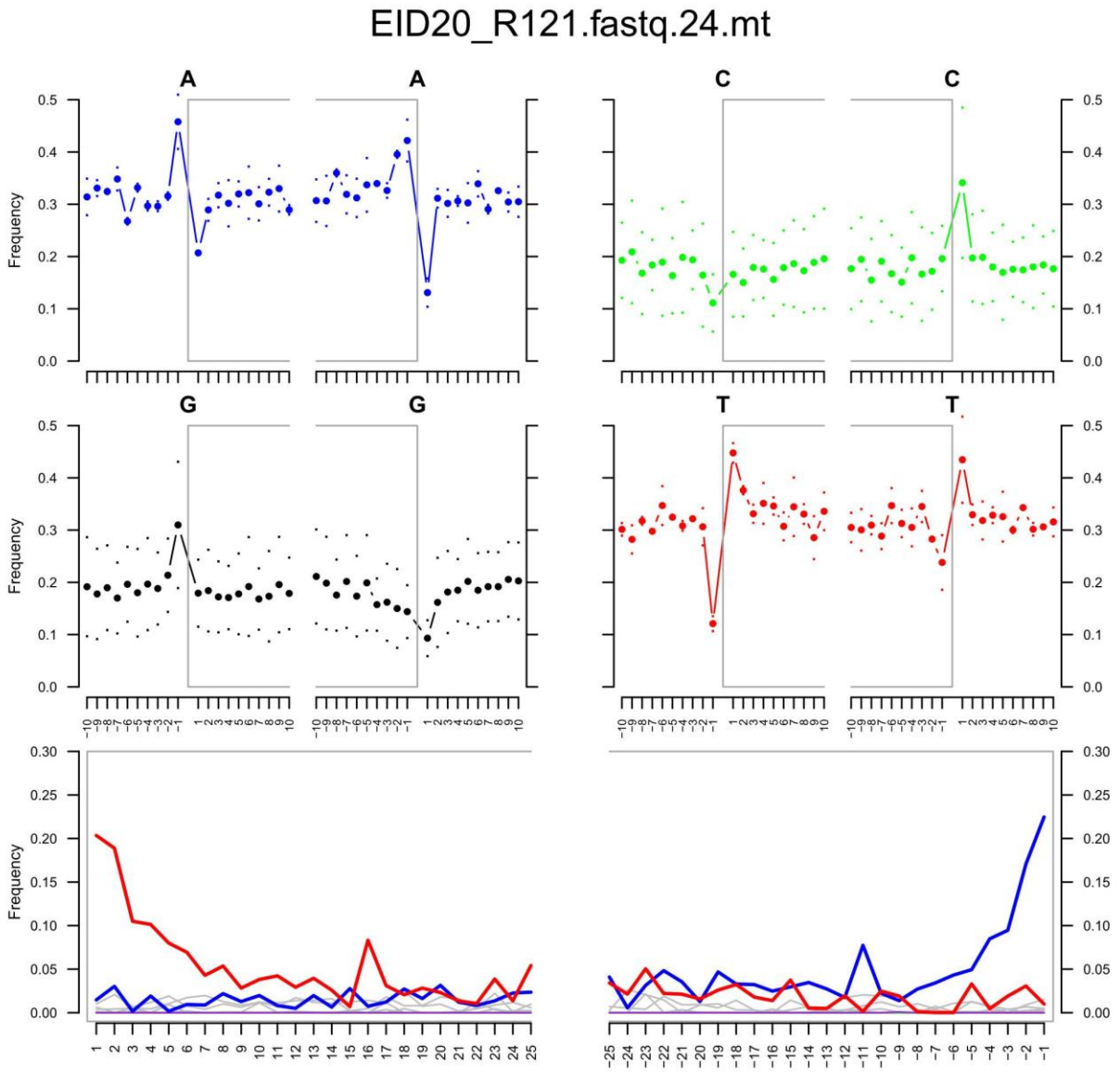
Supplementary Figure 1c. Mammoth 3 (IK-99-524 Alaska Permafrost) fragment misincorporation plot for enriched non-collapsed mapped reads. Generated with mapDamage version 2.0 (39).



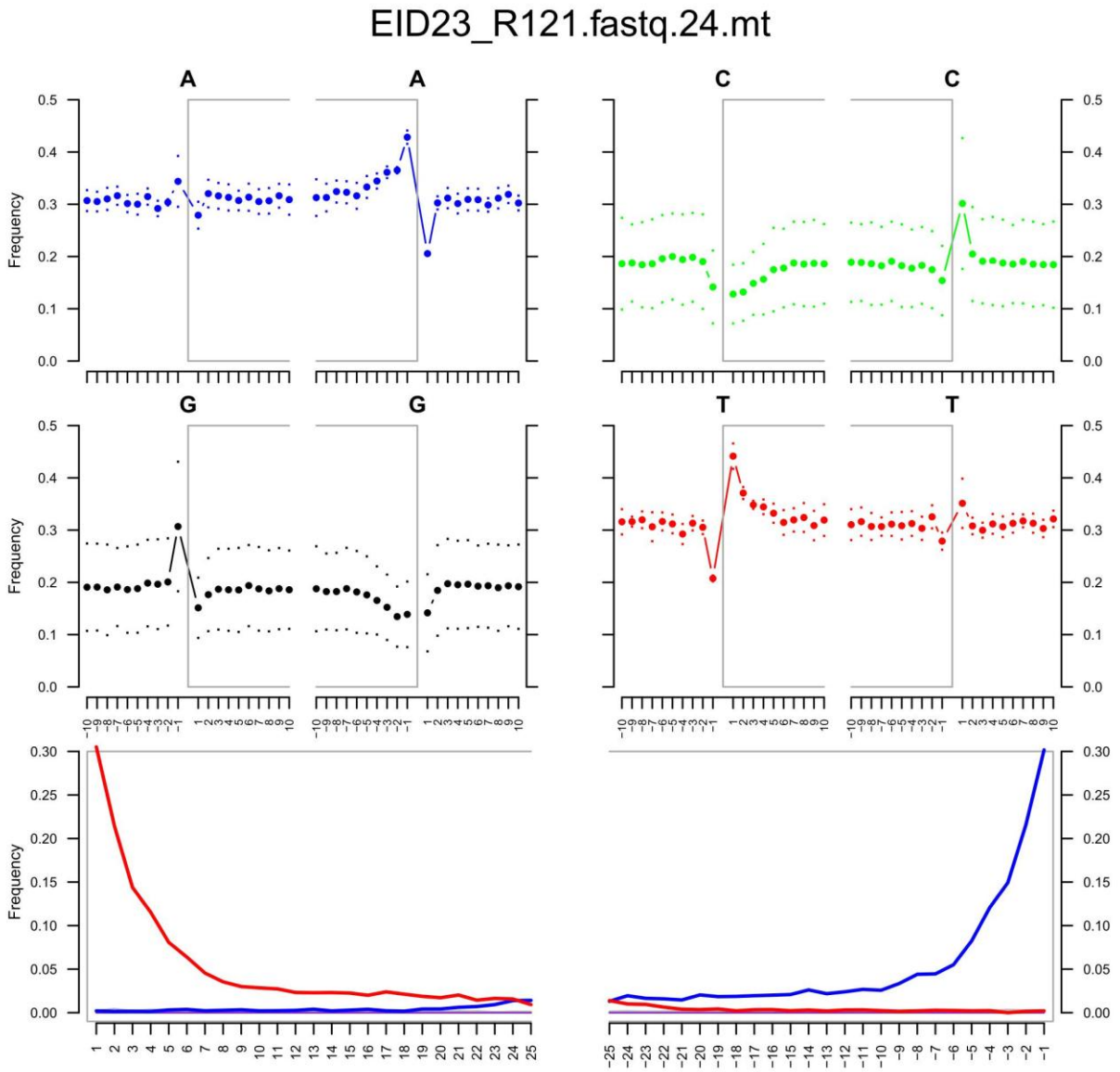
Supplementary Figure 1d. Mammoth 4 (UW20579 Wyoming Temperate) fragment misincorporation plot for enriched non-collapsed mapped reads. Generated with mapDamage version 2.0 (39).



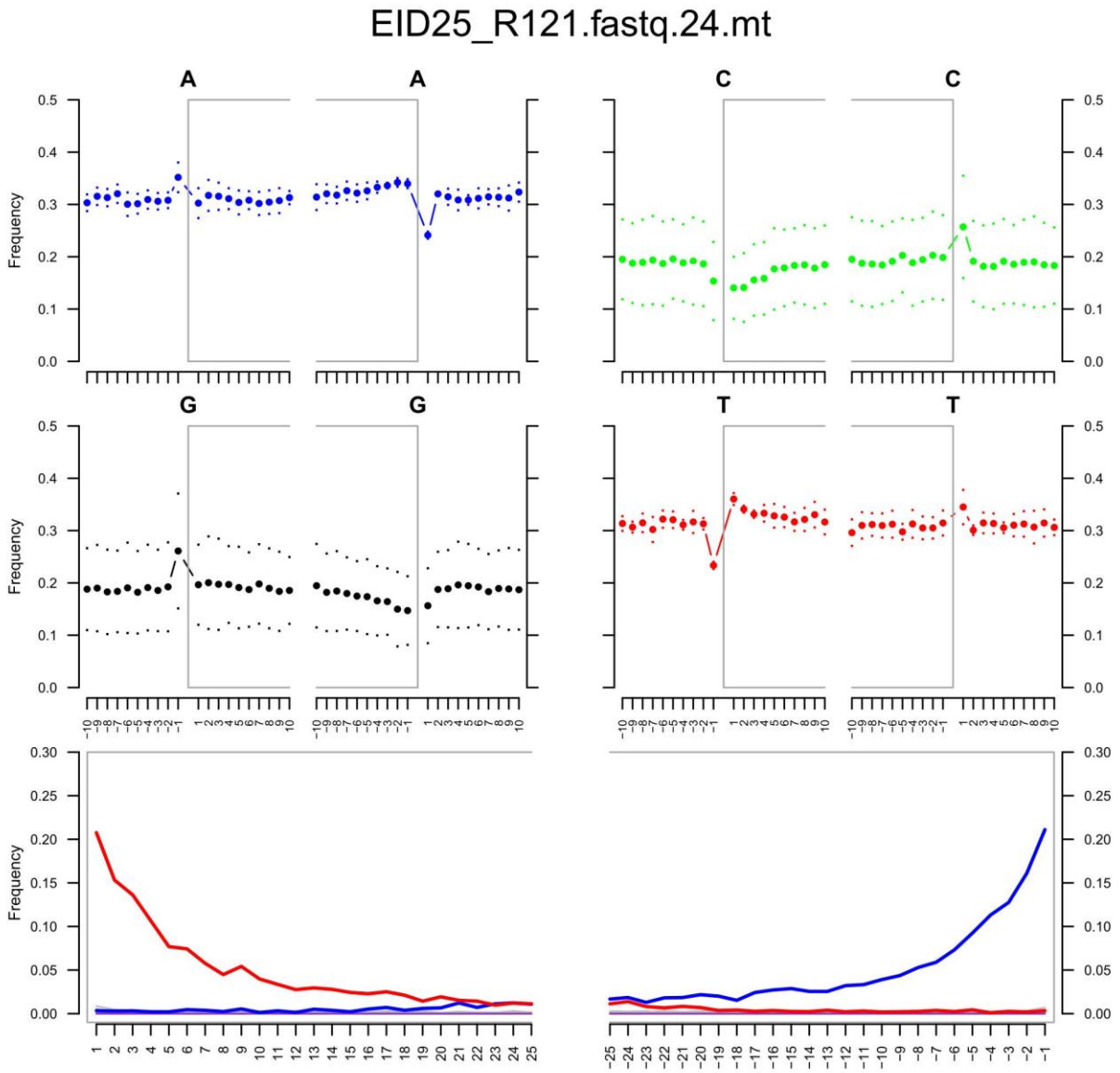
Supplementary Figure 1e. Mammoth 5 (Lyuba Mummy Siberia Permafrost) fragment misincorporation plot for enriched non-collapsed mapped reads. Generated with mapDamage version 2.0 (39).



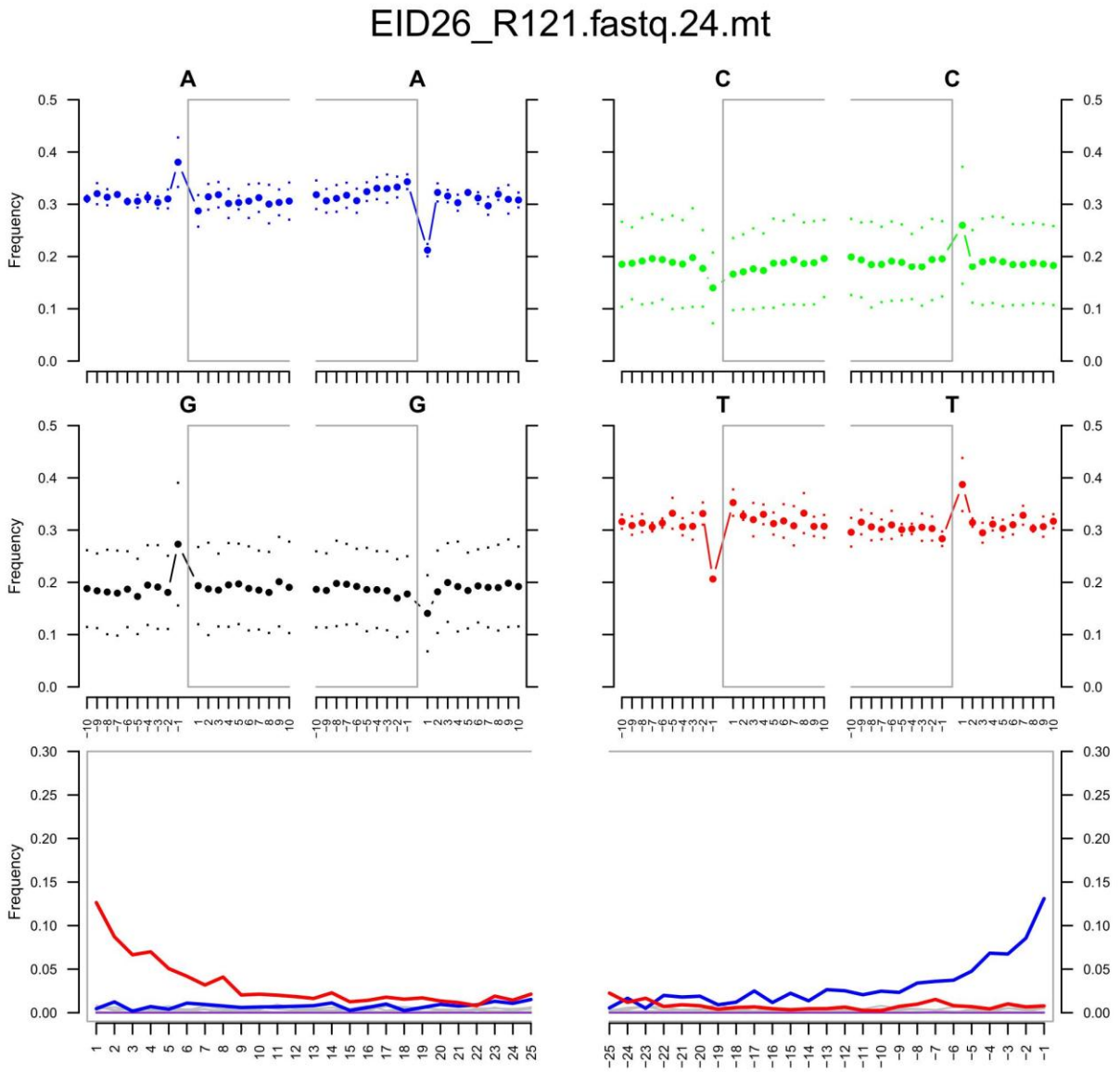
Supplementary Figure 1f. Mammoth 6 (YP180.40 Yukon Permafrost) fragment misincorporation plot for enriched non-collapsed mapped reads. Generated with mapDamage version 2.0 (39).



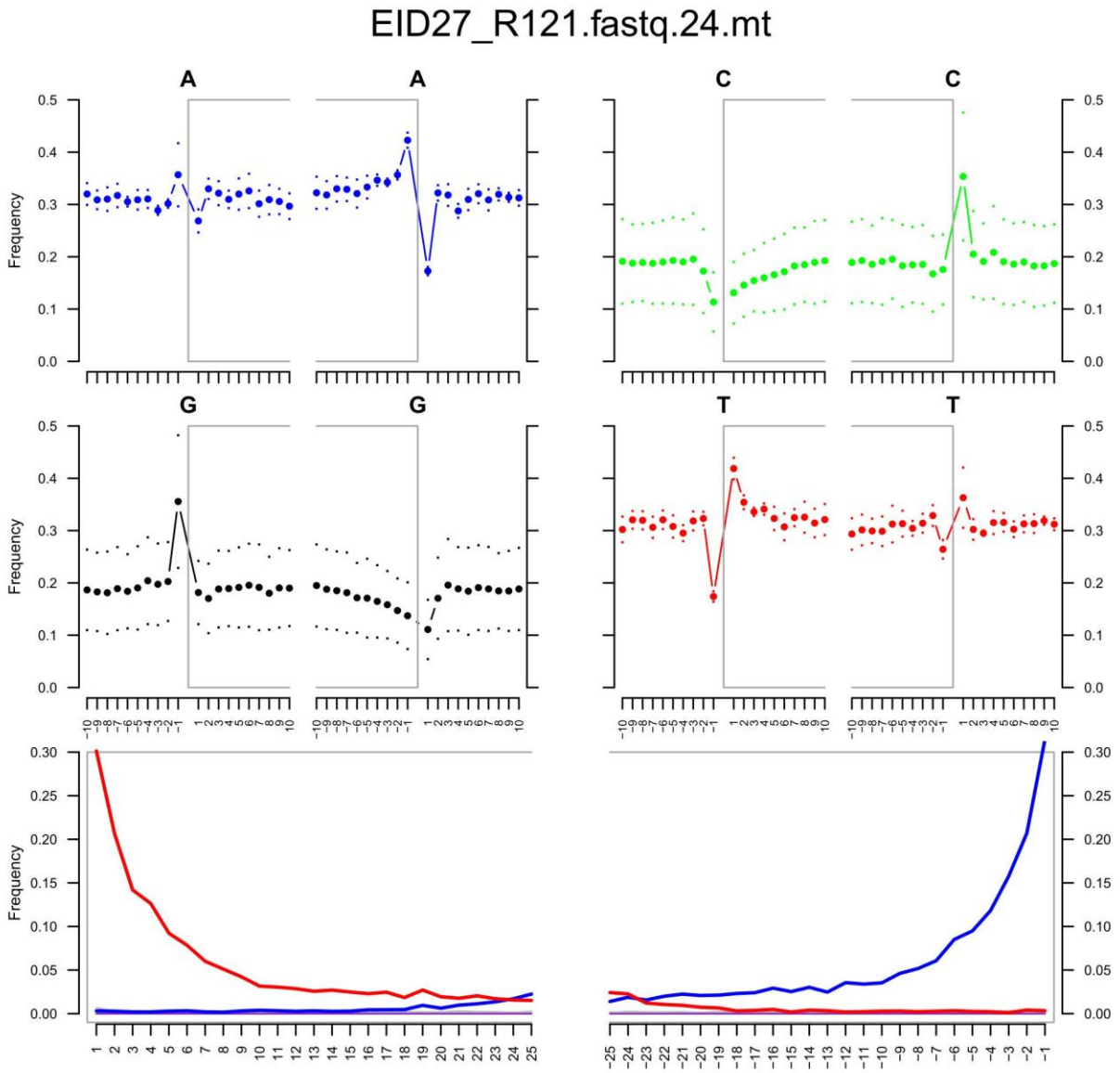
Supplementary Figure 1g. Mammoth 7 (Poyser Indiana Temperate) fragment misincorporation plot for enriched non-collapsed mapped reads. Generated with mapDamage version 2.0 (39).



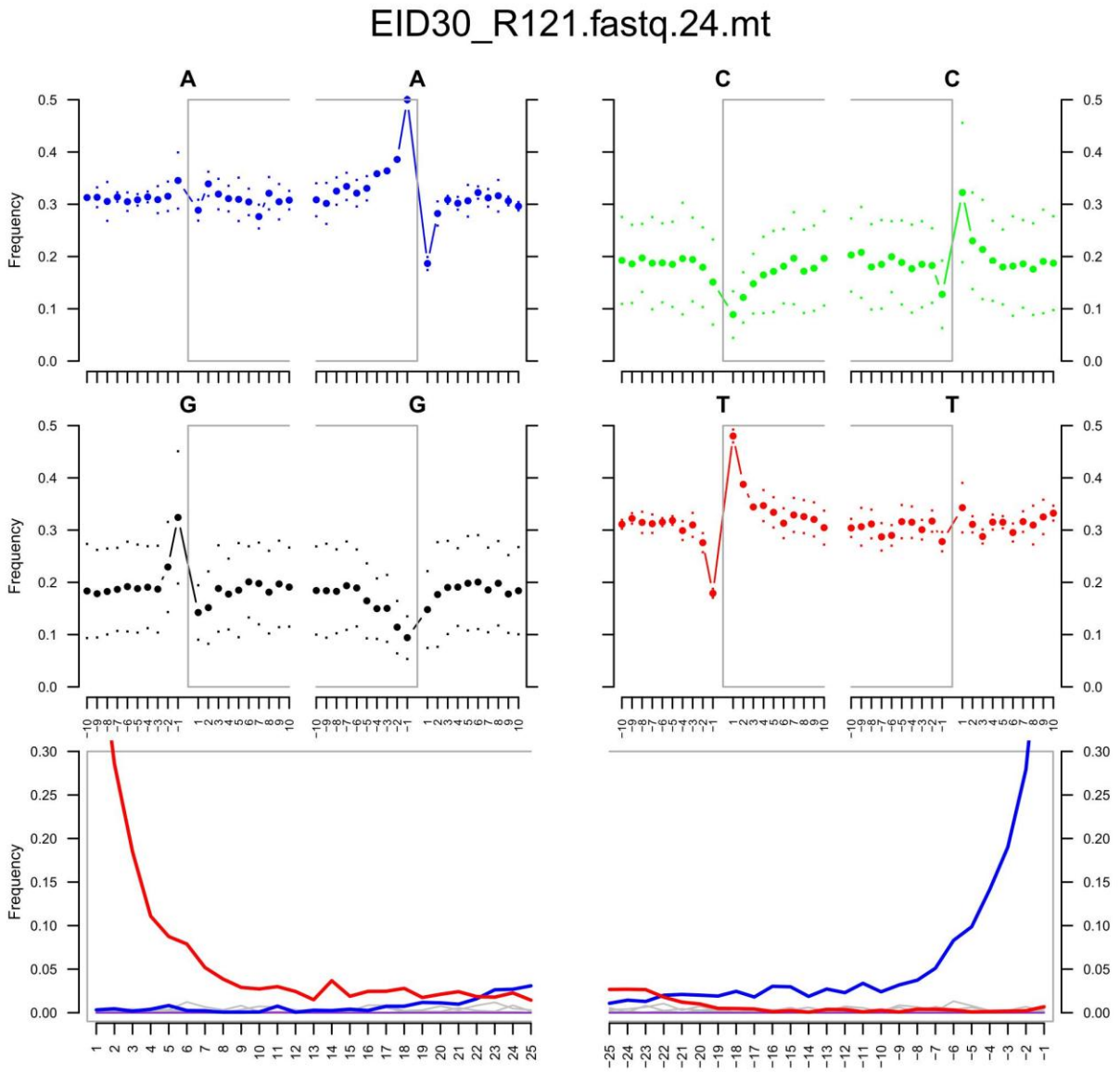
Supplementary Figure 1h. Mammoth 8 (Scarborough Maine Temperate) fragment misincorporation plot for enriched non-collapsed mapped reads. Generated with mapDamage version 2.0 (39).



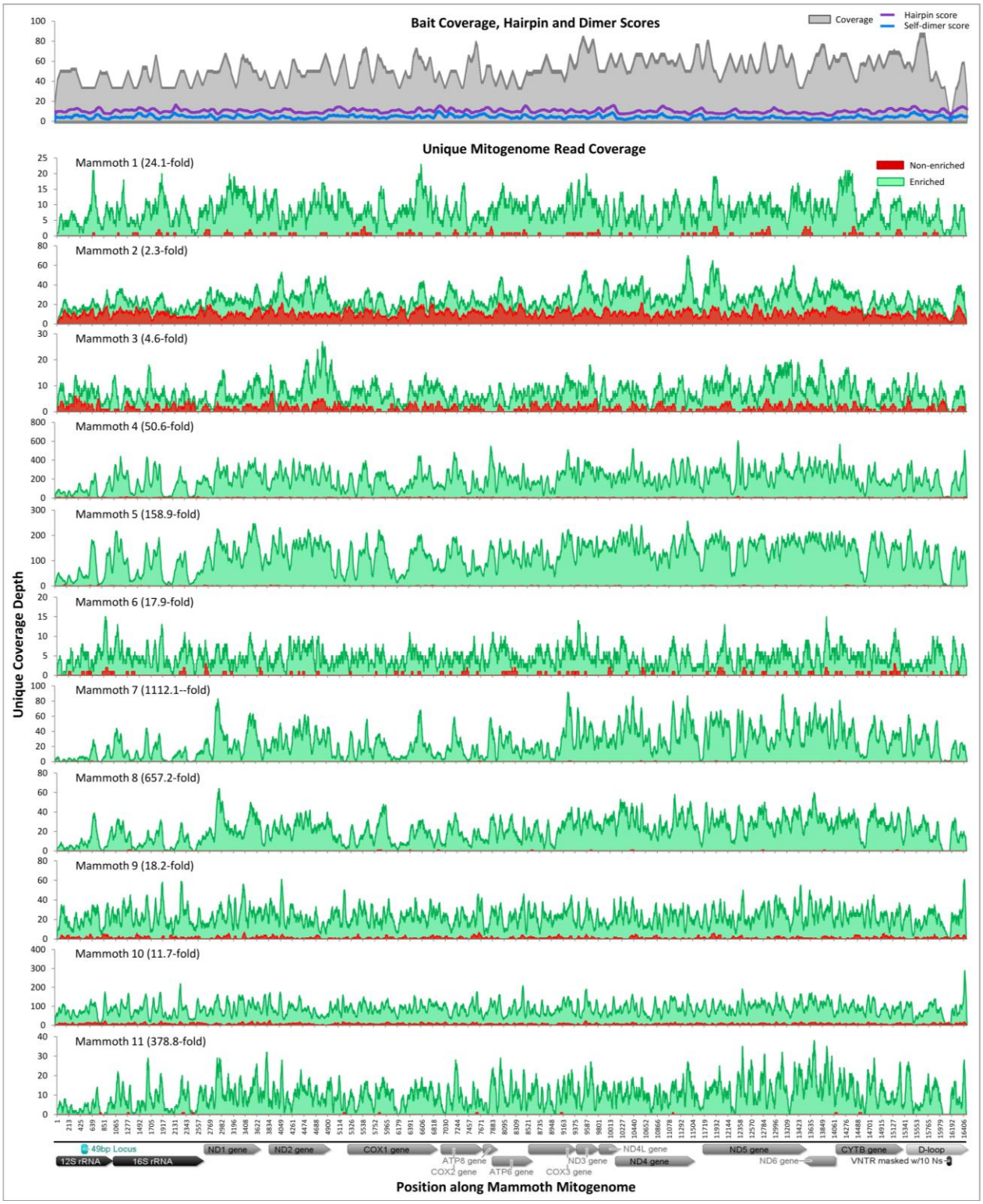
Supplementary Figure 1i. Mammoth 9 (Rawlins Wyoming Temperate) fragment misincorporation plot for enriched non-collapsed mapped reads. Generated with mapDamage version 2.0 (39).



Supplementary Figure 1j. Mammoth 10 (Randolph New York Temperate) fragment misincorporation plot for enriched non-collapsed mapped reads. Generated with mapDamage version 2.0 (39).



Supplementary Figure 1k. Mammoth 11 (Bindloss Alberta Temperate) fragment misincorporation plot for enriched non-collapsed mapped reads. Generated with mapDamage version 2.0 (39).



Supplementary Figure 2: Bait coverage depth, bait secondary structure scores, and unique read coverage depth across the mitogenome. Unique read enrichment rates are indicated in parentheses. (*caption in publication is erroneous, corrigendum in progress*).

2.11 SUPPLEMENTARY TABLES

Supplementary Table 1: Sample and qPCR data

SPECIMEN					EXTRACT				LIBRARY
Sample	**Name**	**^{14}C Age**	**Geog. Origin**	**Deposit. Context**	**49bp, 1X Q1a**	**49bp, 0.1X Q1b**	**% Inhibited**	**Concentration for Library**	**Concentration for Indexing**
Mammoth 1	YPC3.0229	nd	Yukon	Permafrost	246	899	72.7%	1.00	0.10
Mammoth 2	YPC5.0046	22,430	Yukon	Permafrost	2,333	3,569	34.6%	1.00	1.00
Mammoth 3	IK-99-524	>51,000	Alaska	Permafrost	478	691	30.7%	0.45	0.10
Mammoth 4	UW20579	nd	Wyoming	Temperate	1,550	2,238	30.7%	0.82	1.00
Mammoth 5	Lyuba-c	41,910	Siberia	Permafrost	199	23,081	99.1%	0.10	0.40
Mammoth 6	YP180.40	nd	Yukon	Permafrost	21	32	35.6%	1.00	0.10
Mammoth 7	Poyser	22,000	Indiana	Temperate	27	241	88.9%	1.00	1.00
Mammoth 8	Scarborough	12,180	Maine	Temperate	87	360	75.8%	0.73	0.40
Mammoth 9	Rawlins	11,280	Wyoming	Temperate	135	2,504	94.6%	0.10	0.10
Mammoth 10	Randolph	10,350	New York	Temperate	323	2,721	88.1%	0.10	0.40
Mammoth 11	Bindloss 2	10,930	Alberta	Temperate	6	13	54.0%	0.73	0.40
Extr. Blank	E22 Blank	-	-	-	<1	0	-	1.00	0.10
Mylodon	MC01	-	Chile	Cave	0	0	-	1.00	0.40

Measures reported for 0.1X extracts are corrected for dilution factor and thus represent projected copies in the straight

Green italicized numbers indicate post-enrichment measures that were within the range of the enrichment blanks (data not shown), or rates derived from such numbers

Supplementary Table 1 (Continued): Sample and qPCR data

INDEXED LIBRARY			ENRICHED LIBRARY, 2.5ng BAITS, REP 1				ENRICHED LIBRARY, 2.5ng BAITS, REP 2				2.5ng Avg
Total Q2	12S Q2	49:Total Q2	Total Q3a	12S Q3a	12S:Total Q3a	Enr. Rate Q3a	Total Q3b	12S Q3b	12S:Total Q3b	Enr. Rate Q3b	Enr. Rate Q3
6.7E+09	10,423	1.56E-06	2.6E+09	906,800	3.5E-04	227	2.3E+09	714,100	3.1E-04	201	214
1.9E+10	199,500	1.07E-05	1.4E+10	6,211,000	4.4E-04	41	1.8E+10	1,993,000	1.1E-04	11	26
1.3E+10	37,945	2.82E-06	3.2E+09	1,797,000	5.7E-04	200	1.1E+10	2,047,500	1.9E-04	69	135
6.6E+09	5,102	7.76E-07	6.2E+09	104,965	1.7E-05	22	4.3E+09	138,900	3.3E-05	42	32
6.5E+09	5,105	7.84E-07	1.2E+10	395,950	3.4E-05	44	5.7E+09	288,600	5.0E-05	64	54
7.5E+09	161	2.16E-08	3.5E+09	15,735	4.5E-06	210	3.4E+09	4,073	1.2E-06	55	133
3.4E+09	44	1.27E-08	7.0E+08	19,760	2.8E-05	2217	6.6E+08	*4*	5.3E-09	*0*	*1109*
8.7E+09	107	1.23E-08	3.3E+09	19,095	5.7E-06	463	5.3E+09	8,096	1.5E-06	124	294
7.4E+09	123	1.67E-08	5.5E+09	5,742	1.0E-06	62	5.2E+09	6,797	1.3E-06	78	70
1.3E+10	16,505	1.23E-06	4.9E+09	340,200	7.0E-05	57	1.2E+10	319,500	2.6E-05	21	39
2.2E+09	19	8.69E-09	5.1E+08	*9*	1.8E-08	*2*	5.6E+08	*26*	4.7E-08	*5*	*4*
3.0E+09	0	0	8.0E+08	*3*	3.9E-09	-	5.4E+08	*8*	1.5E-08	-	-
2.7E+10	0	0	1.4E+10	*5*	3.4E-10	-	1.1E+10	*18*	1.7E-09	-	-

Measures reported for 0.1X extracts are corrected for dilution factor and thus represent projected copies in the straight

Green italicized numbers indicate post-enrichment measures that were within the range of the enrichment blanks (data not shown), or rates derived from such numbers

Supplementary Table 1 (Continued): Sample and qPCR data

ENRICHED LIBRARY, 25ng BAITS, REP1				ENRICHED LIBRARY, 25ng BAITS, REP2				25ng Avg.
Total Q4a	12S Q4a	12S:Total Q4a	Enr. Rate Q4a	Total Q4b	12S Q4b	12S:Total Q4b	Enr. Rate Q4b	Enr. Rate Q4
7.9E+09	487,150	6.1E-05	79	1.0E+10	246,850	2.4E-05	31	55
4.7E+09	43,935	9.4E-06	435	5.0E+09	30,810	6.2E-06	287	361
9.3E+08	21,455	2.3E-05	1820	8.3E+08	8,502	1.0E-05	803	1311
6.6E+09	36,110	5.5E-06	446	1.3E+10	89,045	7.0E-06	568	507
8.1E+09	15,805	1.9E-06	116	8.1E+09	7,487	9.2E-07	55	86
1.2E+10	869,000	7.0E-05	57	8.8E+09	558,250	6.3E-05	52	54
3.5E+09	48,730	1.4E-05	1601	2.4E+09	24,300	1.0E-05	1155	1378
1.9E+09	*116*	6.1E-08	-	3.4E+09	*194*	5.7E-08	-	-
1.2E+10	*149*	1.2E-08	-	1.7E+10	*75*	4.4E-09	-	-

Supplementary Table 2: Sequence data, mitogenome alignment rates, and enrichment rates

SAMPLE	NON-ENRICHED						
	ALL READS SEQUENCED			3 MILLION READ SUBSET			
	READ 1 n	INSERTS < 10bp		MITOGENOME READS		UNIQUE MITO. READS	
Mammoth 1	15,977,223	105,634	0.66%	83	0.003%	80	0.0027%
Mammoth 2	7,759,694	861,257	11.10%	4,506	0.150%	2,423	0.0808%
Mammoth 3	9,242,230	247,693	2.68%	487	0.016%	399	0.0133%
Mammoth 4	17,219,453	35,794	0.21%	1,654	0.055%	1,635	0.0545%
Mammoth 5	14,544,531	21,790	0.15%	247	0.008%	247	0.0082%
Mammoth 6	21,481,367	163,991	0.76%	86	0.003%	84	0.0028%
Mammoth 7	33,321,984	7,357	0.02%	8	0.000%	8	0.0003%
Mammoth 8	23,765,027	19,289	0.08%	10	0.000%	10	0.0003%
Mammoth 9	34,689,671	263,459	0.76%	463	0.015%	452	0.0151%
Mammoth 10	19,712,228	639,219	3.24%	2,850	0.095%	2,712	0.0904%
Mammoth 11	26,304,817	64,220	0.24%	11	0.000%	11	0.0004%
Extr. Blank	12,383,802	8,516,406	68.77%	1	0.000%	1	0.0000%
Mylodon	31,561,857	20,001	0.06%	10	0.000%	10	0.0003%

Supplementary Table 2 (Continued): Sequence data, mitogenome alignment rates, and enrichment rates

ENRICHED							ENRICHMENT RATE (FOLD)	
ALL READS SEQUENCED			3 MILLION READ SUBSET					
READ 1 n	INSERTS < 10bp		MITOGENOME READS		UNIQUE MITO. READS		MITOGENOME	UNIQUE MITO.
3,308,514	18,500	0.56%	221,776	7.393%	1,928	0.0643%	2672.0	24.10
6,765,718	403,489	5.96%	1,379,489	45.983%	5,517	0.1839%	306.1	2.28
4,204,305	69,158	1.64%	587,951	19.598%	1,819	0.0606%	1207.3	4.56
22,798,422	26,057	0.11%	782,089	26.070%	82,650	2.7550%	472.8	50.55
5,243,990	4,054	0.08%	498,381	16.613%	39,493	1.3164%	2017.7	159.89
16,430,456	56,756	0.35%	82,523	2.751%	1,502	0.0501%	959.6	17.88
16,844,725	3,266	0.02%	11,231	0.374%	8,897	0.2966%	1403.9	1112.13
44,082,587	14,840	0.03%	10,542	0.351%	6,572	0.2191%	1054.2	657.20
40,150,292	142,371	0.35%	158,052	5.268%	8,246	0.2749%	341.4	18.24
10,174,572	160,461	1.58%	499,317	16.644%	31,676	1.0559%	175.2	11.68
17,542,164	19,487	0.11%	19,510	0.650%	4,167	0.1389%	1773.6	378.82
11,302,699	2,205,958	19.52%	104	0.00%	102	0.00%	104.0	102.00
25,373,857	12,671	0.05%	732	0.02%	276	0.01%	73.2	27.60

Supplementary Table 3: Linear correlation matrices between coverage, target and bait properties, analyzed 3 million reads each sample. GC content analyzed at a 41bp sliding window. Missing values have too little variation for correlation coefficient calculation. Correlation cells are conditionally formatted against correlations from all samples.

MAMMOTH 1								YPC3.0229
	NON-ENRICHED			**ENRICHED**				
BASES COVERED (OF 16450 + 10 Ns)	4733			16402				
Non-Enriched Unique Coverage	0.93							
Non-Enriched Duplication Rate	0.31	-0.04						
Enriched Raw Coverage	0.21	0.05	0.48					
Enriched Unique Coverage	0.18	0.16	0.08	0.59				
Enriched Duplication Rate	0.15	-0.02	0.49	0.84	0.21			
Raw Enrichment Rate	-0.15	-0.23	0.19	0.88	0.55	0.81		
Unique Enrichment Rate	-0.39	-0.45	0.12	0.48	0.79	0.23	0.64	
Reference GC Content	-0.07	-0.02	-0.15	-0.20	-0.01	-0.26	-0.21	0.00
Bait Coverage	-0.01	-0.01	0.02	0.35	0.18	0.39	0.45	0.13
Average Bait Hairpin Score	-0.14	-0.06	-0.23	-0.36	-0.16	-0.35	-0.35	-0.12
Average Bait Dimer Score	-0.20	-0.12	-0.24	-0.36	-0.19	-0.34	-0.32	-0.16
AVERAGE:	1.3	1.2	1.0	1052.9	8.4	116.0	1217.7	8.9
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

MAMMOTH 2								YPC5.0046
	NON-ENRICHED			ENRICHED				
BASES COVERED (OF 16450 + 10 Ns)	16441			16452				
Non-Enriched Unique Coverage	0.90							
Non-Enriched Duplication Rate	0.41	0.01						
Enriched Raw Coverage	0.28	0.28	0.06					
Enriched Unique Coverage	0.56	0.61	0.02	0.81				
Enriched Duplication Rate	0.11	0.08	0.07	0.91	0.60			
Raw Enrichment Rate	-0.15	-0.10	-0.16	0.83	0.55	0.87		
Unique Enrichment Rate	-0.32	-0.36	-0.02	0.56	0.45	0.58	0.76	
Reference GC Content	0.11	0.14	-0.03	-0.22	-0.06	-0.27	-0.25	-0.20
Bait Coverage	0.05	0.04	-0.01	0.39	0.28	0.43	0.39	0.28
Average Bait Hairpin Score	-0.07	-0.08	-0.02	-0.39	-0.30	-0.35	-0.35	-0.26
Average Bait Dimer Score	-0.05	-0.09	0.05	-0.37	-0.32	-0.33	-0.33	-0.25
AVERAGE:	19.7	10.6	1.9	6729.4	25.3	232.5	361.8	2.5
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

Supplementary Table 3 (Continued): Linear correlation matrices between coverage, target and bait properties, analyzed 3 million reads each sample.

MAMMOTH 3								IK-99-524
	NON-ENRICHED			ENRICHED				
BASES COVERED (OF 16450 + 10 Ns)	12841			16322				
Non-Enriched Unique Coverage	0.89							
Non-Enriched Duplication Rate	0.41	0.01						
Enriched Raw Coverage	0.20	0.17	0.08					
Enriched Unique Coverage	0.39	0.43	0.03	0.73				
Enriched Duplication Rate	0.06	0.02	0.07	0.80	0.35			
Raw Enrichment Rate	-0.29	-0.27	-0.14	0.73	0.45	0.71		
Unique Enrichment Rate	-0.45	-0.50	0.00	0.40	0.46	0.29	0.71	
Reference GC Content	0.02	0.00	0.05	-0.22	-0.10	-0.26	-0.22	-0.11
Bait Coverage	-0.05	-0.06	0.00	0.34	0.22	0.35	0.32	0.19
Average Bait Hairpin Score	-0.09	-0.06	-0.07	-0.35	-0.23	-0.30	-0.26	-0.15
Average Bait Dimer Score	-0.10	-0.07	-0.07	-0.31	-0.20	-0.30	-0.21	-0.07
AVERAGE:	2.5	2.0	1.2	2595.9	7.8	303.0	1414.0	4.9
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

MAMMOTH 4								UW20579
	NON-ENRICHED			ENRICHED				
BASES COVERED (OF 16450 + 10 Ns)	15798			16452				
Non-Enriched Unique Coverage	0.995							
Non-Enriched Duplication Rate	0.10	0.01						
Enriched Raw Coverage	0.01	0.01	-0.01					
Enriched Unique Coverage	0.24	0.24	-0.02	0.71				
Enriched Duplication Rate	-0.14	-0.14	0.00	0.92	0.51			
Raw Enrichment Rate	-0.39	-0.38	-0.07	0.71	0.39	0.75		
Unique Enrichment Rate	-0.57	-0.57	-0.06	0.43	0.41	0.43	0.80	
Reference GC Content	-0.03	-0.03	-0.01	-0.30	-0.28	-0.31	-0.22	-0.17
Bait Coverage	-0.04	-0.04	-0.03	0.36	0.32	0.38	0.25	0.20
Average Bait Hairpin Score	0.04	0.04	0.04	-0.39	-0.22	-0.38	-0.32	-0.18
Average Bait Dimer Score	0.01	0.01	0.06	-0.39	-0.24	-0.37	-0.30	-0.17
AVERAGE:	4.1	4.0	1.0	2100.9	206.2	8.3	778.1	71.6
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

Supplementary Table 3 (Continued): Linear correlation matrices between coverage, target and bait properties, analyzed 3 million reads each sample.

MAMMOTH 5 — LYUBA

	NON-ENRICHED			ENRICHED				
BASES COVERED (OF 16450 + 10 Ns)	8085			16448				
Non-Enriched Unique Coverage	1.00							
Non-Enriched Duplication Rate								
Enriched Raw Coverage	0.02	0.02						
Enriched Unique Coverage	0.01	0.01		0.76				
Enriched Duplication Rate	0.00	0.00		0.97	0.66			
Raw Enrichment Rate	-0.26	-0.26		0.91	0.68	0.89		
Unique Enrichment Rate	-0.49	-0.49		0.61	0.80	0.54	0.77	
Reference GC Content	0.13	0.13		-0.25	-0.13	-0.31	-0.29	-0.20
Bait Coverage	0.06	0.06		0.38	0.37	0.39	0.36	0.25
Average Bait Hairpin Score	-0.04	-0.04		-0.44	-0.36	-0.38	-0.35	-0.25
Average Bait Dimer Score	-0.10	-0.10		-0.42	-0.35	-0.37	-0.30	-0.20
AVERAGE:	1.4	1.4	1.0	1677.3	123.0	11.1	1387.6	102.1
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

MAMMOTH 6 — YP180.40

	NON-ENRICHED			ENRICHED				
BASES COVERED (OF 16450 + 10 Ns)	3579			16178				
Non-Enriched Unique Coverage	0.94							
Non-Enriched Duplication Rate	0.29	-0.04						
Enriched Raw Coverage	0.15	0.14	0.03					
Enriched Unique Coverage	0.23	0.23	0.04	0.63				
Enriched Duplication Rate	0.01	0.01	-0.01	0.66	0.08			
Raw Enrichment Rate	-0.17	-0.16	-0.06	0.93	0.51	0.70		
Unique Enrichment Rate	-0.25	-0.29	0.07	0.51	0.85	0.07	0.61	
Reference GC Content	0.03	0.00	0.09	-0.20	-0.05	-0.21	-0.26	-0.15
Bait Coverage	0.08	0.04	0.11	0.24	0.02	0.33	0.29	0.10
Average Bait Hairpin Score	-0.09	-0.05	-0.13	-0.36	-0.18	-0.30	-0.27	-0.13
Average Bait Dimer Score	-0.08	-0.05	-0.12	-0.34	-0.23	-0.20	-0.22	-0.16
AVERAGE:	1.1	1.1	1.0	273.8	4.7	58.2	338.4	5.4
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

Supplementary Table 3 (Continued): Linear correlation matrices between coverage, target and bait properties, analyzed 3 million reads each sample.

MAMMOTH 7								POYSER
	NON-ENRICHED			ENRICHED				
BASES COVERED (OF 16450 + 10 Ns)	336			16024				
Non-Enriched Unique Coverage								
Non-Enriched Duplication Rate								
Enriched Raw Coverage								
Enriched Unique Coverage				0.99				
Enriched Duplication Rate				0.20	0.19			
Raw Enrichment Rate				1.00	1.00	0.76		
Unique Enrichment Rate				1.00	1.00	0.76	1.00	
Reference GC Content				-0.10	-0.11	0.03	-0.11	-0.12
Bait Coverage				0.36	0.36	0.03	0.60	0.61
Average Bait Hairpin Score				-0.38	-0.39	-0.06	0.21	0.20
Average Bait Dimer Score				-0.37	-0.39	-0.06	-0.29	-0.28
AVERAGE:	1.0	1.0	1.0	33.3	26.5	1.2	47.8	37.2
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

MAMMOTH 8								SCARBOROUGH
	NON-ENRICHED			ENRICHED				
BASES COVERED (OF 16450 + 10 Ns)	465			16200				
Non-Enriched Unique Coverage								
Non-Enriched Duplication Rate								
Enriched Raw Coverage								
Enriched Unique Coverage				0.97				
Enriched Duplication Rate				0.64	0.60			
Raw Enrichment Rate				1.00	0.97	0.84		
Unique Enrichment Rate				0.97	1.00	0.75	0.97	
Reference GC Content				-0.07	-0.05	-0.15	0.08	0.12
Bait Coverage				0.38	0.37	0.31	0.67	0.65
Average Bait Hairpin Score				-0.41	-0.39	-0.29	-0.62	-0.56
Average Bait Dimer Score				-0.40	-0.38	-0.28	-0.51	-0.45
AVERAGE:	1.0	1.0	1.0	35.9	22.3	1.5	50.1	29.2
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

Supplementary Table 3 (Continued): Linear correlation matrices between coverage, target and bait properties, analyzed 3 million reads each sample.

MAMMOTH 9								RAWLINS
	NON-ENRICHED			ENRICHED				
BASES COVERED (OF 16450 + 10 Ns)	11366			16453				
Non-Enriched Unique Coverage	0.98							
Non-Enriched Duplication Rate	0.08	-0.10						
Enriched Raw Coverage	0.10	0.08	0.08					
Enriched Unique Coverage	0.25	0.25	-0.03	0.46				
Enriched Duplication Rate	-0.07	-0.08	0.05	0.77	-0.08			
Raw Enrichment Rate	-0.45	-0.44	-0.04	0.73	0.21	0.67		
Unique Enrichment Rate	-0.61	-0.62	0.06	0.23	0.49	0.00	0.60	
Reference GC Content	0.15	0.15	-0.02	-0.02	0.41	-0.33	-0.11	0.18
Bait Coverage	-0.01	0.00	-0.08	0.36	0.08	0.43	0.29	0.05
Average Bait Hairpin Score	-0.09	-0.09	-0.08	-0.37	-0.13	-0.31	-0.29	-0.05
Average Bait Dimer Score	-0.03	-0.01	-0.11	-0.38	-0.18	-0.29	-0.31	-0.13
AVERAGE:	1.9	1.8	1.0	470.9	23.0	21.1	322.2	16.6
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

MAMMOTH 10								RANDOLPH
	NON-ENRICHED			ENRICHED				
BASES COVERED (OF 16450 + 10 Ns)	16223			16449				
Non-Enriched Unique Coverage	0.99							
Non-Enriched Duplication Rate	0.10	-0.04						
Enriched Raw Coverage	0.15	0.14	0.06					
Enriched Unique Coverage	0.57	0.57	0.01	0.54				
Enriched Duplication Rate	-0.24	-0.25	0.07	0.77	0.00			
Raw Enrichment Rate	-0.51	-0.51	-0.06	0.46	-0.06	0.65		
Unique Enrichment Rate	-0.59	-0.60	0.01	0.11	-0.01	0.20	0.78	
Reference GC Content	0.33	0.34	-0.03	-0.02	0.35	-0.31	-0.22	-0.08
Bait Coverage	-0.06	-0.07	0.07	0.37	0.14	0.42	0.27	0.13
Average Bait Hairpin Score	-0.12	-0.10	-0.07	-0.39	-0.16	-0.32	-0.19	-0.01
Average Bait Dimer Score	-0.14	-0.13	-0.05	-0.42	-0.22	-0.30	-0.19	-0.01
AVERAGE:	7.7	7.4	1.0	1476.7	88.2	16.9	252.1	14.8
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

Supplementary Table 3 (Continued): Linear correlation matrices between coverage, target and bait properties, analyzed 3 million reads each sample.

MAMMOTH 11							BINDLOSS
	NON-ENRICHED			ENRICHED			
BASES COVERED (OF 16450 + 10 Ns)	415			15984			
Non-Enriched Unique Coverage							
Non-Enriched Duplication Rate							
Enriched Raw Coverage							
Enriched Unique Coverage				0.86			
Enriched Duplication Rate				0.64	0.34		
Raw Enrichment Rate				1.00	0.63	0.90	
Unique Enrichment Rate				0.63	1.00	0.40	0.71
Reference GC Content				-0.10	0.05	-0.31	-0.42
Bait Coverage				0.28	0.20	0.32	0.36
Average Bait Hairpin Score				-0.35	-0.30	-0.25	-0.31
Average Bait Dimer Score				-0.37	-0.31	-0.27	-0.61
AVERAGE:	1.0	1.0	1.0	52.1	10.7	4.6	29.4

Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***
7.8							

Supplementary Table 4: Linear correlation matrices between coverage, target, and bait properties, using the full read set (3.3m - 44.2m) available for each sample. GC content analyzed at a 41bp sliding window. Missing values have too little variation for correlation coefficient calculation. Correlation cells are conditionally formatted against data from all samples. Enrichment rates are calculated in proportional, rather than absolute, increases.

MAMMOTH 1							YPC3.0229
	NON-ENRICHED			**ENRICHED**			
BASES COVERED (OF 16450 + 10 Ns)	12605			16402			
Non-Enriched Unique Coverage	0.91						
Non-Enriched Duplication Rate	0.28	-0.09					
Enriched Raw Coverage	0.32	0.29	0.07				
Enriched Unique Coverage	0.51	0.52	0.00	0.61			
Enriched Duplication Rate	0.09	0.06	0.05	0.84	0.23		
Raw Enrichment Rate	-0.29	-0.26	-0.15	0.66	0.24	0.74	
Unique Enrichment Rate	-0.50	-0.57	0.10	0.18	0.27	0.12	0.57
Reference GC Content	0.05	0.05	-0.03	-0.20	-0.01	-0.26	-0.18
Bait Coverage	0.04	0.05	-0.02	0.36	0.20	0.39	0.32
Average Bait Hairpin Score	-0.11	-0.13	0.07	-0.36	-0.18	-0.34	-0.33
Average Bait Dimer Score	-0.15	-0.16	0.07	-0.36	-0.21	-0.34	-0.30
AVERAGE:	2.6	2.2	1.2	1156.5	8.5	124.6	2901.2

Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***
24.4							

Supplementary Table 4 (Continued): Linear correlation matrices between coverage, target, and bait properties, using the full read set (3.3m - 44.2m) available for each sample.

MAMMOTH 2								YPC5.0046
	NON-ENRICHED			**ENRICHED**				
BASES COVERED (OF 16450 + 10 Ns)	16443			16452				
Non-Enriched Unique Coverage	0.89							
Non-Enriched Duplication Rate	0.46	0.03						
Enriched Raw Coverage	0.33	0.29	0.13					
Enriched Unique Coverage	0.55	0.57	0.08	0.87				
Enriched Duplication Rate	0.16	0.08	0.15	0.89	0.64			
Raw Enrichment Rate	-0.05	-0.04	-0.07	0.87	0.67	0.89		
Unique Enrichment Rate	-0.11	-0.17	0.04	0.72	0.67	0.68	0.85	
Reference GC Content	0.11	0.15	-0.05	-0.22	-0.08	-0.28	-0.26	-0.22
Bait Coverage	0.05	0.05	-0.02	0.39	0.30	0.46	0.42	0.34
Average Bait Hairpin Score	-0.09	-0.05	-0.08	-0.39	-0.34	-0.32	-0.37	-0.38
Average Bait Dimer Score	-0.08	-0.08	-0.01	-0.37	-0.35	-0.31	-0.34	-0.34
AVERAGE:	45.5	14.1	3.2	14367.9	31.5	385.5	368.4	2.6
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

MAMMOTH 3								IK-99-524
	NON-ENRICHED			**ENRICHED**				
BASES COVERED (OF 16450 + 10 Ns)	15650			16357				
Non-Enriched Unique Coverage	0.87							
Non-Enriched Duplication Rate	0.41	-0.01						
Enriched Raw Coverage	0.25	0.30	0.02					
Enriched Unique Coverage	0.54	0.61	0.06	0.75				
Enriched Duplication Rate	-0.02	0.02	-0.03	0.79	0.35			
Raw Enrichment Rate	-0.30	-0.22	-0.21	0.64	0.32	0.75		
Unique Enrichment Rate	-0.37	-0.45	0.13	0.27	0.26	0.24	0.60	
Reference GC Content	0.03	0.02	0.01	-0.22	-0.11	-0.26	-0.21	-0.10
Bait Coverage	-0.12	-0.07	-0.09	0.34	0.24	0.36	0.37	0.23
Average Bait Hairpin Score	-0.07	-0.02	-0.11	-0.35	-0.23	-0.30	-0.25	-0.18
Average Bait Dimer Score	-0.03	0.01	-0.09	-0.31	-0.19	-0.29	-0.23	-0.17
AVERAGE:	6.3	3.6	1.8	3576.2	8.3	376.8	1594.5	6.1
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

Supplementary Table 4 (Continued): Linear correlation matrices between coverage, target, and bait properties, using the full read set (3.3m - 44.2m) available for each sample.

MAMMOTH 4								UW20579
	NON-ENRICHED			**ENRICHED**				
BASES COVERED (OF 16450 + 10 Ns)	16455			16459				
Non-Enriched Unique Coverage	0.99							
Non-Enriched Duplication Rate	0.16	0.05						
Enriched Raw Coverage	0.08	0.08	-0.01					
Enriched Unique Coverage	0.51	0.51	0.00	0.60				
Enriched Duplication Rate	-0.15	-0.15	-0.02	0.92	0.39			
Raw Enrichment Rate	-0.29	-0.29	-0.09	0.82	0.32	0.91		
Unique Enrichment Rate	-0.46	-0.46	-0.04	0.47	0.42	0.54	0.72	
Reference GC Content	-0.02	-0.01	-0.02	-0.30	-0.25	-0.31	-0.25	-0.22
Bait Coverage	-0.04	-0.03	-0.14	0.35	0.29	0.39	0.36	0.32
Average Bait Hairpin Score	0.02	0.03	-0.05	-0.39	-0.16	-0.38	-0.37	-0.19
Average Bait Dimer Score	0.01	0.01	0.02	-0.39	-0.18	-0.37	-0.36	-0.20
AVERAGE:	22.0	20.8	1.1	15952.7	248.6	55.4	633.8	9.8
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

MAMMOTH 5								LYUBA
	NON-ENRICHED			**ENRICHED**				
BASES COVERED (OF 16450 + 10 Ns)	15910			16448				
Non-Enriched Unique Coverage	0.99							
Non-Enriched Duplication Rate	0.08	-0.02						
Enriched Raw Coverage	-0.02	-0.02	0.01					
Enriched Unique Coverage	0.01	0.02	-0.02	0.74				
Enriched Duplication Rate	-0.06	-0.06	0.02	0.97	0.64			
Raw Enrichment Rate	-0.42	-0.42	-0.03	0.73	0.51	0.74		
Unique Enrichment Rate	-0.65	-0.65	0.00	0.40	0.49	0.38	0.79	
Reference GC Content	0.12	0.11	0.03	-0.24	-0.12	-0.31	-0.20	-0.10
Bait Coverage	0.00	0.00	-0.01	0.38	0.37	0.40	0.27	0.18
Average Bait Hairpin Score	0.03	0.03	0.00	-0.44	-0.35	-0.38	-0.35	-0.23
Average Bait Dimer Score	-0.01	-0.01	-0.01	-0.42	-0.35	-0.37	-0.32	-0.20
AVERAGE:	3.5	3.4	1.0	2928.9	128.2	18.8	3239.5	141.8
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

Supplementary Table 4 (Continued): Linear correlation matrices between coverage, target, and bait properties, using the full read set (3.3m - 44.2m) available for each sample.

MAMMOTH 6								YP180.40
	NON-ENRICHED			**ENRICHED**				
BASES COVERED (OF 16450 + 10 Ns)	12079			16281				
Non-Enriched Unique Coverage	0.87							
Non-Enriched Duplication Rate	0.46	0.03						
Enriched Raw Coverage	0.33	0.33	0.10					
Enriched Unique Coverage	0.39	0.42	0.08	0.70				
Enriched Duplication Rate	0.11	0.09	0.09	0.71	0.15			
Raw Enrichment Rate	-0.33	-0.28	-0.21	0.63	0.37	0.56		
Unique Enrichment Rate	-0.40	-0.49	0.05	0.26	0.50	0.00	0.64	
Reference GC Content	-0.09	-0.06	-0.06	-0.20	-0.08	-0.22	-0.15	-0.04
Bait Coverage	0.01	-0.02	0.02	0.25	0.12	0.30	0.24	0.13
Average Bait Hairpin Score	-0.05	-0.05	-0.05	-0.35	-0.24	-0.24	-0.28	-0.14
Average Bait Dimer Score	-0.05	-0.04	-0.08	-0.33	-0.31	-0.17	-0.27	-0.24
AVERAGE:	2.5	1.9	1.3	1483.5	6.4	221.0	1110.6	5.7
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

MAMMOTH 7								POYSER
	NON-ENRICHED			**ENRICHED**				
BASES COVERED (OF 16450 + 10 Ns)	3489			16315				
Non-Enriched Unique Coverage	0.97							
Non-Enriched Duplication Rate	0.37	0.14						
Enriched Raw Coverage	0.16	0.16	0.04					
Enriched Unique Coverage	0.14	0.13	0.06	0.99				
Enriched Duplication Rate	0.05	0.06	-0.04	0.45	0.38			
Raw Enrichment Rate	-0.19	-0.18	-0.06	0.91	0.92	0.22		
Unique Enrichment Rate	-0.21	-0.22	0.01	0.89	0.92	0.16	0.99	
Reference GC Content	0.03	0.03	0.03	-0.11	-0.12	-0.04	-0.23	-0.22
Bait Coverage	0.01	-0.01	0.08	0.40	0.40	0.22	0.35	0.35
Average Bait Hairpin Score	-0.06	-0.09	0.08	-0.39	-0.38	-0.20	-0.22	-0.21
Average Bait Dimer Score	-0.04	-0.06	0.06	-0.39	-0.38	-0.18	-0.26	-0.25
AVERAGE:	1.2	1.2	1.0	184.8	71.3	2.5	341.4	131.4
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

Supplementary Table 4 (Continued): Linear correlation matrices between coverage, target, and bait properties, using the full read set (3.3m - 44.2m) available for each sample.

MAMMOTH 8 SCARBOROUGH

	NON-ENRICHED			ENRICHED				
BASES COVERED (OF 16450 + 10 Ns)	4597			16444				
Non-Enriched Unique Coverage	1.00							
Non-Enriched Duplication Rate								
Enriched Raw Coverage	0.01	0.01						
Enriched Unique Coverage	0.01	0.01		0.89				
Enriched Duplication Rate	-0.03	-0.03		0.88	0.65			
Raw Enrichment Rate	-0.27	-0.27		0.95	0.86	0.85		
Unique Enrichment Rate	-0.37	-0.37		0.83	0.91	0.65	0.91	
Reference GC Content	0.02	0.02		-0.08	0.02	-0.25	-0.03	0.02
Bait Coverage	-0.01	-0.01		0.43	0.38	0.43	0.39	0.35
Average Bait Hairpin Score	0.06	0.06		-0.40	-0.34	-0.31	-0.49	-0.43
Average Bait Dimer Score	0.01	0.01		-0.41	-0.39	-0.28	-0.50	-0.43
AVERAGE:	1.1	1.1	1.0	523.4	49.5	9.6	281.1	26.1
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

MAMMOTH 9 RAWLINS

	NON-ENRICHED			ENRICHED				
BASES COVERED (OF 16450 + 10 Ns)	16423			16460				
Non-Enriched Unique Coverage	0.95							
Non-Enriched Duplication Rate	0.23	-0.06						
Enriched Raw Coverage	0.16	0.13	0.05					
Enriched Unique Coverage	0.74	0.79	-0.06	0.40				
Enriched Duplication Rate	-0.24	-0.28	0.09	0.80	-0.12			
Raw Enrichment Rate	-0.49	-0.47	-0.21	0.52	-0.15	0.69		
Unique Enrichment Rate	-0.50	-0.52	-0.05	0.10	-0.10	0.17	0.72	
Reference GC Content	0.40	0.45	-0.08	-0.03	0.49	-0.30	-0.28	-0.15
Bait Coverage	-0.06	-0.10	0.09	0.36	0.04	0.41	0.29	0.15
Average Bait Hairpin Score	-0.08	-0.08	-0.02	-0.38	-0.16	-0.30	-0.21	-0.04
Average Bait Dimer Score	-0.09	-0.09	-0.04	-0.38	-0.22	-0.28	-0.20	-0.06
AVERAGE:	14.6	10.5	1.4	6257.0	29.1	220.9	459.6	2.7
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

Supplementary Table 4 (Continued): Linear correlation matrices between coverage, target, and bait properties, using the full read set (3.3m - 44.2m) available for each sample.

MAMMOTH 10								RANDOLPH
	NON-ENRICHED			**ENRICHED**				
BASES COVERED (OF 16450 + 10 Ns)	16459			16459				
Non-Enriched Unique Coverage	0.98							
Non-Enriched Duplication Rate	0.25	0.07						
Enriched Raw Coverage	0.19	0.18	0.10					
Enriched Unique Coverage	0.85	0.86	0.08	0.41				
Enriched Duplication Rate	-0.31	-0.33	0.05	0.78	-0.14			
Raw Enrichment Rate	-0.43	-0.43	-0.13	0.69	-0.17	0.92		
Unique Enrichment Rate	-0.43	-0.45	-0.04	0.27	0.00	0.36	0.58	
Reference GC Content	0.46	0.48	-0.02	-0.02	0.44	-0.31	-0.30	-0.17
Bait Coverage	-0.01	-0.04	0.13	0.37	0.08	0.42	0.37	0.20
Average Bait Hairpin Score	-0.07	-0.07	-0.04	-0.39	-0.13	-0.32	-0.29	-0.06
Average Bait Dimer Score	-0.14	-0.14	-0.04	-0.41	-0.19	-0.30	-0.26	-0.03
AVERAGE:	47.6	35.8	1.3	4936.8	98.2	51.9	226.6	5.5
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

MAMMOTH 11								BINDLOSS
	NON-ENRICHED			**ENRICHED**				
BASES COVERED (OF 16450 + 10 Ns)	2800			16222				
Non-Enriched Unique Coverage	0.92							
Non-Enriched Duplication Rate	0.50	0.14						
Enriched Raw Coverage	-0.07	-0.07	-0.01					
Enriched Unique Coverage	0.02	0.04	-0.05	0.80				
Enriched Duplication Rate	-0.09	-0.15	0.10	0.75	0.34			
Raw Enrichment Rate	-0.22	-0.21	-0.11	0.97	0.75	0.78		
Unique Enrichment Rate	-0.26	-0.26	-0.09	0.78	0.93	0.39	0.81	
Reference GC Content	0.08	0.09	0.00	-0.10	0.09	-0.34	-0.13	-0.02
Bait Coverage	0.05	-0.01	0.14	0.28	0.20	0.32	0.18	0.15
Average Bait Hairpin Score	0.06	0.09	-0.05	-0.37	-0.28	-0.27	-0.38	-0.25
Average Bait Dimer Score	-0.01	0.06	-0.14	-0.37	-0.28	-0.30	-0.37	-0.24
AVERAGE:	1.1	1.1	1.0	299.7	13.5	20.2	481.3	21.6
	Non-Enriched Raw Coverage	***Non-Enriched Unique Coverage***	***Non-Enriched Duplication Rate***	***Enriched Raw Coverage***	***Enriched Unique Coverage***	***Enriched Duplication Rate***	***Raw Enrichment Rate***	***Unique Enrichment Rate***

Supplementary Table 5: Alignment and enrichment rates of the repeat-masked nuclear targets

SAMPLE	NON-ENRICHED 3 MILLION READ SUBSET				ENRICHED 3 MILLION READ SUBSET				ENRICHMENT RATE (FOLD INCREASE)	
	NUCLEAR READS		UNIQUE NUC. READS		NUCLEAR READS		UNIQUE NUC. READS		NUC-LEAR	UNIQUE NUC.
Mammoth 1	1	0.00003%	1	0.00003%	18	0.00060%	7	0.00023%	18.0	7.00
Mammoth 2	16	0.00053%	10	0.00033%	1233	0.04110%	18	0.00060%	77.1	1.80
Mammoth 3	11	0.00037%	10	0.00033%	2802	0.09340%	18	0.00060%	254.7	1.80
Mammoth 4	30	0.00100%	30	0.00100%	2739	0.09130%	592	0.01973%	91.3	19.73
Mammoth 5	6	0.00020%	6	0.00020%	841	0.02803%	123	0.00410%	140.2	20.50
Mammoth 6	5	0.00017%	4	0.00013%	276	0.00920%	11	0.00037%	55.2	2.75
Mammoth 7	0	0.00000%	0	0.00000%	11	0.00037%	8	0.00027%	-	-
Mammoth 8	1	0.00003%	1	0.00003%	26	0.00087%	15	0.00050%	26.0	15.00
Mammoth 9	25	0.00083%	25	0.00083%	1345	0.04483%	149	0.00497%	53.8	5.96
Mammoth 10	36	0.00120%	35	0.00117%	1709	0.05697%	183	0.00610%	47.5	5.23
Mammoth 11	0	0.00000%	0	0.00000%	104	0.00347%	26	0.00087%	-	-
Extr. Blank	0	0.00000%	0	0.00000%	1	0.00003%	1	0.00003%	-	-
Mylodon	0	0.00000%	0	0.00000%	12	0.00040%	5	0.00017%	-	-

Supplementary Table 6: Quantitative PCR and indexing amplification conditions.

Extract 12S quantifications (20µL)					
Reagent	**Concentration**	**Thermal profile**			
Ampli*Taq* Gold® (Applied Biosystems) PCR Buffer II	1X	Initial Denaturation	95°C	5m	
MgCl2	2.5mM	Melt	95°C	20s	
Bovine Serum Albumin	1mg/mL	Anneal	54°C	20s	×55
dNTPs	250mM ea	Extend	72°C	20s	
Primers	200nM ea	Final extension	72°C	1m	
Ampli*Taq*	0.1 U/µL				
SYBR Green® (Invitrogen)	0.167X				
Library 12S quantifications (10µL)					
Reagent	**Concentration**	**Thermal profile**			
Ampli*Taq* Gold® (Applied Biosystems) PCR Buffer II	1X	Initial Denaturation	95°C	5m	
MgCl2	2.5mM	Melt	95°C	20s	
dNTPs	250mM ea	Anneal	54°C	20s	×55
Primers	200nM ea	Extend	72°C	20s	
Ampli*Taq* Gold polym.	0.05U/µL	Final extension	72°C	1m	
EvaGreen® dye (Biotium)	0.5X				

Supplementary Table 6 (Continued): Quantitative PCR and indexing amplification conditions.

Indexed library total quantifications (10µL)					
Reagent	**Concentration**	**Thermal profile**			
Ampli*Taq* Gold® (Applied Biosystems) PCR Buffer II	1X	Initial Denaturation	95°C	5m	
MgCl2	2.5mM	Melt	95°C	20s	
dNTPs	250mM ea	Anneal	62°C	20s	×35
Primers	200nM ea	Extend	72°C	20s	
Ampli*Taq*	0.05U/µL	Final extension	72°C	1m	
EvaGreen® dye (Biotium)	0.5X				
Indexing amplification (50µL)					
Reagent	**Concentration**	**Thermal profile**			
Accuprime™ *Pfx* (Invitrogen) reaction mix	1X	Initial Denaturation	95°C	2m	
Primers	500nM ea	Melt	95°C	15s	
Accuprime™ *Pfx* polym.	0.042U/µL	Anneal	60°C	30s	×11-23
EvaGreen® dye (Biotium)	0.75X	Extend	68°C	45s	
Post-enrichment amplification (24µL)					
Reagent	**Concentration**	**Thermal profile**			
Phusion™ High Fidelity Mastermix (Finnzymes)	1X	Initial Denaturation	98°C	2m	
Primers	300nM ea	Melt	98°C	15s	
EvaGreen® dye (Biotium)	0.625X	Anneal	62°C	30s	×20
		Extend	72°C	30s	

2.12 REFERENCES

1.) **Meyer, M., M. Kircher, M.T. Gansauge, H. Li, F. Racimo, S. Mallick, J.G. Schraiber, F. Jay, et al.** 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science *338*:222-226.

2.) **Poinar, H.N., C. Schwarz, J. Qi, B. Shapiro, R.D. Macphee, B. Buigues, A. Tikhonov, D.H. Huson, et al.** 2006. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science *311*:392-394.

3.) **Green, R.E., J. Krause, A.W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, et al.** 2010. A Draft Sequence of the Neandertal Genome. Science *328*:710-722.

4.) **Avila-Arcos, M.C., E. Cappellini, J.A. Romero-Navarro, N. Wales, J.V. Moreno-Mayar, M. Rasmussen, S.L. Fordyce, R. Montiel, et al.** 2011. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. Sci Rep-Uk *1*.

5.) **Bos, K.I., V.J. Schuenemann, G.B. Golding, H.A. Burbano, N. Waglechner, B.K. Coombes, J.B. McPhee, S.N. DeWitte, et al.** 2011. A draft genome of Yersinia pestis from victims of the Black Death. Nature *478*:506-510.

6.) **Briggs, A.W., J.M. Good, R.E. Green, J. Krause, T. Maricic, U. Stenzel, C. Lalueza-Fox, P. Rudan, et al.** 2009. Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. Science *325*:318-321.

7.) **Burbano, H.A., E. Hodges, R.E. Green, A.W. Briggs, J. Krause, M. Meyer, J.M. Good, T. Maricic, et al.** 2010. Targeted investigation of the Neandertal genome by array-based sequence capture. Science *328*:723-725.

8.) **Fu, Q., M. Meyer, X. Gao, U. Stenzel, H.A. Burbano, J. Kelso, and S. Paabo.** 2013. DNA analysis of an early modern human from Tianyuan Cave, China. Proceedings of the National Academy of Sciences of the United States of America *110*:2223-2227.

9.) **Horn, S.** 2012. Case study: enrichment of ancient mitochondrial DNA by hybridization capture. Methods Mol Biol *840*:189-195.
10.) **Krause, J., A.W. Briggs, M. Kircher, T. Maricic, N. Zwyns, A. Derevianko, and S. Paabo.** 2010. A complete mtDNA genome of an early modern human from Kostenki, Russia. Current biology : CB *20*:231-236.
11.) **Krause, J., Q. Fu, J.M. Good, B. Viola, M.V. Shunkov, A.P. Derevianko, and S. Paabo.** 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature *464*:894-897.
12.) **Mason, V.C., G. Li, K.M. Helgen, and W.J. Murphy.** 2011. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. Genome research *21*:1695-1704.
13.) **Reich, D., R.E. Green, M. Kircher, J. Krause, N. Patterson, E.Y. Durand, B. Viola, A.W. Briggs, et al.** 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature *468*:1053-1060.
14.) **Schuenemann, V.J., K. Bos, S. DeWitte, S. Schmedes, J. Jamieson, A. Mittnik, S. Forrest, B.K. Coombes, et al.** 2011. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. Proceedings of the National Academy of Sciences of the United States of America *108*:E746-752.
15.) **Sawyer, S., J. Krause, K. Guschanski, V. Savolainen, and S. Paabo.** 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PloS one *7*:e34131.
16.) **Bouwman, A.S., S.L. Kennedy, R. Muller, R.H. Stephens, M. Holst, A.C. Caffell, C.A. Roberts, and T.A. Brown.** 2012. Genotype of a historic strain of Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences of the United States of America *109*:18511-18516.
17.) **Vilstrup, J.T., A. Seguin-Orlando, M. Stiller, A. Ginolhac, M. Raghavan, S.C. Nielsen, J. Weinstock, D. Froese, et al.** 2013. Mitochondrial phylogenomics of modern and ancient equids. PloS one *8*:e55950.
18.) **Fu, Q., A. Mittnik, P.L. Johnson, K. Bos, M. Lari, R. Bollongino, C. Sun, L. Giemsch, et al.** 2013. A revised timescale for human evolution based on ancient mitochondrial genomes. Current biology : CB *23*:553-559.
19.) **Farrell Jr, R.E.** 2010. Chapter 13 - Practical Nucleic Acid Hybridization, p. 283-299. RNA Methodologies (4th Edition). Academic Press, San Diego.
20.) **Harrison, A., H. Binder, A. Buhot, C.J. Burden, E. Carlon, C. Gibas, L.J. Gamble, A. Halperin, et al.** 2013. Physico-chemical foundations underpinning microarray and next-generation sequencing experiments. Nucleic acids research *41*:2779-2796.
21.) **Allentoft, M.E., M. Collins, D. Harker, J. Haile, C.L. Oskam, M.L. Hale, P.F. Campos, J.A. Samaniego, et al.** 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proceedings. Biological sciences / The Royal Society *279*:4724-4733.
22.) **Deagle, B.E., J.P. Eveson, and S.N. Jarman.** 2006. Quantification of damage in DNA recovered from highly degraded samples--a case study on DNA in faeces. Frontiers in zoology *3*:11.
23.) **Schwarz, C., R. Debruyne, M. Kuch, E. McNally, H. Schwarcz, A.D. Aubrey, J. Bada, and H. Poinar.** 2009. New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. Nucleic acids research *37*:3215-3229.
24.) **King, C., R. Debruyne, M. Kuch, C. Schwarz, and H. Poinar.** 2009. A quantitative approach to detect and overcome PCR inhibition in ancient DNA extracts. BioTechniques *47*:941-949.
25.) **Wales, N., J.A. Romero-Navarro, E. Cappellini, and M.T.P. Gilbert.** 2012. Choosing the Best Plant for the Job: A Cost-Effective Assay to Prescreen Ancient Plant Remains Destined for Shotgun Sequencing. PloS one *7*:e45644.
26.) **Meyer, M. and M. Kircher.** 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor protocols *2010*:pdb prot5448.
27.) **Kircher, M., S. Sawyer, and M. Meyer.** 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic acids research *40*:e3.
28.) **Barnes, I., B. Shapiro, A. Lister, T. Kuznetsova, A. Sher, D. Guthrie, and M.G. Thomas.** 2007. Genetic structure and extinction of the woolly mammoth, Mammuthus primigenius. Current biology : CB *17*:1072-1075.
29.) **Debruyne, R., G. Chu, C.E. King, K. Bos, M. Kuch, C. Schwarz, P. Szpak, D.R. Grocke, et al.** 2008. Out of America: Ancient DNA evidence for a new world origin of late quaternary woolly mammoths. Current Biology *18*:1320-1326.

30.) **Gilbert, M.T.P., D.I. Drautz, A.M. Lesk, S.Y.W. Ho, J. Qi, A. Ratan, C.H. Hsu, A. Sher, et al.** 2008. Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. Proceedings of the National Academy of Sciences of the United States of America *105*:8327-8332.

31.) **Krause, J., P.H. Dear, J.L. Pollack, M. Slatkin, H. Spriggs, I. Barnes, A.M. Lister, I. Ebersberger, et al.** 2006. Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. Nature *439*:724-727.

32.) **Enk, J., A. Devault, R. Debruyne, C.E. King, T. Treangen, D. O'Rourke, S.L. Salzberg, D. Fisher, et al.** 2011. Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. Genome biology *12*:R51.

33.) **Rohland, N., A.S. Malaspinas, J.L. Pollack, M. Slatkin, P. Matheus, and M. Hofreiter.** 2007. Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup. Plos Biol *5*:1663-1671.

34.) **Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, and S.G. Rozen.** 2012. Primer3--new capabilities and interfaces. Nucleic acids research *40*:e115.

35.) **Martin, M.** 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*:10-12.

36.) **Magoc, T. and S.L. Salzberg.** 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*:2957-2963.

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

38.) **Neph, S., M.S. Kuehn, A.P. Reynolds, E. Haugen, R.E. Thurman, A.K. Johnson, E. Rynes, M.T. Maurano, et al.** 2012. BEDOPS: high-performance genomic feature operations. Bioinformatics *28*:1919-1920.

39.) **Ginolhac, A., M. Rasmussen, M.T. Gilbert, E. Willerslev, and L. Orlando.** 2011. mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics *27*:2153-2155.

40.) **Dabney, J. and M. Meyer.** 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. BioTechniques *52*:87-94.

41.) **Mokry, M., H. Feitsma, I.J. Nijman, E. de Bruijn, P.J. van der Zaag, V. Guryev, and E. Cuppen.** 2010. Accurate SNP and mutation detection by targeted custom microarray-based genomic enrichment of short-fragment sequencing libraries. Nucleic acids research *38*:e116.

42.) **Kozarewa, I., Z. Ning, M.A. Quail, M.J. Sanders, M. Berriman, and D.J. Turner.** 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nature methods *6*:291-295.

43.) **Kozarewa, I. and D.J. Turner.** 2011. Amplification-free library preparation for paired-end Illumina sequencing. Methods Mol Biol *733*:257-266.

44.) **Aird, D., M.G. Ross, W.S. Chen, M. Danielsson, T. Fennell, C. Russ, D.B. Jaffe, C. Nusbaum, and A. Gnirke.** 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome biology *12*:R18.

45.) **Kihana, M., F. Mizuno, R. Sawafuji, L. Wang, and S. Ueda.** 2013. Emulsion PCR-coupled target enrichment: An effective fishing method for high-throughput sequencing of poorly preserved ancient DNA. Gene *528*:347-351.

46.) **Colotte, M., V. Couallier, S. Tuffet, and J. Bonnet.** 2009. Simultaneous assessment of average fragment size and amount in minute samples of degraded DNA. Analytical biochemistry *388*:345-347.

Chapter 3

Mitogenomic phylogeography of *Mammuthus* in North America

JACOB ENK, ALISON DEVAULT, CHRIS WIDGA, JEFFREY SAUNDERS, PAUL SZPAK, JOHN SOUTHON, JEAN-MARIE ROUILLARD, GRANT ZAZULA, DUANE FROESE, ROSS MACPHEE, DANIEL FISHER, and HENDRIK POINAR
In preparation for submission to *Molecular Ecology*

3.1 PREFACE

Here I test systematic hypotheses put forward in Chapter 1 by capturing and sequencing dozens of mammoth mitochondrial genomes from southern areas of their range. This simultaneously allows me to examine whether mammoths in both the periglacial and temperate Pleistocene biomes of North America underwent disparate population histories. I conclude that of various methods for calibrating the mammoth molecular clock, one that uses solely the geologic ages the samples are probably misleading. This further refines biogeographic and mammoth behavioral hypotheses that can be tested when nuclear data are retrieved from very poorly-preserved remains like those mitogenomically characterized here.

3.2 ABSTRACT

After evolving in Africa during the Miocene, *Mammuthus* (mammoths) spread through much of the northern hemisphere and diversified as they entered multiple major habitats. In Pleistocene North America alone, several mammoth species are recognized paleontologically, occupying both the cold tundra-steppe of the north and the arid grasslands and temperate savanna-parklands of the south. However, mammoth genetic and phylogeographic studies have been overwhelmingly focused on the permafrost-preserved remains of Late Pleistocene arctic and subarctic populations of just one mammoth species, *M. primigenius* (woolly mammoth). This restriction to just northern mammoth populations likely biases our understanding of the global population dynamics of this keystone megaherbivore before their ultimate extinction. Here we overcome the taphonomic and technological limitations responsible for this bias, and take a multi-biome survey of Late Pleistocene mammoth genetic diversity across North America. Using a sensitive targeted enrichment technique we sequenced 67 complete mitochondrial genomes of mammoth specimens distributed throughout their ecological range, including from poorly-preserved specimens of *M. columbi* (Columbian mammoth), *M. jeffersonii* (Jeffersonian mammoth), and *M. exilis* (pygmy mammoth). While we uncovered clear phylogeographic structure in mammoth matrilines, their genus-wide mitochondrial phylogeny is not immediately compatible with conventional interpretations of the mammoth paleontological record. It instead suggests that various mammoth paleontological species were likely biologically conspecific, and therefore ecomorphotypes of one widespread and highly variable species. We hypothesize that at least two distinct stages of interbreeding between paleontological species are likely responsible for this pattern – one between Siberian woolly mammoths and resident American populations that introduced woolly mammoth phenotypes into the continent, and another between distinct southern ecomorphotypes (woolly and Columbian mammoths) in North America. This also suggests that mammoth mitochondrial phylogeny reconstructed to date has a much deeper chronology, and thus taxonomic context, than is suggested when sample ages are alone used to calibrate the mammoth molecular clock.

3.3 INTRODUCTION

Mammuthus (mammoth) evolution spans millions of years and much of the planet. After emerging in Africa in the Late Miocene, mammoths spread throughout the northern hemisphere, reaching the Eurasian subarctic and the New World by the Early Pleistocene (Lister 1996; Maglio 1973; Sanders 2010; Todd & Roth 1996). Current

interpretations of the paleontological record specify at least two and potentially three immigrations of *Mammuthus* from Siberia into North America over the Bering Land Bridge during the Pleistocene (Agenbroad 2005; Harington 1984; Kurtén & Anderson 1980; Lister & Bahn 2007; Lister & Sher 2001; MacFadden & Hulbert 2009; Sanders 2010) (Figure 1). In these scenarios, an Early to Middle Pleistocene immigrant, *M. trogontherii* (steppe mammoth), reached well into the southern latitudes and then evolved locally into *M. columbi* (Columbian mammoth) and other endemic North American species *M. exilis* (pygmy mammoth) and *M. jeffersonii* (Jeffersonian mammoth). These species primarily occupied the savanna-parkland and arid grasslands of what were at the time and remain today temperate regions of the continent. Meanwhile, steppe mammoths in the tundra-steppe of Siberia evolved into *M. primigenius* (woolly mammoth), which immigrated to North America sometime in the Late Pleistocene and spread southward along the Laurentide periglacial steppe into the Great Lakes regions and Atlantic Seaboard (Agenbroad 2005; Fisher 2009; Harington 1984; Saunders *et al.* 2010). Subarctic woolly mammoths are one of the best studied of the extinct Pleistocene megafauna at the genetic level, thanks to two decades of research of mostly their mitochondrial DNA (Barnes *et al.* 2007; Debruyne *et al.* 2008; Gilbert *et al.* 2008; Hagelberg *et al.* 1994; Höss *et al.* 1994; Krause *et al.* 2006; Nystrom *et al.* 2010; Ozawa *et al.* 1997; Palkopoulou *et al.* 2013). These studies uncovered three highly divergent mitochondrial lineages in their populations (clades I, II, and II; equivalent to haplogroups C-D-E, A, and B), and a phylogeography that suggests a dynamic population history of immigrations, contractions, expansions and replacements of endemic matrilines (Debruyne *et al.* 2008; Gilbert *et al.* 2008; Palkopoulou *et al.* 2013).

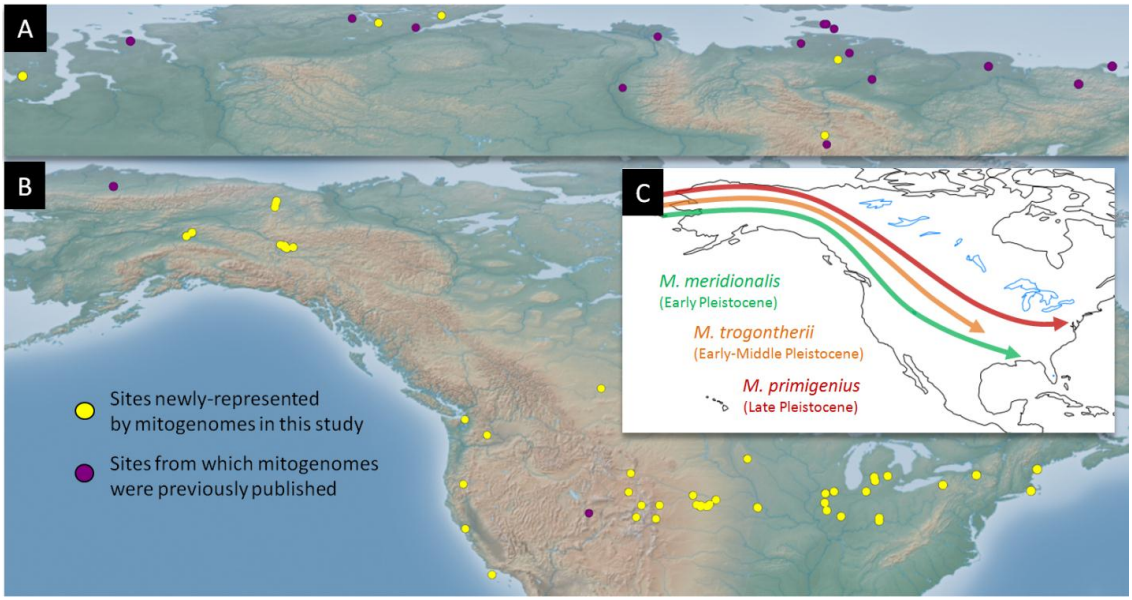


Figure 1: Site locations represented in this study and schematic models of *Mammuthus* immigrations into North America. Sites in both the Old (A) and New (B) Worlds are depicted. The site in the Alaskan North Slope had been previously-represented by a complete mitogenome but we add several more in this study. (C) is a schematic illustration of three separate Pleistocene immigrations of *Mammuthus* species; some authors support a two-immigration model of only *M. trogontherii* and *M. primigenius*, while others support an earlier immigration of *M. meridionalis* in the Early Pleistocene, as indicated by plesiomorphic forms found in Idaho, California, Nebraska and Florida. While older models suggest that *M. columbi* descend from *M. meridionalis*, more recent models suggest that *M. columbi* descended from *M. trogontherii*.

In sharp contrast, very little is known about the genetic diversity and structure of mammoth populations south of northwestern Canada. In an initial investigation, Enk et al. (2011) found that two terminal Pleistocene Columbian

mammoths were mitochondrially closely related to North American woolly mammoths (closest to Clade I / haplogroup C) – more closely related than woolly mammoths on both continents were to each other. This suggests a more complex genetic history for *Mammuthus* in North America than is implied by the conventional model of *M. columbi* and *M. primigenius* evolving independently for over a million years. Rather it suggests that at some point, the two paleontological species exchanged genes. However, since this signal derives from the mitochondrial DNA of just two very late Pleistocene Columbian mammoths, it is not immediately clear whether this close woolly-Columbian mitogenomic affinity was typical or exceptional. If exceptional, it is possible that those Columbian mammoths studied previously had acquired woolly mammoth-like mitochondrial genomes through relatively isolated hybridization events, for which there is behavioral precedent in extant elephants (Roca *et al.* 2005) and which some authors suggest might explain intermediate woolly-Columbian morphologies of certain mammoth specimens found in Pleistocene ecotonal regions (Fisher 2009; Hoyle *et al.* 2004). On the other hand, if this close mitogenomic affinity between the two taxa is not exceptional but rather typical, it is more likely that North American woolly mammoths acquired their mitochondrial genomes from endemic North American populations that descended from the initial *M. trogontherii* immigration in the Early/Middle Pleistocene. In order to address whether the signal retrieved by Enk et al. (2011) is typical or exceptional for Columbian mammoths, and therefore to determine its most likely origin, a survey of mitochondrial diversity in non-woolly North American mammoths is needed. In addition to addressing questions of mammoth systematics and behavior, such a survey would also potentially allow us to examine the population structure and dynamics of *Mammuthus* in previously-uncharacterized areas of their range. As keystone megaherbivores, mammoths greatly impacted the floral and faunal communities with which they coexisted, and so such an expansion could potentially provide a much more comprehensive vision of how Pleistocene ecosystems evolved and eventually went extinct.

However, it has only recently become viable to assemble large phylogeographic datasets from Pleistocene specimens found in the thermolabile burial contexts of lower latitudes. The relatively dry and thermostable depositional contexts of northern contexts more readily preserve DNA in fragment lengths practical for gene resequencing with overlapping PCR amplicons (Kircher 2012; Mitchell *et al.* 2005; Smith *et al.* 2003), which is the typical approach in ancient DNA science. Constraint to PCR-based approaches has also led to the predominant use of relatively short contiguous sequences (< 800bp) for reconstructing megafaunal phylogeography, including mammoths (Barnes *et al.* 2007; Debruyne *et al.* 2008; Nystrom *et al.* 2010; Palkopoulou *et al.* 2013). Fortunately, recent technological advances in targeted enrichment and high throughput sequencing allow heavily fragmented DNA found in more exposed and thermolabile contexts to be efficiently sequenced (Bos *et al.* 2011; Carpenter *et al.* 2013; Enk *et al.* 2013; Wagner *et al.* 2014). Here we use these technologies to capture, sequence and analyze the complete mitochondrial genomes of relatively poorly-preserved mammoth remains that died along and south of the Laurentide ice margin. As such we broadly survey North American mammoth mitogenomic diversity, as well as reconstruct aspects of their population structure and dynamics during the Late Pleistocene.

3.4 MATERIALS & METHODS

3.4.1 Summary

We collected 234 *Mammuthus* specimens from sites in the U.S. west coast, mountain west, Great Plains, Great Lakes, and east coast. These include *M. primigenius*, *M. columbi*, *M. exilis* or pygmy mammoth from the Channel Islands off the coast of California, and *M. jeffersonii* or Jeffersonian mammoth, a Late Pleistocene paleontological species exhibiting intermediate *M. primigenius* – *M. columbi* morphology (Kurtén & Anderson 1980; Pasenko & Schubert 2004; Saunders *et al.* 2010). We also included 42 specimens from northern latitudes that had either been previously characterized for the cytochrome *b* – HVR region (Debruyne *et al.* 2008) or have yet to be analyzed genetically. We then extracted DNA and screened the extracts with a 49bp proboscidean mitochondrion-specific

quantitative PCR assay. Specimens exhibiting consistent positive amplification or of particular geographic interest were then converted to Illumina sequencing libraries. The majority of these were subsequently enriched in solution once or twice for the complete mitochondrial genome with a set of biotinylated RNA oligonucleotide baits. We then sequenced these in multiplex on an Illumina Hiseq 1500 platform, curated the read data and aligned the reads to a mammoth mitogenome reference, and called consensus sequences. We AMS radiocarbon dated several specimens that yielded complete or near-complete mitochondrial genomes but had not been dated previously. Following consensus generation and combining our new dataset with previously-published mammoth mitogenome sequences, we performed a series of phylogenetic analyses to estimate lineage relationships, cladogenic chronology, and population dynamics.

3.4.2 Sample collection

Mammuthus hard and soft tissue specimens were collected from the University of Michigan Museum of Natural History, the Illinois State Museum, the University of Nebraska State Museum, the Denver Museum of Nature and Science, the University of California Museum of Paleontology, the Santa Barbara Museum of Natural History, Central Washington University, Northern Arizona University, the Geological Museum of the University of Wyoming, the Canadian Museum of Nature, and the Yukon Paleontological collections. Specimens came from 25 U.S. states, two Canadian provinces, and Siberia. Several of these specimens have been formally described and allocated to taxon on the basis of morphological traits to one or another of the nominal taxa *M. primigenius*, *M. columbi*, *M. jeffersonii*, and *M. exilis*. In some cases, the basis for allocation was unclear and we did not make an effort to revise them, while for others we allocated specimens to taxon following Maglio (1973). We emphasize that, in light of the genetic results provided here, a thorough re-evaluation of the morphological bases for discriminated taxa within *Mammuthus* would be beneficial. Locality information, taxon, and radiocarbon ages (when available) of the specimens yielding complete or near-complete mitogenomes are included in Supplemental Table 1A.

3.4.3 DNA extraction

Specimens were initially subsampled by the authors or contributing scientists at the site of curation. Once at the McMaster Ancient DNA Centre, specimens were handled in dedicated ancient DNA laboratory facilities, where we further subsampled 50-500mg of cementum, dentin, bone, coprolite and muscle tissue remains and pulverized these with a hammer to particle sizes ranging from powder to 1-5mm crumbles. A bone specimen from a Pleistocene *Mylodon darwinii* was also subsampled and included in all extraction sets in a ratio of at least 1:16 mylodon:mammoth to serve as a negative control for monitoring contamination. Subsamples were then subjected to an initial wash for 0.5 to 1.5h in 0.5M EDTA (pH 8.0) with agitation at room temperature, centrifuged and decanted. These pre-washed pellets were then demineralized with EDTA overnight at room temperature, and the supernatants removed following centrifugation. Then the pellets were either demineralized again or digested with a Tris-HCl-based proteinase K digestion solution with between 0 to 0.5% sodium lauryl sarcosine (Fisher Scientific), 0 to 1% polyvinylpolypyrrolidone (PVP, Fisher scientific), 0 to 50mM dithiothreitol, 0 to 2.5mM N-phenacyl thiazolium bromide (PTB, Prime Organics), and 2.5mM calcium chloride (CaCl2). These proteinase digestions were performed at room temperature overnight, or between 37 and 55°C for 3-10h with agitation. Following centrifugation the digestion supernatants were removed and in most cases pooled with the demineralization supernatants, although in some cases they were kept separate. For several pellets that remained undigested, we repeated this process, pooling with the original rounds or keeping them separate for use in shotgun sequencing. Pooled or individual supernatants were then extracted of organics using phenol:chloroform:isoamyl alcohol (PCI, 25:24:1) and the resulting post-centrifugation aqueous again extracted with chloroform:isoamyl alcohol (25:1). We then concentrated the final aqueous phases with 10kDA or 30kDA Amicon centrifuge filters (Millipore) at 7k to 10k x g, with up to four washes with 0.1X or 1X TE buffer (pH 7-8.5) to provide final desalted concentrates of 25-100µL.

3.4.4 Quantitative PCR screens

It has been previously shown that a short single-locus quantitative PCR assay can be used to predict on-target ancient DNA high-throughput sequencing read counts both before and after targeted enrichment with some accuracy (Enk *et al.* 2013). With this in mind we screened the extracts in duplicate for a short 49bp portion of the mammoth mitochondrial *12S* gene using the quantitative PCR protocol in Supplemental Table 2A using 1µL of 0.1X concentration of the extracts diluted with 0.1X or 1X Tween-TE (buffer TE with 0.05% Tween-20). Successful amplification was strongly biased towards specimens from northern latitudes and eastern longitudes, with no specimens from Oklahoma, Texas, New Mexico, Nevada, South Carolina or Florida yielding amplifiable DNA. Only three of 70+ specimens from California showed consistent amplification of the 49bp locus, with an increase in successful amplification rates of the whole North American dataset moving eastward. Specimens from deposits just south of or immediately adjacent to the Laurentide ice showed good rates of amplification, consistent with the notion that the thermostability of their immediate depositional context played a role in long-term survival of DNA, despite relatively variable post-Pleistocene thermal histories.

3.4.5 Library preparation

For further treatment we chose extracts exhibiting consistent duplicate amplification, some extracts that did not amplify consistently but were geographically or taxonomically interesting to this project, and associated mylodon control extractions. We purified these with MinElute columns (QIAGEN) to 20-40µL EBT and converted them to double-stranded, UDG-treated Illumina sequencing libraries (Kircher *et al.* 2012; Meyer & Kircher 2010) according to the recipes in Supplemental Table 2B. A portion of these resulting libraries were then double-indexed with P5 and P7 indexing primers (Kircher *et al.* 2012) and purified again with Minelute to 13µL EBT. For most samples we screened their indexed libraries again with the 49bp locus, with these metrics and extract qPCR values available in Supplemental Table 3. The majority of extracts that screened positive for the 49bp locus were again consistently positive following indexing and purification; the target amplified in no mylodon controls.

3.4.6 Targeted enrichment

We designed a set of 100bp baits tiled every 5 bases across six mitochondrial genome sequences, including one representative of each mammoth haplogroup known at the time (GenBank Accession #NC015529, EU153447, EU153453, EU153456, and a mitogenome for the specimen Lyuba we generated early in this project) as well as the mitogenome of *Mammut americanum* (NC009547). The variable tandem repeat section (VNTR) of the D-loop was masked with 10 Ns prior to bait design, as it is too long to resolve with short read sequencing. In light of evidence that bait coverage across targets can result in coverage biases in target read coverage (Enk *et al.* 2013; Mokry *et al.* 2010), we chose not to collapse baits of identical sequence prior to manufacture. Baits were then synthesized at MYcroarray as part of several MYbaits targeted enrichment kits.

We used 10µL of each indexed library in 36-38 hour hybridization reactions at 48°C following the MYbaits targeted enrichment protocol, replacing some kit constituents with higher concentration versions to accommodate the extra library input volume. Phosphate-group end-blocked oligonucleotides matching one strand of the regions flanking the 7bp indexes of the library adapters were included. We used 50ng of baits per reaction, which is at least double what we expect to be sufficient for very sensitive capture of a target region of this size using short ancient DNAs (Enk *et al.* 2013). Following hybridization the reactions were cleaned according to the kit protocol except with 200µL rather than 500µL volumes of wash buffers for each wash step, to accommodate 96-well plate-format manipulation. Hot washes were performed at 48°C. These enriched libraries were eluted and then purified with MinElute to 13µL EBT, which we then re-amplified according to the protocol in Supplemental Table 2A and again purified to 13µL EBT.

3.4.7 Illumina sequencing

Enriched libraries were pooled and sequenced on an Illumina HiSeq® 1500 platform using a 2 x 70bp ("SE"-prefix libraries), 2 x 80 ("VE"- and "SVE"-prefix libraries), or 2 x 85bp ("EE"-prefix libraries) paired-end double-index protocol. Both cBot cluster generation and sequencing employed the v3 chemistry and a dual 7bp indexing protocol, using the alternative primer mixes from the TruSeq Dual Index Sequencing Primer Kit (Paired End). We included a dedicated control lane with the PhiX control kit v3 and a 1% PhiX spike in each lane. Raw data was processed with HCS version 1.5.15.1 and RTA version 1.13.48.0. File conversion and demultiplexing using each 7bp reverse index (requiring a 100% match) was performed with CASAVA version 1.8.2.

3.4.8 Libraries with prefix "EID"

Some of the mitogenomes analysed in this study were generated simultaneously with those described elsewhere (Enk *et al.* 2013), and are indicated with an "EID" prefix in Supplemental Table 3. Major differences between those experiments and the ones described here are that those libraries were prepared without UDG treatment, index-amplified for varying numbers of cycles, and enriched once at 45°C with a duplicate bait-collapsed proboscidean mitogenome bait set.

3.4.9 Shotgun sequencing

Two libraries included here (Oimyakon and Lyuba / SIDs04-11 and 36-10) were prepped and indexed at the McMaster Ancient DNA Centre and then sent to Harvard Medical School for further processing. There they size-selected the libraries for inserts longer than 40bp with gel electrophoresis, purified them with the QiaQuick Gel Purification kit (QIAGEN) and reamplified them. They then sequenced these libraries on an Illumina MiSeq® platform using a paired-end 2 x 75bp single-index protocol, with the read data post-processed using the default settings in the MiSeq® Control Software. For another specimen (2005/915), four libraries generated from different extract fractions were sequenced on an Illumina HiSeq 1500 platform using a 2 x 64bp paired-end double-index protocol and post-processed as with the other HiSeq data. Data from these four 2005/915 libraries were combined, and all shotgun read sets were then carried through with analysis pipeline of the enriched library read sets, described below.

3.4.10 Data curation

Read 1 and Read 2 sequence files were trimmed of adapter sequence, simultaneously merged, and filtered of short reads with SeqPrep (https://github.com/jstjohn/SeqPrep), keeping only reads 24bp and above (-L 24), requiring an 11bp overlap between paired reads (-o 11) and searching for the 13bp universal linker on both the P5 and P7 adapters (-A AGATCGGAAGAGC). Merged reads were then combined with the non-mergeable adapter-trimmed Read 1 reads for a final read dataset used for alignment. This inclusion of unmerged reads are expected to simultaneously inflate and deflate unique read counts, with a bias towards deflation due to inserts of variable length potentially mapping to identical 5'-3' coordinates due to only partial sequencing of those inserts.

3.4.11 Alignment and consensus calling

We used BWA version 0.6.1-r104 (Li & Durbin 2009) to align reads to an *M. columbi* mitochondrial genome sequence (GenBank Accession #NC015529) using the parameters described in Schubert et al. (2012). We then collapsed reads with identicial 5' and 3' coordinates and strand, keeping reads with the highest mapping quality scores, to generate final unique read alignments. Alignments were then inspected in Geneious Pro version R6.1.6 (Kearse *et al.* 2012) and consensuses were generated for positions covered by at least 3 unique reads, while

positions with fewer than 3 unique reads coverage were called as N. We repeated this process for the 3'-5' junction of the mitochondrial genome as well as the regions flanking the VNTR to resolve as many positions as possible. Strict consensuses were called requiring more than 50% of reads for each base to match. Consensus sequences were combined with a number of other complete mammoth mitogenome sequences previously published (Supplemental Table 1B) then aligned to an *Elephas maximus* mitogenome sequence (GenBank Accession #EF588275) in Geneious and manually adjusted at apparently misaligned regions, referring back to the original read alignments to confirm the proper configuration. The final to-*Elephas* alignment served as input for phylogenetic analysis.

3.4.12 Second-round enrichment and sequencing

Mitogenomes were considered "complete" if at least 90% of bases were covered by at least 3 unique reads. Several libraries did not provide this after the first round of enrichment and sequencing, and so for these we reconstructed complexity curves to gauge whether all unique mitochondrial molecules had been characterized in the first round. To do this we first converted the pre-collapsed alignments to BED files with BEDOPS version 2.2.0 (Neph *et al.* 2012) and then randomly sampled these in 1% increments and collapsed to unique reads. In cases where the complexity curves had reached plateau, we calculated whether the mitogenome could be obtained by enriching and sequencing the second half of the original libraries, or whether residual extract itself would have to be re-prepped and enriched and sequenced. In cases where the complexity curves had not reached plateau but were non-linear, we used preseq version 0.0.3 (Daley & Smith 2013) to predict the unique read yield with deeper sequencing. In cases where sufficient numbers of unique molecules could be obtained with an economical amount of additional sequencing, we sequenced these libraries further in a second Illumina run ("SSE"-prefix) or indexed more of the original library and enriched them once ("SSEP/D####" format). In cases where only very deep sequencing would be required for resolving the mitogenome to sufficient depth, we enriched these libraries again using an identical protocol as the first round. These doubly-enriched libraries were then sequenced again ("SVEP/D##" name format), with mapping rates after this second round in Supplemental Table 3. Reads from the same specimen were then combined and the read alignment, collapse, consensus generation, and final alignment were repeated.

3.4.13 AMS radiocarbon dating

Specimens yielding complete or near-complete mitogenome sequences that had not been previously radiocarbon dated were further subsampled. Collagen from these was then extracted and purified at the University of Western Ontario, and AMS-dated at University of California, Irvine. Radiocarbon ages are included in Supplemental Table 1A. Although some specimens (ISM01, UCMP04, UCMP17, UNSM08, UNSM42, and UW20579) yielded carbon:nitrogen ratios > 3.00, which may be indicative of carbon contamination, for the purposes of this paper we treat the estimated dates as appropriate estimates of their geological ages. To calibrate AMS dates to calendar years we used OxCal 4.2 (Ramsey & Lee 2013) with the IntCal13 (Reimer *et al.* 2013) calibration curve, reporting the median calendar date estimates in Supplemental Table 1A even in cases where they are beyond the range of calibration.

3.4.14 Maximum likelihood tree estimation

We used jModelTest 2 (Darriba *et al.* 2012) to choose the nucleotide substitution model (with 5 gamma-rate categories) that best fit the data according to the corrected Aikake Information Criterion [AICc, (Hurvich & Tsai 1989)] for the final alignment of our 67 new complete mitochondrial genome sequences and 21 previously-published elephantid mitogenome sequences. We then used IQ-TREE version 0.9.6 (Minh *et al.* 2013) on the full dataset to estimate a maximum likelihood (ML) tree, using the jModelTest-estimated AICc best model (GTR+I+G5)

gamma distribution (-a 0.568) and proportion of invariable sites (-i 0.579). One hundred bootstrap iterations were performed (-b 100). We then estimated the Bayesian posterior probability support for the nodes in the ML consensus tree using BEAST version 1.8 (Drummond *et al.* 2012) forcing monophyly among *Mammuthus,* a calibration-free strict clock and a general time-reversible model with a proportion of invariant sites and a gamma distribution on the substitution rates (GTR+I+G) with 5 gamma categories and broad uniform priors on both kappa parameters the alpha parameter (0 to 1000 and 0 to 10, respectively). We ran this for 10m generations sampling every 1k generation, and following a 10% burn-in the posterior distribution of the tree likelihood securely converged with an effective sample size of 2310. Priors and posteriors for these analyses are in Supplemental Table 4 (currently tabulated based on original radiocarbon dates, not calendar dates, to-be-updated for final publication). We then annotated the BEAST-estimated trees to the ML tree using treeannotator. The ML tree with nodal bootstrap and posterior probability support are depicted in Figure 2. We repeated this process with an alignment that included the partial mitochondrial genome consensuses called requiring a minimum of either 3x or 2x unique depth, depicted in Supplemental Figure 1.

3.4.15 Mutation rate and tMRCA estimations

To evaluate the temporal signal in our data we performed several 10M generation date randomization tests (Ho *et al.* 2008) in BEAST on an alignment including only specimens with finite AMS ^{14}C dates, the best model for which was again chosen according to the AICc in jModelTest 2 [Tamura-Nei (Tamura & Nei 1993)+ I + G]. These analyses used only the median calibrated ages of the samples as single-point calibration points (a.k.a, a "tips-only" calibration scheme). Each of the 95% highest posterior densities (HPDs) of the estimated mutation rates using shuffled tip dates fell outside of the HPD of that rate when using the actual tip dates, suggesting strong temporal signal in the data (Supplemental Figure 2).

To estimate the times to most recent common ancestry (tMRCAs) as represented by the various nodes in the reconstructed phylogeny, we performed a series of BEAST analyses under a tips-only scheme as well as two other calibration schemes that using internal nodal calibrations gleaned from fossil record chronology. In one, only the age of the estimated *Elephas-Mammuthus* split was used for temporal calibration ("root only"), set to that estimated by Rohland *et al*. (2007) as a normal prior distribution centering on 6.7e6 years ago with a standard deviation of 5e5 years. In a third scheme, both this root prior and individual tip priors were used ("root + tip"). In all cases the TN+I+G model was employed with an uncorrelated lognormal clock and constant population size prior. Prior distributions, the numbers of generations used, mean posterior values, 95% highest posterior density bounds (HPDs), and posterior effective sample sizes on all relevant parameters are indicated in Supplemental Table 4 (currently tabulated based on original radiocarbon dates, not calendar dates, to-be-updated for final publication).

We performed another series of temporal analyses using a single additional sample, CMNH40031, a tooth found *in situ* at an Old Crow (Yukon) locality below the Chester Bluff tephra. This horizon is estimated to be at least MIS7 (~250kya) in age, potentially even older, but younger than 780kya. Therefore we assigned a conservative tip date to this specimen of 200kya to gauge its effect on the tMRCA estimates.

3.4.16 Bayesian Skygrid analyses

In order to explore potential population size dynamics recovered from the mitogenome phylogeny, we performed a series of Bayesian Skygrid (Gill *et al.* 2013) analyses using just the mammoths sequences. For temporal calibration we employed two separate calibration schemes. In one, tip dates alone were used. In another, internal nodal dates estimated from the tMRCA analyses with *Elephas* included (above section) were used as nodal temporal priors in addition to tip dates. All other priors were identical for those used for tMRCA estimations

(where in common). In addition to using the alignment with all finite-dated mammoth specimens, we also performed individual analyses to compare trajectories between each member of three separate paired groups; specimens from northern (Alaska, Yukon, and Siberia) vs. southern localities, periglacial (northern and Great Lakes specimens) vs. temperate localities, and old world vs. new world localities. Analyses used the substitution model identified in jModelTest for each individual partition, and were run for sufficient generations for convergence of all group size posteriors, with final ESS values of 200 or above. The results of these analyses are tabulated in Supplemental Table 5 (based on original radiocarbon, not calendar dates, to-be-adjusted) and depicted in Figure 3.

3.5 RESULTS & DISCUSSION

We screened 234 DNA extracts from mammoth specimens found in low latitudes and 42 found in high latitudes with a quantitative PCR assay targeting a 49bp locus of the *12S* rRNA mitochondrial gene. Of the lower latitude set, 78 (33%) amplified in at least two replicates from at least one attempted extraction, while 37 (88%) of the northern specimens amplified in duplicate attempts. We converted most specimens with consistent positive amplification to Illumina sequencing libraries and screened them again. Except for three that were sequenced with a shotgun approach, those that consistently amplified for the target were then enriched for the complete mitochondrial genome in one or two rounds of in-solution hybridization capture . From this set we generated 67 complete or near-complete mitogenome sequences (>90% of the genome covered to at least 3x unique read depth) as well as more than a dozen partial mitogenomes (10% to 90% covered to at least 3x unique depth). 57 of these are from mammoths found at sites south of 51 degrees north latitude. We combined these with previously-published complete mammoth mitogenomes (Enk *et al.* 2011; Gilbert *et al.* 2008; Krause *et al.* 2006; Rogaev *et al.* 2006), making a final dataset of 87 complete mammoth mitogenome sequences, 68 of which have finite direct AMS radiocarbon dates. Localities represented by this combined dataset are depicted in Figure 1, with other sample information and experimental results available in Supplemental Tables 1 and 3.

3.5.1 Mitogenome phylogeography

Using the full set of complete mitochondrial genomes and an Asian elephant (*Elephas maximus*) mitogenome sequence as an outgroup, we generated a maximum likelihood tree with 100 bootstrap iterations, and further estimated its Bayesian posterior probability support. The resulting phylogeny is depicted in Figure 2, with the morphology-based nominal taxonomic designations of each specimen indicated along with their geographic origin.

The woolly mammoth portion of the revealed matriline topology is largely consistent with previous estimations using both short cytochrome *b*–HVR sequences (Barnes *et al.* 2007; Debruyne *et al.* 2008; Palkopoulou *et al.* 2013) and complete mitochondrial genomes (Enk *et al.* 2011; Gilbert *et al.* 2008), indicating at least three highly divergent matrilines (Clades I, II, and III). The mitogenomes of all non-woolly mammoths sampled here fall within Clade I, with most sharing a most recent common ancestor with the majority of attributed woolly mammoth samples from North America. For ease of discussion, we name a new haplogroup (F) that comprises the majority of the mitogenomes within this predominantly southern group, excluding only the Columbian and indeterminate mammoth specimens from Bindloss, Alberta and San Antonio Creek, California, which stem from a more basal position in the Clade I phylogeny. When we generated an ML tree using several additional partial mitochondrial genomes we reconstructed, including those of a pygmy mammoth and an unallocated specimen from the coast of Washington, we find that this basal root of the Californian lineage is not likely to be spurious (Supplemental Figure 1). This feature and the high overall diversity of North American lineages is consistent with the 'Out of America' (Debruyne *et al.* 2008) model, which suggests that the extremely widespread and last-surviving Old World woolly mammoth matrilines, haplogroups D and E, descend from a North American female ancestor.

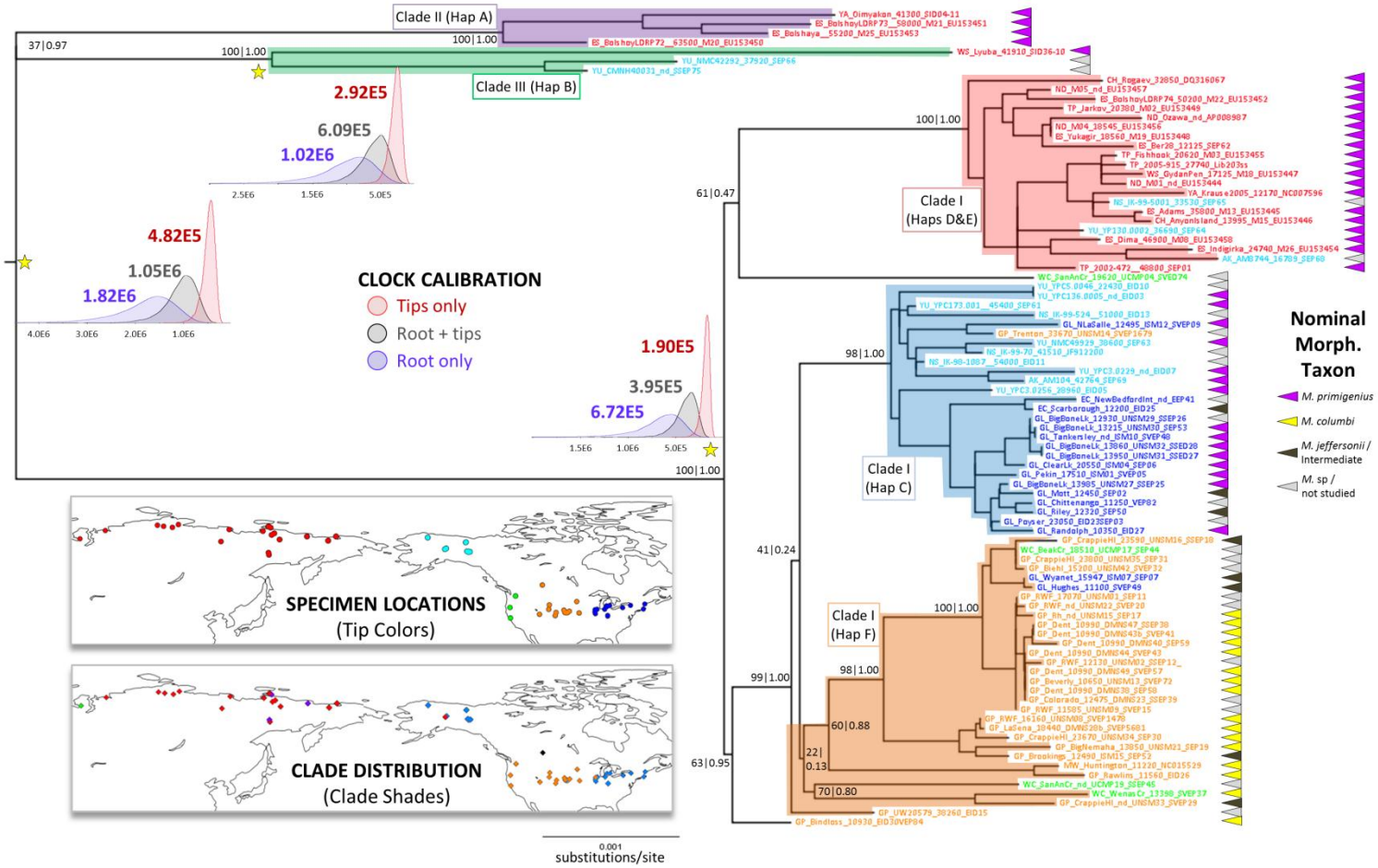


Figure 2: Maximum Likelihood tree with spatial distribution of samples and clades, and posterior distributions of estimated times to most recent common ancestry for three nodes. Tree generated with a 100-bootstrap analysis in IQ-TREE using the nucleotide substitution model, alpha and invariant site proportion parameters chosen by the corrected Aikake Information Criterion (AICc) in jModelTest. Nodal support was further assessed with a Bayesian coalescent approach in BEAST. Bootstrap and posterior probability support are depicted as (bootstrap)|(posterior). Tip names follow those in Supplemental Table 1A and are color-coded by location, which are depicted on the upper map. Clades discussed in the text are shaded by color and their distribution is depicted on the lower map. Posterior distributions of times to most recent common ancestry estimated under three different temporal calibration schemes are depicted, with mean values indicated.

Among the exclusively North American constituents of Clade I, there is clear geographic and morphological structure in mitogenome relatedness. Haplogroup C, previously identified in Beringian woolly mammoths, extends well into the southern latitudes, but is largely restricted to what were periglacial regions during the Pleistocene (the Great Lakes and northeast coast), which are regions where woolly mammoths are the most abundant taxon identified paleontologically (Fisher 2009). Haplogroup F is largely restricted to the Great Plains, mountain west and west coast, again consistent with the paleontological predominance of Columbian mammoths in these regions. However, mitogenomic relatedness and region of occurrence do not always correspond. For instance, the Hughes and Wyanet mammoths were found in periglacial regions but have mitogenomes more typical of Great Plains groups. Interestingly, both of these have been taxonomically identified as Jeffersonian mammoths due to dental and/or cranial traits that are essentially intermediate between *M. primigenius* or *M. columbi* (Saunders et al. 2010; Skeels 1962). These and the occurrences of several other morphologically intermediate animals in haplogroups C and F are consistent with the notion that gene flow occurred between the two dominant morphological taxa, probably at savanna-steppe ecotones, and that this sometimes manifested mitogenomically.

3.5.2 Chronology and mammoth biogeography

Estimated times of most recent common ancestry (tMRCAs) in mammoth phylogeny have been used to infer the chronology of their population processes. Early studies used only the paleontologically-estimated split between *Elephas* and *Mammuthus* or even deeper paleontological divergences to calibrate the mammoth molecular clock (Ozawa *et al.* 1997; Rogaev *et al.* 2006; Rohland *et al.* 2007). More recent studies use Bayesian coalescent-based approaches that take into account the absolute ages of the samples ("tip dates") themselves (Barnes *et al.* 2007; Debruyne *et al.* 2008; Gilbert *et al.* 2008; Palkopoulou *et al.* 2013). However, different groups have used tip dates in different ways, based in part on the assumed taxonomic identity of the reconstructed nodes. In some studies, only tip dates have been used to estimate the age of nodes deeper in the tree (Barnes *et al.* 2007; Palkopoulou *et al.* 2013); in others, both the *Elephas-Mammuthus* split and the tip dates have been combined to calibrate the molecular clock (Debruyne *et al.* 2008; Gilbert *et al.* 2008). Each approach returns very different chronologies of cladogenesis, and consequently very different inferences about both the timing and taxonomic contexts of corresponding population processes, and thereby the climatic events to which those may or may not be related.

To illustrate the impact that each calibration method has on the inferred cladogenic chronology, we used BEAST to estimate the ages of various nodes using all three calibration schemes, restricting the analysis to specimens with finite radiocarbon ages. In Figure 2 we include the marginal posterior densities of the estimated dates for three key nodes in the phylogeny. Unsurprisingly our estimations are similar to those of other studies with similar calibration schemes. Which approach is 'correct', however, is not immediately clear. The apparent time-dependency of molecular rates (Ho *et al.* 2005; Ho *et al.* 2007) is potentially influenced not only by actual biological and demographic processes [such as purifying selection and variation in generation times, though see Woodhams (2006)] but also by the sampling schemes used, both in terms of the number of samples and information content of the DNA sequences analyzed (Bandelt 2008; Debruyne & Poinar 2009; Emerson 2007). Our date randomization tests using tip dates alone (Supplemental Figure 2) suggest a strong temporal signal in our data. However, estimations of deeper tMRCAs using tip dates alone are very sensitive to the distribution of the tips in time. For example, when we include the mitogenome of a mammoth that is stratigraphically considered at least 200 thousand years old (CMNH40031, from the Old Crow River Basin) in a tips-only analysis, it increases the mean nodal date estimates in Figure 1 by as much as a factor of 1.34 (+~28 thousand years for tMRCA of Clade I; 99ky for Clade III, and 107ky for all mammoths). We suspect that additional deep temporal sampling would push these dates back even further when using a tips-only calibration scheme. By the same token this suggests that older samples will be necessary to firmly date these events.

One of the primary goals of this study was to determine whether the apparently close affinity of Columbian mammoths to North American woolly mammoths was typical rather than exceptional (Enk *et al.* 2011). Here we present good evidence that it is the former, which we think significantly impacts both the taxonomic and chronological interpretation of mammoth mitochondrial phylogeny reconstructed to date. Given the consensus paleontological model that Columbian mammoths descend from an Early to Middle Pleistocene immigrant to North America (Agenbroad 2005; Lister & Bahn 2007), and the pattern of diversity that points to a North American origin of Clade I, we find that the most parsimonious reconciliation of the genetic evidence with the paleontological record requires the conclusion that the mitochondrial tMRCA is much more ancient than a tips-only calibration scheme would estimate. As depicted in Figure 3A, we suspect that the divergence of Clade I from Clades II+III occurred near the time when *M. trogontherii* entered North America, with the independent trajectories of Clade I and Clades II+III corresponding to the allopatric speciation of *M. primigenius* and *M. columbi*. This necessarily implies that the morphological identity of Clade I mammoths as *M. primigenius* is the consequence of an introgression event (or events) that introduced woolly mammoth morphology into North America, but left little matrilineal evidence. It is possible that this morphology was carried into North America by Clade III mammoths, which were nearly as widespread as last-surviving Haplogroups D and E of Clade I. However, despite the clear influence of sampling scheme on tips-only dating, we cannot completely rule out the more recent chronologies suggested by that calibration scheme. Thus it is possible that the close mitogenomic affinity to woolly mammoths of the Columbian and other southern mammoths we sampled here are the consequence of a massive replacement event (Figure 3B). In this scheme, which was emphasized by a previous investigation (Enk *et al.* 2011), the ancestral Columbian mammoth matriline remains unsampled by our dataset, but Columbian mammoth morphology was retained through invading woolly mammoths interbreeding with resident descendents of *M. trogontherii*. More thorough sampling of older mammoths from areas further from the Laurentide ice margin may yet uncover such a matriline. However, assuming it existed and is eventually detected, it would have to be carefully ascertained whether it descended from *M. trogontherii* or from an earlier immigrant like Early Pleistocene *M. meridionalis*. This latter species has been detected at some localities, and there is evidence that it may have co-occurred with more advanced forms (McDaniel & Jefferson 2003).

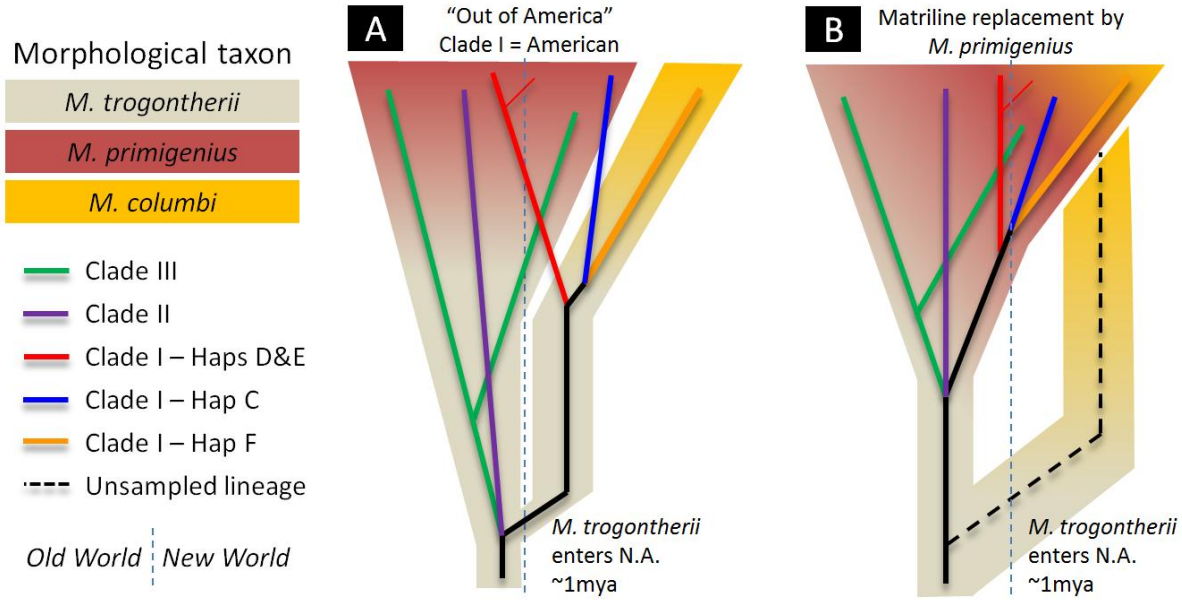


Figure 3: Species/morphological tree schematics. Schematic versions of the estimated mitochondrial phylogeny are overlain on species/morphological trees. One version (A) is consistent with the chronology estimated with a root+tip calibration scheme and a paleontological model where *M. columbi* descends from North American *M. trogontherii*. The other (B) is more consistent with a tip-only calibration scheme and assumes that the reconstructed ancestors in the phylogeny are taxonomically *M. primigenius*. Both models assume nuclear gene interchange between divergent lineages that resulted in morphological similarity; convergence is not entertained.

3.5.3 Population size dynamics

In previous investigations, reconstructed population size trajectories of subarctic woolly mammoths indicated that they experienced relatively constant size throughout the Pleistocene on both sides of the Bering Strait (Debruyne *et al.* 2008). In reconstructions that include Holocene sequences from Wrangel Island, trajectories indicate an unsurprising decline in population size shortly before continental mammoths go extinct (Nystrom *et al.* 2010; Palkopoulou *et al.* 2013). In order to explore how adding our southern mitogenomes impacts this impression, we reconstructed Bayesian Skygrids (Gill *et al.* 2013) using the complete dataset of mammoth mitogenomes, as well as for three geographic or ecological paired partitions under two temporal calibration schemes (Figure 4 for tips-only, Supplemental Figure 3 for node + tips analysis.)

Though the highest posterior densities (HPDs) are exceptionally broad for our reconstructed trajectories, there is some indication that mammoth population size increased prior to an end-Pleistocene decline. Comparing paired regional/ecological partitions shows that it is primarily the new world populations that drive this detected increase, while the Old World and periglacial groups remain relatively constant in size. Interestingly this increase in the southern portion is roughly coincident with a slight (and potentially illusory) decline in northern populations, suggesting complex responses of mammoths to the same global climate background. For most temporal calibration and partitioning schemes, this increase in southern populations/decrease in northern populations apparently follows the Sangamonian (Eemian or Riss-Würm) interglacial stage, which was followed by the Wisconsinan glacial stage prior to the Holocene. The southern partitions also apparently drive the detected decline following this increase, which is consistent with the inferred extinction chronology in Beringia versus continental North America (Agenbroad 2005; Stuart *et al.* 2002).

However, in addition to having broad HPDs, we find that both the shape and chronology of the reconstructed trajectories are likely biased by a number of variables, and thus should be interpreted with caution. Not only do the Skygrid projections reflect the nodal topology of the tree (i.e., a long branch leading to Clade I followed by widespread cladogenesis) but also the distribution of radiocarbon dates, at least qualitatively. Where the sampling distribution and estimated population sizes do not correspond, samples come from the same location and are closely related (e.g., Dent, Colorado and Big Bone Lick, Kentucky). As such more thorough and even sampling across both time and space may uncover less (or more) distinct population histories in the populations sampled here. In addition, and as alluded to previously, in some cases the partitioning scheme has a substantial impact on the chronology of the trajectories (supplemental figure 3, trajectories from high vs. low-latitude or periglacial vs. temperate schemes).

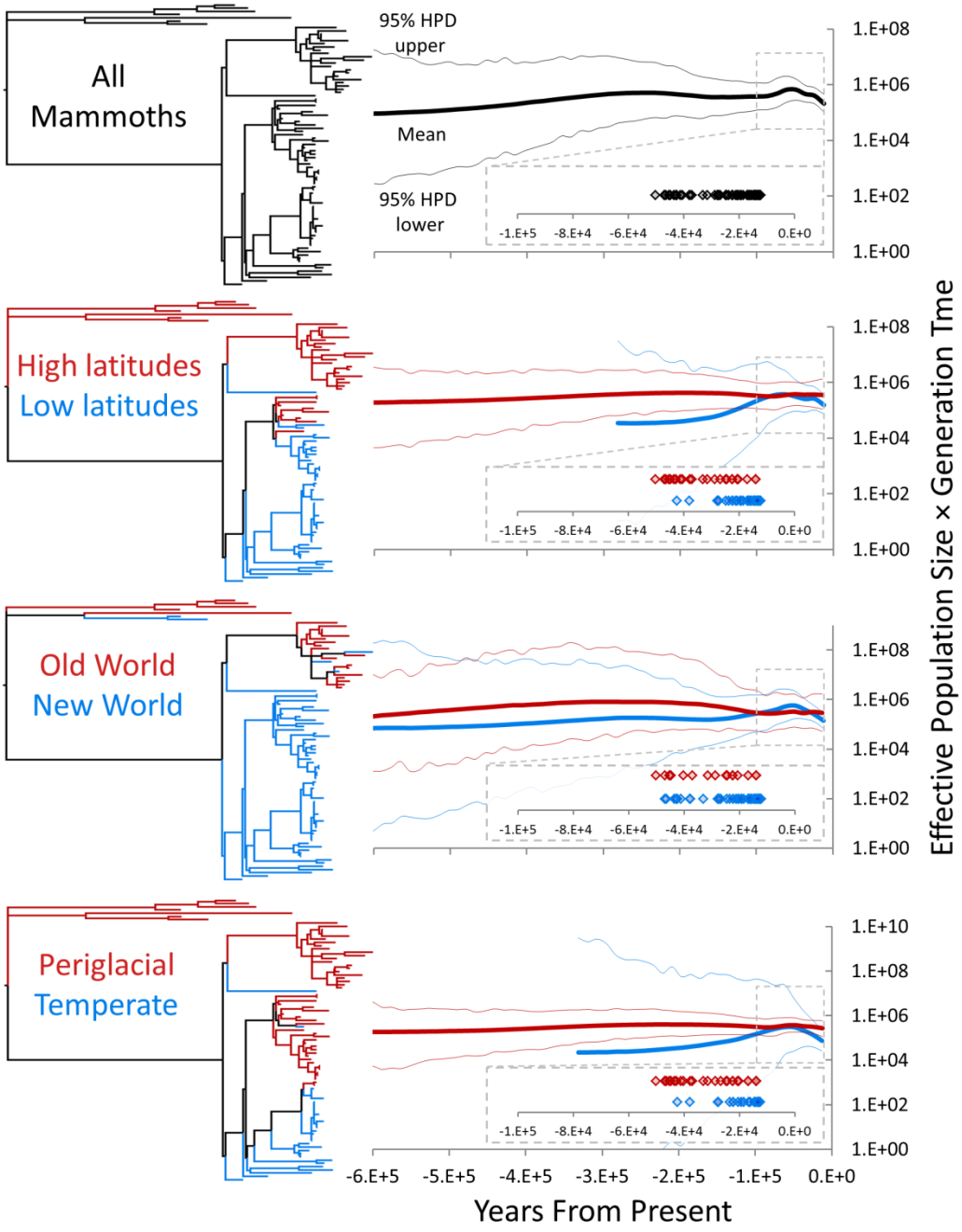


Figure 4: Bayesian Skygrid analyses of the whole dataset and under three paired partition schemes. Each projection was estimated in BEAST using the partition-specific best model indicated by the AICc in jModelTest. Trees adjacent to each Skygrid are color-coded by which specimens were included in the individual projections. "High" latitude samples are those from the Old World and Alaska and the Yukon, while "low" latitude samples are those from the Great Lakes, east coast, Great Plains, mountain west and west coast. "Periglacial" samples are those in high latitudes as well as the Great Lakes and east coast, contexts nearby the Laurentide ice sheet during the Late Pleistocene, while "temperate" specimens are the rest. All Skygrid population sizes are associated with posterior effective sample sizes > 200.

3.6 CONCLUSIONS

The patterns of cladogenesis and mitogenomic diversity documented here point to a complex genetic history of mammoths in North America, one likely characterized by bouts of limited local divergence followed by the introgression of foreign morphologies. By confirming that southern mammoth taxa most likely descend from the same matrilines that were previously observed in woolly mammoths, we provide evidence supporting an ancient chronology for both our and previous mammoth mitogenomic phylogeny. There is also some preliminary evidence that mammoths in the north and south underwent different population size trajectories during the Late Pleistocene, implying that the periarctic history of megafaunal population dynamics are unlikely to serve as a proxy for their dynamics worldwide. While geographically broader and temporally deeper sampling of mitogenomes from North America could help clarify the exact biogeographic processes underlying mammoth evolution, ultimately nuclear sequence data will allow us to directly test specific introgression scenarios, and perhaps improved estimating the timing of various divergence and introgression events. Fortunately, extraction and library preparation of very degraded DNA (Dabney *et al.* 2013; Meyer *et al.* 2014; Orlando *et al.* 2013), targeted DNA capture (Carpenter *et al.* 2013; Wagner *et al.* 2014), and ancient DNA bioinformatic analysis (Schubert *et al.* 2012) have experienced tremendous improvement in recent years. As these technologies continue to improve, we can expect to see genetic data from more specimens preserved in less DNA-friendly taphonomic contexts like the ones analyzed here, and as such more thoroughly gauge their responses to and effects on the now-extinct Pleistocene environment.

3.7 DATA AVAILABILITY

Raw sequence data will be uploaded to the NCBI Sequence Read Archive (SRA) and consensus to NCBI GenBank upon acceptance for review.

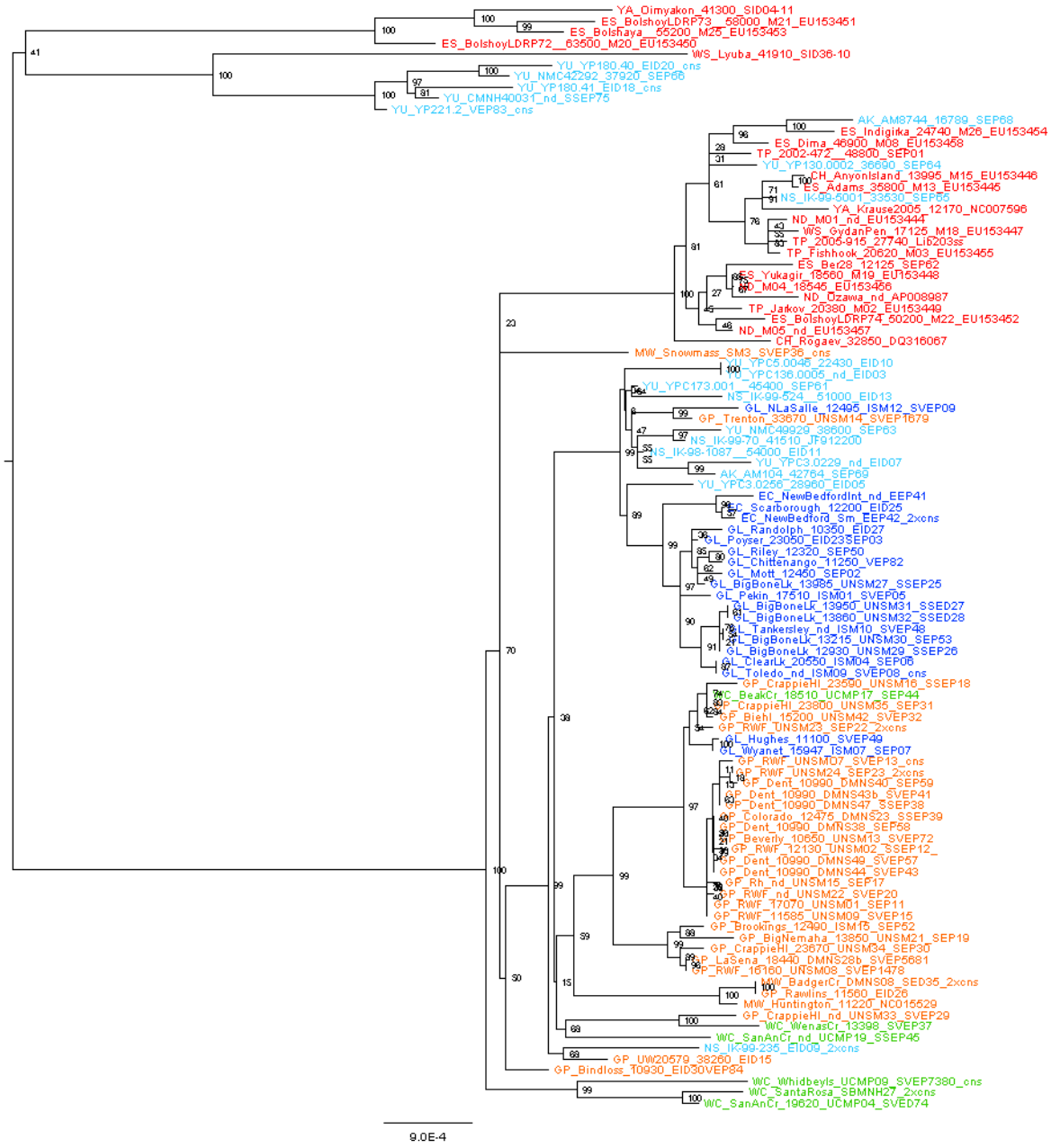
3.8 ACKNOWLEDGMENTS

Thanks especially to sample providers, including Jim Mead of Northern Arizona University, Bernard Buigues of Cerpoplex, Dale Guthrie and Dick Harington of the Canadian Museum of Nature, Ross Secord and George Corner of the University of Nebraska State Museum, Richard Stucky and Steve Holen of the Denver Museum of Nature and Science, Pat Holroyd and Katilin McGuire of the University of California Museum of Paleontology, Paul Collins of the Santa Barbara Museum of Natural History, Pat Lubinski of Central Washington University, Mark Clementz of the University of Wyoming, Kathy Kavanagh of the University of Massachusetts at Dartmouth, and Donald Benson of New Bedford, Massachusetts. Special thanks to Nadin Rohland and David Reich of the Harvard Medical School for assistance in sequencing specimens Oimyakon and Lyuba. Thanks to Christine King, Leanne Blanchard and Lilie De Sousa of the Farncombe Metagenomics Facility at McMaster Children's Hospital for assistance in library quality control and sequencing. Thanks to Regis Debruyne, Carsten Schwartz, Hariharan Manogaralingam and Jocelyn Ip for assistance with laboratory work. This work was supported by generous contributions from MYcroarray, and an NSERC Discovery and Accelerator Grant and CRC to HP.

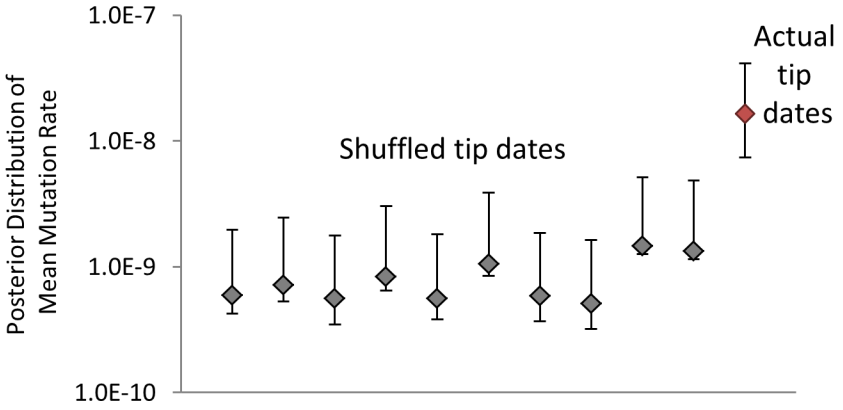
3.9 AUTHOR CONTRIBUTIONS

JE and HP conceived of and designed the study. JE and AD performed laboratory work. PS and JS performed AMS radiocarbon dating. CW, JS, and DF performed morphological taxon identifications. JE analyzed the data. CW, JS, JMR, GZ, DF, RM and DF assisted in data interpretation. JE wrote the manuscript with assistance from all authors.

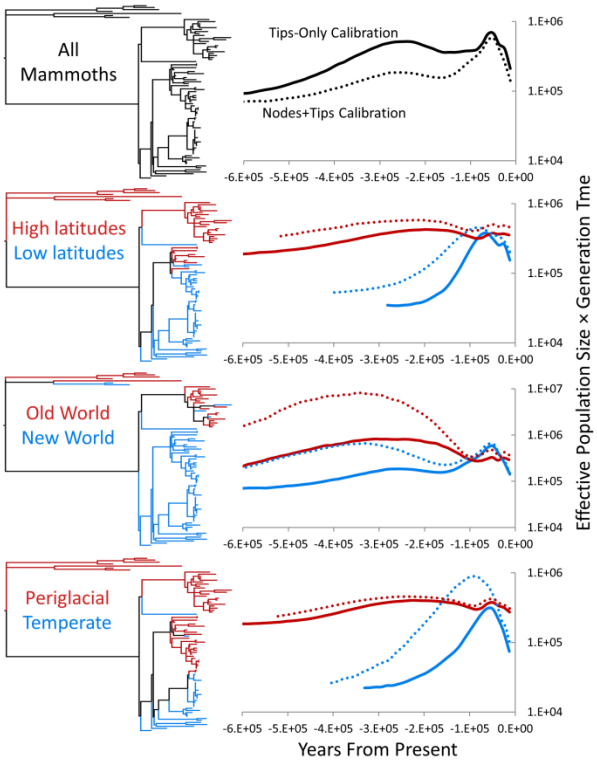
3.10 SUPPLEMENTAL FIGURES



Supplemental Figure 1: ML tree with partial mitochondrial genome consensuses included. Tips ending in "cns" are consensus called for bases covered by at least 3 unique reads, while tips ending in "2xcns" were called for bases covered by at least 2 unique reads. Tips are again colored by region of origin according to the depiction in Figure 2 in the main text. Nodal bootstrap values are depicted.



Supplemental Figure 2: Posterior distributions of the estimated mean mutation rate after date randomization. Diamonds correspond to the mean posterior probability, while error bars correspond to the 95% highest posterior densities.



Supplemental Figure 3: Mean Skygrid population sizes for two temporal calibration schemes. Solid line trajectories depict mean population size when using a tips-only calibration scheme. Dotted lines depict mean population size when using a nodes + tips calibration scheme, where the internal nodes were given priors corresponding to the posterior probability-estimated ages derived from the root + tip tMRCA estimation analysis.

3.11 SUPPLEMENTAL TABLES

Supplemental Table 1A: Mammoth specimens from which mitogenomes were sequenced in this study. Cells highlighted in orange designate radiocarbon date measurements with a C:N ratio of >3.00, which may indicate carbon contamination.

ML Tree Tip Prefix	MADC#	Museum Catalog# / Name	Field #	St/Prv/Reg	Location/Locality	Lat.	Long.	Locality reference	Species	AMS Lab #	AMS ^{14}C Date	Sigma	Date reference	Median Cal. yBP	Mito Status
TP_2002-472	2002/472	-	-	SI	Taimyr Peninsula	74.42	107.75	(Debruyne *et al.* 2008)	*M. primigenius*	UCIAMS38677	>48800	na	(Debruyne *et al.* 2008)	-	Complete
TP_2005-915	2005/915	-	-	SI	Taimyr Peninsula	73.75	102.00	(Debruyne *et al.* 2008)	*M. primigenius*	Beta-210777	27740	220	(Debruyne *et al.* 2008)	31501	Complete
AK_AM104	AM104	-	-	AK	Cleary Creek	65.17	-147.50	(Debruyne *et al.* 2008)	*M. primigenius*	AA14906	42764	1737	(Debruyne *et al.* 2008)	46455	Complete
AK_AM8744	AM8744	-	-	AK	Ester Creek	64.83	-148.00	(Debruyne *et al.* 2008)	*M. primigenius*	AA14896	16789	108	(Debruyne *et al.* 2008)	20256	Complete
ES_Ber28	Ber28	-	-	SI	Berelekh	70.40	143.95	(Debruyne *et al.* 2008)	*M. primigenius*	UCIAMS38670	12125	30	(Debruyne *et al.* 2008)	14011	Complete
GL_Chittenango	HP1133	Chittenango	-	NY	Near Chittenango	43.05	-75.87	guess, (Feranec & Kozlowski 2012)	*M.* sp	NOSAMS OS-93430	11250	65	(Feranec & Kozlowski 2012)	13116	Complete
YU_CMNH40031	CMNH40031	-	-	YU	Old Crow	68.06	-139.78	approximate	*M.* sp	nd	-	-	-	-	Complete
GP_Colorado	DMNS23	1473	-	CO	nd	39.07	-105.13	middle of Colorado	*M.* sp	UCIAMS131692	12475	40	-	14661	Complete
GP_LaSena	DMNS28b	La Sena	-	NE	La Sena	40.38	-100.23	(Holen 2006)	*M. columbi*	AA-6972	18440	145	(Holen 2006)	22298	Complete
GP_Dent	DMNS38	3995	-	CO	Dent	40.30	-104.80	(Surovell & Waguespack 2008)	*M. columbi*	Average of 3	10990	25	(Waters & Stafford 2007)	12827	Complete
GP_Dent	DMNS40	1899	-	CO	Dent	40.30	-104.80	(Surovell & Waguespack 2008)	*M. columbi*	Average of 3	10990	25	(Waters & Stafford 2007)	12827	Complete
GP_Dent	DMNS43b	1893-198	-	CO	Dent	40.30	-104.80	(Surovell & Waguespack 2008)	*M. columbi*	Average of 3	10990	25	(Waters & Stafford 2007)	12827	Complete
GP_Dent	DMNS44	3802	-	CO	Dent	40.30	-104.80	(Surovell & Waguespack 2008)	*M. columbi*	Average of 3	10990	25	(Waters & Stafford 2007)	12827	Complete
GP_Dent	DMNS47	3810	-	CO	Dent	40.30	-104.80	(Surovell & Waguespack 2008)	*M. columbi*	Average of 3	10990	25	(Waters & Stafford 2007)	12827	Complete
GP_Dent	DMNS49	1636	-	CO	Dent	40.30	-104.80	(Surovell & Waguespack 2008)	*M. columbi*	Average of 3	10990	25	(Waters & Stafford 2007)	12827	Complete
NS_IK-98-1087	IK-98-1087	-	-	AK	Upper Ikpikpuk R., North Slope	69.37	-154.67	(Debruyne *et al.* 2008)	*M.* sp	CAMS 91795	>54000	na	(Debruyne *et al.* 2008)	-	Complete
NS_IK-99-5001	IK-99-5001	-	-	AK	Upper Ikpikpuk R., North Slope	69.37	-154.67	(Debruyne *et al.* 2008)	*M.* sp	CAMS 91968	33530	340	(Debruyne *et al.* 2008)	37856	Complete
NS_IK-99-524	IK-99-524	-	-	AK	Upper Ikpikpuk R., North Slope	69.37	-154.67	(Debruyne *et al.* 2008)	*M.* sp	CAMS 91811	>51000	na	(Debruyne *et al.* 2008)	-	Complete
GL_Pekin	ISM01	497399	-	IL	Near Pekin	40.52	-89.72	This study	*M. primigenius*	UCIAMS131694	17510	70	This study; C:N >3.00	21151	Complete
GL_ClearLk	ISM04		-	IL	Gravel Pit near Clear Lake	39.82	-89.53	(Parmalee 1967)	*M. primigenius*	NZA 32590	20550	100	This study	24745	Complete
GL_Wyanet	ISM07	494012	-	IL	Wyanet	41.37	-89.65	(Saunders *et al.* 2010)	*M. jeffersonii*	NZA 28851	15947	60	(Saunders *et al.* 2010)	19237	Complete
GL_Tankersley	ISM10	Tankersley	-	OH	Near Cleves	39.15	-84.75	approximate, Tankersley pers. comm.	*M.* sp	nd	-	-	-	-	Complete
GL_NLaSalle	ISM12	408764	-	IL	North LaSalle County	41.55	-88.87	(Parmalee 1967)	*M.* sp	UCIAMS131695	12495	45	This study	14724	Complete
GP_Brookings	ISM15	-	-	SD	Near Brookings	44.46	-96.88	This study	*M.* sp (intermediate)	CURL8895	12490	35	(Mandel 2004)	14716	Complete
WS_Lyuba	HP1095	Lyuba	-	RU	Yuribei R., Yamal Peninsula	68.90	69.50	approximate, (Fisher *et al.* 2012)	*M. cf primigenius*	GrA-41246	41910	(+550-450)	(Fisher *et al.* 2012)	45294	Complete
GL_Mott	HP1135	Mott	-	MI	Clayton Township	42.97	-83.97	http://www.mcc.edu/science_math/museum/exhibits.html	*M.* sp	UCIAMS131696	12450	40	This study	14585	Complete
EC_NewBedfodInt	HP2133		-	MA	New Bedford Sound	41.62	-70.88	approximate	*M.* sp	nd	-	-	-	-	Complete
GP_Bindloss	NMC17845	Bindloss	-	AB	Bindloss	50.95	-110.13	(Hills & Harington 2003)	*M. columbi*	TO-8514	10930	100	(Hills & Harington 2003)	12825	Complete
YU_NMC42292	NMC-42292	-	-	YU	Dawson Area	64.05	-139.42	(Debruyne *et al.* 2008)	*M. primigenius*	AA17535	37920	2700	(Debruyne *et al.* 2008)	42815	Complete
YU_NMC49929	NMC-49929	-	-	YU	Dawson Area	64.05	-139.42	(Debruyne *et al.* 2008)	*M. primigenius*	AA17553	38600	2900	(Debruyne *et al.* 2008)	43496	Complete
YA_Oimyakon	2006/001-2	Oimyakon	-	SI	Yakutia	63.50	142.75	(Debruyne *et al.* 2008)	*M. primigenius*	GrA-30727	41300	(+900-650)	(Debruyne *et al.* 2008)	44806	Complete
GL_Poyser	HP1138	Poyser	-	IN	Near Goshen	41.55	-85.93	Sand & gravel pit near Goshen, Indiana	*M.* sp	BETA #92870	23050	180	This study	27364	Complete
GL_Randolph	HP1134	Randolph	-	NY	East Randolph Fish Hatchery	42.15	-78.93	(Feranec & Kozlowski 2012)	*M. primigenius*	NOSAMS OS-93354	10350	45	(Feranec & Kozlowski 2012)	12210	Complete
GP_Rawlins	UW6368	Rawlins	-	WY	Near Rawlins	41.50	-107.63	approximate, (Haynes *et al.* 2013)	*M. columbi*	Average of 4	11560	60	(Haynes *et al.* 2013)	13392	Complete
GL_Riley	HP1726	Riley	-	MI	Morrison Lake Country Club, Saranac	42.87	-85.20	appx., Lake Creek	*M.* sp (intermediate)	BETA #282797	12320	50	This study	1949	Complete

Supplemental Table 1A (Continued): Mammoth specimens from which mitogenomes were sequenced in this study. Cells highlighted in orange designate radiocarbon date measurements with a C:N ratio of >3.00, which may indicate carbon contamination.

ML Tree Tip Prefix	MADC#	Museum Catalog# / Name	Field #	St/Prv/Reg	Location/Locality	Lat.	Long.	Locality reference	Species	AMS Lab #	AMS ^{14}C Date	Sigma	Date reference	Median Calendar yBP	Mito Status
EC_Scarborough	HP1137	Scarborough	-	ME	Near Scarborough	43.58	-70.32	approximate, (Hoyle *et al.* 2004)	*M.* sp (intermediate)	OS-5636	12200	55	(Hoyle *et al.* 2004)	14091	Complete
WC_SanAnCr	UCMP04	29171	-	CA	San Antonio Creek	38.16	-122.53	guess	*M.* sp	UCIAMS131697	19620	120	This study; C:N >3.00	23642	Complete
WC_BeakCr	UCMP17	21284	-	OR	Beak (Bear?) Creek	42.22	-122.72	approximate; "Bear" Creek is near Ashland, home of discoverer	*M.* sp	UCIAMS131699	18510	100	This study; C:N >3.00	22387	Complete
WC_SanAnCr	UCMP19	29170	-	CA	San Antonio Creek	38.16	-122.53	guess	*M.* sp	nd	-	-	-	-	Complete
GL_Hughes	UM22798	Hughes	-	MI	North of Assyria	42.52	-85.12	approximate, (Skeels 1962)	*M. jeffersonii*	UCIAMS131693	11100	35	This study	12985	Complete
GP_RWF	UNSM01	49826	1122-83	NE	Red Willow Fauna, Rw-102	40.22	-100.37	approximate, (Corner 1977)	*M. sp*	UCIAMS131700	17070	70	This study	20590	Complete
GP_RWF	UNSM02	49736	342-82	NE	Red Willow Fauna, Rw-102	40.22	-100.37	approximate, (Corner 1977)	*M. sp*	UCIAMS131701	12130	35	This study	14014	Complete
GP_RWF	UNSM08	2022	9-13-6-33	NE	Red Willow Fauna, Rw-102	40.22	-100.37	approximate, (Corner 1977)	*M. columbi*	UCIAMS131702	16160	80	This study; C:N >3.00	19502	Complete
GP_RWF	UNSM09	49414	1155-81	NE	Little Sand Pit near McCook, Rw-110	40.20	-100.50	approximate, Corner pers. Comm	*M. sp*	UCIAMS131703	11585	35	This study	13420	Complete
GP_Beverly	UNSM13	2125	24-3-31	NE	Beverly Gravel Pits	40.30	-101.03	approximate	*M. columbi*	UCIAMS131704	10650	30	This study	12633	Complete
GP_Trenton	UNSM14	1811	2019-74	NE	Trenton Reservoir	40.17	-101.07	approximate, Corner pers. Comm	*M. sp*	UCIAMS131705	33670	450	This study	37997	Complete
GP_Rh	UNSM15	1785	4000-75	NE	Richardson Co, Big Nemaha R.	40.12	-95.87	approximate, Corner pers. Comm	*M. columbi*	nd	-	-	-	-	Complete
GP_CrappieHl	UNSM16	49033	6001-007	NE	Crappie Hole	41.20	-101.75	approximate, (Voorhies & Corner 1984)	*M.* sp (intermediate)	UCIAMS131706	23590	130	This study	27713	Complete
GP_BigNemaha	UNSM21	88535	5001-96	NE	South Fork Big Nemaha R.	40.07	-95.82	guess	*M. columbi*	UCIAMS131707	13850	45	This study	16770	Complete
GP_RWF	UNSM22	1861	1095-65	NE	Palisade Sand Pit	40.35	-101.42	approximate, Corner pers. Comm	*M.* sp	nd	-	-	-	-	Complete
GL_BigBoneLk	UNSM27	-	18S-5E	KY	Big Bone Lick	38.88	-84.75		*M. primigenius*	UCIAMS131708	13985	45	This study	16982	Complete
GL_BigBoneLk	UNSM29	-	3704-66	KY	Big Bone Lick	38.88	-84.75		*M. primigenius*	UCIAMS131709	12930	40	This study	15445	Complete
GL_BigBoneLk	UNSM30	-	4287-63	KY	Big Bone Lick	38.88	-84.75		*M. primigenius*	UCIAMS131710	13215	40	This study	15884	Complete
GL_BigBoneLk	UNSM31	-	18S-5E 4.2BBM	KY	Big Bone Lick	38.88	-84.75		*M. primigenius*	UCIAMS131711	13950	45	This study	16926	Complete
GL_BigBoneLk	UNSM32	-	4084-63	KY	Big Bone Lick	38.88	-84.75		*M. primigenius*	UCIAMS131712	13860	40	This study	16786	Complete
GP_CrappieHl	UNSM33	-	1335-87	NE	Crappie Hole	41.20	-101.75	approximate, (Voorhies & Corner 1984)	*M.* sp (intermediate)	nd	-	-	-	-	Complete
GP_CrappieHl	UNSM34	-	1185-82	NE	Crappie Hole	41.20	-101.75	approximate, (Voorhies & Corner 1984)	*M. columbi*	UCIAMS131713	23670	190	This study	27778	Complete
GP_CrappieHl	UNSM35	91547	1210-76	NE	Crappie Hole	41.20	-101.75	approximate, (Voorhies & Corner 1984)	*M.* sp	UCIAMS131714	23800	140	This study	27863	Complete
GP_Biehl	UNSM42	1505	1096-73	NE	Biehl Farm	40.80	-99.65	guess, Biehl Cattle Company	*M.* sp	UCIAMS131715	15200	60	This study; C:N >3.00	18470	Complete
GP_UW20579	UW20579	-	-	WY	nd	43.22	-107.43	middle of Wyoming	*M.* sp	UCIAMS131716	38260	790	This study; C:N >3.00	42451	Complete
WC_WenasCr	WAST_01	-	-	WA	Wenas Creek	46.70	-120.55	approximate, (Lubinski *et al.* 2009)	*M.* sp	WK-18064	13398	58	(Lubinski *et al.* 2009)	16122	Complete
YU_YPC130.0002	YPC130.0002	-	-	YU	Quartz Creek	63.82	-139.03	(Debruyne *et al.* 2008)	*M.* sp	UCIAMS39891	36690	810	(Debruyne *et al.* 2008)	41230	Complete
YU_YPC136.0005	YPC136.0005	-	-	YU	Sulphur Creek	63.73	-138.83	(Debruyne *et al.* 2008)	*M. primigenius*	nd	-	-	-	-	Complete
YU_YPC173.001	YPC173.001	-	-	YU	Ch'ijee's Bluff	67.48	-139.92	(Debruyne *et al.* 2008)	*M. primigenius*	UCIAMS41492	>45400	na	(Debruyne *et al.* 2008)	-	Complete
YU_YPC3.0229	YPC3.0229	-	-	YU	Finning	63.83	-138.25	(Debruyne *et al.* 2008)	*M. primigenius*	nd	-	-	-	-	Complete
YU_YPC3.0256	YPC3.0256	-	-	YU	Finning	63.83	-138.25	(Debruyne *et al.* 2008)	*M. primigenius*	UCIAMS39115	28960	310	(Debruyne *et al.* 2008)	33125	Complete
YU_YPC5.0046	YPC5.0046	-	-	YU	Hunker Creek	63.98	-139.03	(Debruyne *et al.* 2008)	*M.* sp	UCIAMS41487	22430	140	(Debruyne *et al.* 2008)	26742	Complete

Supplemental Table 1A (Continued): Mammoth specimens from which mitogenomes were sequenced in this study. Cells highlighted in orange designate radiocarbon date measurements with a C:N ratio of >3.00, which may indicate carbon contamination.

ML Tree Tip Prefix	MADC#	Museum Catalog# / Name	Field #	St/Prv/Reg	Location/Locality	Lat.	Long.	Locality reference	Species	AMS Lab #	AMS 14C Date	Sigma	Date reference	Median Calendar yBP	Mito Status
MW_BadgerCr	DMNS08	4530	-	CO	Badger Creek	40.29	-106.45	guess	*M.* sp	nd	-	-	-	-	Partial
NS_IK-99-235	IK-99-235	-	-	AK	Upper Ikpikpuk R., North Slope	69.37	-154.67	(Debruyne *et al.* 2008)	*M.* sp	CAMS 91803	40870	820	(Debruyne *et al.* 2008)	46389	Partial
GL_Toledo	ISM09	497415	-	IL	Near Toledo	39.28	-88.23	This study	*M.* cf *primigenius*	nd	-	-	-	-	Partial
EC_NewBedfordSm	HP2134		-	MA	New Bedford Sound	41.62	-70.88	approximate	*M.* sp	nd	-	-	-	-	Partial
WC_Pymy	SBMNH27	999	4222	CA	Santa Rosa Island	33.97	-120.10	approximate	*M. exilis*	nd	-	-	-	-	Partial
MW_Snowmass	SM3	"Snowy"	-	CO	Snowmass Site	39.21	-106.93	approximate, (Pigati *et al.* 2014)	*M. columbi*	nd	-	-	-	-	Partial
WC_WhidbeyIs	UCMP09	21293	-	WA	Whidbey Island	48.12	-122.58	approximate	*M.* sp	UCIAMS131698	19200	120	This	25088	Partial
RWF	UNSM07	2099	505-46	NE	Red Willow Fauna, Rw-102	40.22	-100.37	approximate, (Corner 1977)	*M.* sp	nd	-	-	-	-	Partial
RWF	UNSM23	1827	1049-68	NE	Palisade Sand Pit	40.35	-101.42	approximate, Corner pers. comm	*M.* sp	nd	-	-	-	-	Partial
RWF	UNSM24	1829	1049-68	NE	Palisade Sand Pit	40.35	-101.42	approximate, Corner pers. comm	*M.* sp	nd	-	-	-	-	Partial
YU_YP180.40	YP180.40	-	-	YU	Old Crow; CRH-94	68.06	-139.78	approximate	*M.* sp	nd	-	-	-	-	Partial
YU_YP180.41	YP180.41	-	-	YU	Old Crow; CRH-94	68.06	-139.78	approximate	*M.* sp	nd	-	-	-	-	Partial
YU_YP221.2	YP221.2	-	-	YU	Old Crow; CRH-11A	67.84	-139.85	approximate	*M.* sp	nd	-	-	-	-	Partial

Supplemental Table 1B: Previously-published mitogenomes used in this study

ML Tree Tip Prefix	Ref. ID	GenBank ID	Field #	St/Prv/Reg	Location/Locality	Lat.	Long.	Locality reference	Species	AMS Lab #	AMS Date	Sigma	Date / mitogenome reference	Median Cal yBP	Mito Status
CH_AnyonIsland	M15	EU153446	-	SI	Chukotka	69.80	169.00	(Palkopoulou *et al.* 2013)	*M. primigenius*	OxA-17106	13995	55	(Gilbert *et al.* 2008)	16996	Complete
CH_Rogaev	-	DQ316067	-	SI	Chukotka	68.17	165.93	(Palkopoulou *et al.* 2013)	*M. primigenius*	MAG-1000	32850	900	(Rogaev *et al.* 2006)	37144	Complete
ES_Adams	M13	EU153445	-	SI	Northeast	72.50	127.50	(Palkopoulou *et al.* 2013)	*M. primigenius*	T-171	35800	1200	(Gilbert *et al.* 2008)	40418	Complete
ES_Bolshaya	M25	EU153453	-	SI	Novosibirsk Islands	69.79	157.70	(Palkopoulou *et al.* 2013)	*M. primigenius*	OxA-17113	>55200	-	(Gilbert *et al.* 2008)	-	Complete
ES_BolshoyLDRP72	M20	EU153450	-	SI	Novosibirsk Islands	73.64	142.89	(Palkopoulou *et al.* 2013)	*M. primigenius*	OxA-17109	>58000	-	(Gilbert *et al.* 2008)	-	Complete
ES_BolshoyLDRP73	M21	EU153451	-	SI	Novosibirsk Islands	73.21	143.60	(Palkopoulou *et al.* 2013)	*M. primigenius*	OxA-17110	>63500	-	(Gilbert *et al.* 2008)	-	Complete
ES_BolshoyLDRP74	M22	EU153452	-	SI	Novosibirsk Islands	73.64	142.67	(Palkopoulou *et al.* 2013)	*M. primigenius*	OxA-17111	50200	900	(Gilbert *et al.* 2008)	50304	Complete
ES_Dima	M08	EU153458	-	SI	Magadan	62.67	142.93	(Palkopoulou *et al.* 2013)	*M. primigenius*	OxA-17102	46900	700	(Gilbert *et al.* 2008)	46962	Complete
ES_Indigirka	M26	EU153454	-	SI	Indigirka	68.60	147.06	(Palkopoulou *et al.* 2013)	*M. primigenius*	OxA-17114	24740	110	(Gilbert *et al.* 2008)	28769	Complete
ES_Yukagir	M19	EU153448	-	SI	Yakutsk	71.87	140.58	(Palkopoulou *et al.* 2013)	*M. primigenius*	Avg of 3 GRNs	18560	50	(Gilbert *et al.* 2008)	22434	Complete
MW_Huntington	HUN.T.01	NC015529	-	UT	Huntington Reservoir	39.58	-111.25	approximate, (Gillette & Madsen 1993)	*M. columbi*	AA 4936	11220	110	(Gillette & Madsen 1993)	13082	Complete
ND_M01	M01	EU153444	-	SI	nd	-	-	(Palkopoulou *et al.* 2013)	*M. primigenius*	-	-	-	(Gilbert *et al.* 2008)	-	Complete
ND_M04	M04	EU153456	-	SI	nd	67.83	124.29	(Palkopoulou *et al.* 2013)	*M. primigenius*	OxA-17098	18545	70	(Gilbert *et al.* 2008)	22422	Complete
ND_M05	M05	EU153457	-	SI	nd	-	-	(Palkopoulou *et al.* 2013)	*M. primigenius*	-	-	-	(Gilbert *et al.* 2008)	-	Complete
ND_Ozawa	-	AP008987	-	OW	nd	-	-	(Palkopoulou *et al.* 2013)	*M. primigenius*	-	-	-	-	-	Complete
NS_IK-99-70	IK-99-70	JF912200	-	AK	Upper Ikpikpuk R., North Slope	69.37	-154.67	(Debruyne *et al.* 2008)	*M.* sp	BETA #264909	41510	480	(Enk *et al.* 2011)	46913	Complete
TP_Fishhook	M03	EU153455	-	SI	Taimyr Peninsula	74.15	99.59	(Palkopoulou *et al.* 2013)	*M. primigenius*	BETA #148647	20620	70	(Gilbert *et al.* 2008)	24833	Complete
TP_Jarkov	M02	EU153449	-	SI	Taimyr Peninsula	73.32	105.40	(Palkopoulou *et al.* 2013)	*M. primigenius*	UtC-8138	20380	140	(Gilbert *et al.* 2008)	24507	Complete
WS_GydanPen	M18	EU153447	-	SI	Gydan Peninsula	72.09	79.35	(Palkopoulou *et al.* 2013)	*M. primigenius*	OxA-17116	17125	70	(Gilbert *et al.* 2008)	20655	Complete
YA_Krause	-	NC007596	-	SI	Yakutia	71.00	145.00	(Palkopoulou *et al.* 2013)	*M. primigenius*	KIA-25289	12170	50	(Krause *et al.* 2006)	14056	Complete
AS_Elephas	-	EF588275	-	AS	Thailand	-	-	-	*E. maximus*	-	-	-	-	0	Complete

Supplemental Table 2A: Quantitative PCR and library amplification protocols

49bp 12S quantifications (9µL mastermix + 1µL template)					
Forward primer:	CCCTAAACTTTGATAGCTACC				
Reverse primer:	GTAGTTCTCTGGCGGATAGC				
Reagent	**Concentration**	**Thermal profile**			
Ampli*Taq* Gold® PCR Buffer II (Applied Biosystems)	1X	Initial Denaturation	95°C	5m	
MgCl2	2.5mM	Melt	95°C	20s	
Bovine Serum Albumin	1mg/mL	Anneal	54°C	20s	×55
dNTPs	250µM ea	Extend + Read	72°C	20s	
Primers	200nM ea	Final extension	72°C	1m	
AmpliTaq Gold®	0.1 U/µL				
EvaGreen® dye (Biotium)	1X				
Indexing amplification - Accuprime version (25-40µL mastermix + 10-25µL template)					Library sets 2-8
Forward primer:	AATGATACGGCGACCACCGAGATCTACACXXXXXXXXACACTCTTTCCCTACACGACGCTCTT				
Reverse primer:	CAAGCAGAAGACGGCATACGAGATTATXXXXXXXXACTGGAGTTCAGACGTGT				
Reagent	**Concentration**	**Thermal profile**			XXXXXXX = index sequence
Accuprime™ *Pfx* reaction mix (Invitrogen)	1X	Initial Denaturation	95°C	2m	
Primers	0.4-1uM ea	Melt	95°C	15s	
Accuprime™ *Pfx* polym.	0.042U/µL	Anneal + Read	60°C	30s	x10
EvaGreen® dye (Biotium)	0.5-0.75X	Extend	68°C	1m	
		Final extension	68°C	2m	

Supplemental Table 2A (Continued): Quantitative PCR and library amplification protocols

Indexing amplification - Herculase version (25-35µL mastermix + 15-25µL template)						Library sets 9 and above
Forward primer:	AATGATACGGCGACCACCGAGATCTACACXXXXXXXXACACTCTTTCCCTACACGACGCTCTT					
Reverse primer:	CAAGCAGAAGACGGCATACGAGATTATXXXXXXXXACTGGAGTTCAGACGTGT					
Reagent	**Concentration**	**Thermal profile**				XXXXXXX = index sequence
Herculase™ II Fusion buffer (Agilent)	1X	Initial Denaturation	95°C	2m		
Primers	75-150nM ea	Melt	95°C	15s		
dNTPs	250µM ea	Anneal + Read	60°C	30s	x10	
Herculase II Fusion polym.	0.0025U/µL	Extend	68°C	1m		
EvaGreen® dye (Biotium)	1X	Final extension	68°C	2m		
Post-enrichment amplification (37µL mastermix + 13µL template)						
Forward primer:	AATGATACGGCGACCACCGA					
Reverse primer:	CAAGCAGAAGACGGCATACGA					
Reagent	**Concentration**	**Thermal profile**				
Herculase™ II Fusion buffer (Agilent)	1X	Initial Denaturation	95°C	2m		
Primers	100nM ea	Melt	95°C	15s		
dNTPs	250µM ea	Anneal + Read	60°C	30s	x10	
Herculase II Fusion polym.	0.0025U/µL	Extend	68°C	1m		
EvaGreen® dye (Biotium)	0.5X	Final extension	68°C	2m		

Supplemental Table 2B: Library preparation protocols

Reagent	Concentration	Thermal profile			Library sets 2 & 7
END REPAIR					20uL temp + 30uL mm
NE Buffer 2	1X	PNK/UDG/Endo	37°C	3h	
BSA	100µg/µL	stop	4°C	hold	
ATP (Fisher)	1mM				
dNTPs	100µM	Then add T4 DNA polymerase, and			
T4 PNK	0.4 U/µL	PNK	25°C	15m	
Uracil-DNA glycosylase	0.1 U/µL	T4 pol	12°C	15m	
Endonuclease VIII	0.4 U/µL	stop	4°C	hold	
T4 DNA Polymerase	0.2 U/µL				MinElute to 20uL
ADAPTER LIGATION					20uL temp + 20uL mm
T4 DNA ligase buffer	1X	T4 Ligase	25°C	15m	
PEG-4000	5%	stop	4°C	hold	
Adapters	1µM				
T4 DNA ligase	0.13 U/µL				MinElute to 20uL
FILL-IN					20uL temp + 20uL mm
ThermoPol rxn buffer	1X	Bst Pol	37°C	30m	
dNTPs	0.25mM	Bst deactivate	80°C	20m	
Bst polymerase, lg frag	0.4 U/µL	stop	4°C	hold	
Reagent	**Concentration**	**Thermal profile**			Library sets 3-6 & 8
END REPAIR					20uL temp + 30uL mm
NE Buffer 2	1X	PNK/UDG/Endo	37°C	3h	
BSA	100µg/µL	stop	4°C	hold	
ATP (Fisher)	1mM				
dNTPs	100µM	Then add T4 DNA polymerase, and			
T4 PNK	0.4 U/µL	PNK	25°C	15m	
Uracil-DNA glycosylase	0.1 U/µL	T4 pol	12°C	15m	
Endonuclease VIII	0.4 U/µL	stop	4°C	hold	
T4 DNA Polymerase	0.15 U/µL				MinElute to 20uL
ADAPTER LIGATION					20uL temp + 20uL mm
T4 DNA ligase buffer	1X	T4 Ligase	25°C	15m	
PEG-4000	5%	stop	4°C	hold	
Adapters	1µM				
T4 DNA ligase	0.13 U/µL				MinElute to 20uL
FILL-IN					20uL temp + 20uL mm
ThermoPol rxn buffer	1X	Bst Pol	37°C	30m	
dNTPs	0.25mM	Bst deactivate	80°C	20m	
Bst polymerase, lg frag	0.4 U/µL	stop	4°C	hold	

Supplemental Table 2B (Continued): Library preparation protocols

Reagent	Concentration	Thermal profile			Library sets 9 & 10
END REPAIR					25uL temp + 25uL mm
NE Buffer 2	1X	PNK/UDG/Endo	37°C	3h	
BSA	100µg/µL	stop	4°C	hold	
ATP (Fisher)	1mM				
dNTPs	100µM	Then add T4 DNA polymerase, and			
T4 PNK	0.4 U/µL	PNK	25°C	15m	
Uracil-DNA glycosylase	0.1 U/µL	T4 pol	12°C	15m	
Endonuclease VIII	0.4 U/µL	stop	4°C	hold	
T4 DNA Polymerase	0.12 U/µL				MinElute to 20uL
ADAPTER LIGATION					20uL temp + 20uL mm
T4 DNA ligase buffer	1X	T4 Ligase	25°C	15m	
PEG-4000	5%	stop	4°C	hold	
Adapters	250nM				
T4 DNA ligase	0.13 U/µL				MinElute to 20uL
FILL-IN					20uL temp + 20uL mm
ThermoPol rxn buffer	1X	Bst Pol	37°C	30m	
dNTPs	0.25mM	stop	4°C	hold	
Bst polymerase, lg frag	0.4 U/µL				MinElute to 25uL

Reagent	Concentration	Thermal profile			Library set 12
END REPAIR					25uL temp + 25uL mm
NE Buffer 2	1X	PNK/UDG/Endo	37°C	3h	
BSA	100µg/µL	stop	4°C	hold	
ATP (Fisher)	1mM				
dNTPs	100µM	Then add T4 DNA polymerase, and			
T4 PNK	0.4 U/µL	PNK	25°C	15m	
Uracil-DNA glycosylase	0.1 U/µL	T4 pol	12°C	15m	
Endonuclease VIII	0.4 U/µL	stop	4°C	hold	
T4 DNA Polymerase	0.12 U/µL				MinElute to 20uL
ADAPTER LIGATION					20uL temp + 20uL mm
T4 DNA ligase buffer	1X	T4 Ligase	25°C	15m	
PEG-4000	5%	stop	4°C	hold	
Adapters	500nM				
T4 DNA ligase	0.13 U/µL				MinElute to 20uL
FILL-IN					20uL temp + 20uL mm
ThermoPol rxn buffer	1X	Bst Pol	37°C	30m	
dNTPs	0.25mM	Bst deactivate	80°C	20m	
Bst polymerase, lg frag	0.4 U/µL	stop	4°C	hold	

Supplemental Table 3: Quantitative PCR, enrichment and sequencing results

Tip Prefix	MADC#	St/Prv/Reg	Location/Locality	Species	AMS Date	Lib Set	Extract 49bp 0.1X rep1	Extract 49bp 0.1X rep2	IDLib 49bp 0.1X rep1	IDLib 49bp 0.1X rep2	Sequ. Strategy	Library Name	Round 1 on-target	Round 2 on-target	Unique Mito Reads	Mean X unique covg	Mitogen. @3X uniq	Mitogen. @2X uniq
TP_2002-472	2002/472	SI	Taimyr Peninsula	*M. primigenius*	>48800	2	nd	nd	nd	nd	Enriched	SEP01	53.25%	-	149554	786.6	100.00%	-
TP_2005-915	2005/915	SI	Taimyr Peninsula	*M. primigenius*	27740	7	121.50	213.20	715.40	651.70	Shotgun	L200s	0.06%	-	9457	38.9	100.00%	-
AK_AM104	AM104	AK	Cleary Creek	*M. primigenius*	42764	6	667.70	931.20	5201.00	6537.00	Enriched	SEP69	7.01%	-	60217	243.8	100.00%	-
AK_AM8744	AM8744	AK	Ester Creek	*M. primigenius*	16789	6	51.62	47.77	260.30	563.60	Enriched	SEP68	31.89%	-	49012	179.8	100.00%	-
ES_Ber28	Ber28	SI	Berelekh	*M. primigenius*	12125	6	148.6	187.5	2250.00	2325.00	Enriched	SEP62	34.07%	-	45800	178.0	100.00%	-
GL_Chittenango	HP1133	NY	Near Chittenango	*M.* sp	11250	1	429.2	712.3	70930.00	83500.00	Enriched	VEP82	41.25%	-	137997	416.3	100.00%	-
YU_CMNH40031	CMNH40031	YU	Old Crow	*M.* sp	-	8	13.5	15.3	25.96	31.01	Enriched	SSEP75	2.14%	-	6458	24.0	96.08%	-
GP_Colorado	DMNS23	CO	nd	*M.* sp	12475	4	5.12	11.67	131.70	101.90	Enriched	SSEP39	3.00%	-	36395	105.0	99.84%	-
GP_LaSena	DMNS28b	NE	La Sena	*M. columbi*	18440	6 & 8	0.89	2.03	4.41	3.53	Enriched	SVEP5681	0.30%	-	4300	13.2	99.66%	-
GP_Dent	DMNS38	CO	Dent	*M. columbi*	10990	6	1.60	2.46	14.14	20.07	Enriched	SEP58	0.93%	-	4795	16.4	99.70%	-
GP_Dent	DMNS40	CO	Dent	*M. columbi*	10990	6	0.78	0.96	6.41	18.50	Enriched	SEP59	2.12%	-	16578	45.8	99.90%	-
GP_Dent	DMNS43b	CO	Dent	*M. columbi*	10990	5	0.23	0.38	9.34	7.74	Enriched	SVEP41	0.17%	7.34%	17063	53.2	99.93%	-
GP_Dent	DMNS44	CO	Dent	*M. columbi*	10990	5	0.36	0.73	0.54	0.62	Enriched	SVEP43	0.05%	3.01%	21637	65.7	99.98%	-
GP_Dent	DMNS47	CO	Dent	*M. columbi*	10990	4	3.71	3.80	28.37	24.73	Enriched	SSEP38	1.49%	-	22842	59.6	99.81%	-
GP_Dent	DMNS49	CO	Dent	*M. columbi*	10990	6	1.12	1.12	2.72	5.86	Enriched	SVEP57	0.08%	6.10%	14199	40.2	99.77%	-
NS_IK-98-1087	IK-98-1087	AK	Upper Ikpikpuk R., North Slope	*M.* sp	>54000	1	4.9	3.5	170.30	170.50	Enriched	EID11	1.67%	-	3920	13.2	95.99%	-
NS_IK-99-5001	IK-99-5001	AK	Upper Ikpikpuk R., North Slope	*M.* sp	33530	6	12.62	10.98	22.63	aberrant	Enriched	SEP65	23.39%	-	12806	50.0	99.89%	-
NS_IK-99-524	IK-99-524	AK	Upper Ikpikpuk R., North Slope	*M.* sp	>51000	1	91.3	52.3	3921.00	4740.00	Enriched	EID13	18.94%	-	1764	7.7	92.17%	-
GL_Pekin	ISM01	IL	Near Pekin	*M. primigenius*	17510	3	0.60	2.95	0.68	0.04	Enriched	SVEP05	0.16%	9.94%	9125	21.4	98.53%	-
GL_ClearLk	ISM04	IL	Gravel Pit near Clear Lake	*M. primigenius*	20550	3	883.60	1166.00	153.10	252.60	Enriched	SEP06	3.50%	-	9226	25.6	99.28%	-
GL_Wyanet	ISM07	IL	Wyanet	*M. jeffersonii*	15947	3	15.47	18.22	84.02	111.10	Enriched	SEP07	3.96%	-	11104	33.9	99.73%	-
GL_Tankersley	ISM10	OH	Near Cleves	*M.* sp	-	5	0.57	0.75	0.10	0.18	Enriched	SVEP48	0.01%	1.19%	8124	19.9	96.82%	-
GL_NLaSalle	ISM12	IL	North LaSalle County	*M.* sp	12495	3	3.05	8.15	5.77	9.23	Enriched	SVEP09	0.12%	5.56%	14137	38.0	99.56%	-
GP_Brookings	ISM15	SD	Near Brookings	*M.* sp (intermediate)	12490	5	155.1	44.5	137.30	91.97	Enriched	SEP52	2.09%	-	39568	153.4	100.00%	-
WS_Lyuba	HP1095	RU	Yuribei R., Yamal Peninsula	*M. cf primigenius*	41910	2	8179.00	7745.0	1.23E+05	1.76E+05	Shotgun	SID36-10	5.53%	-	7697	35.0	100.00%	-
GL_Mott	HP1135	MI	Clayton Township	*M.* sp	12450	3	5.67	8.88	49.67	72.57	Enriched	SEP02	8.01%	-	12164	40.3	99.94%	-
EC_NewBedfordInt	HP2133	MA	New Bedford Sound	*M.* sp	-	12	6.18	8.08	nd	nd	Enriched	EEP41	0.63%	-	3192	10.7	97.98%	-
GP_Bindloss	NMC17845	AB	Bindloss	*M. columbi*	10930	1	1.17	1.50	5.16	2.74	Enriched	EID30VEP84	0.60%	-	5798	14.1	98.77%	-
YU_NMC42292	NMC-42292	YU	Dawson Area	*M. primigenius*	37920	6	148.90	232.80	aberrant	994.50	Enriched	SEP66	21.99%	-	73166	249.6	100.00%	-
YU_NMC49929	NMC-49929	YU	Dawson Area	*M. primigenius*	38600	6	121.8	138.4	1100.00	1099.00	Enriched	SEP63	35.66%	-	67616	253.8	100.00%	-
YA_Oimyakon	2006/001-2	SI	Yakutia	*M. primigenius*	41300	2	48270	64300	2.81E+05	4.71E+05	Shotgun	SID04-11	0.38%	-	9707	11.4	100.00%	-
GL_Poyser	HP1138	IN	Near Goshen	*M.* sp	23050	1	24.8	23.4	13.78	14.30	Enriched	EID23SEP03	0.36%	-	23306	67.9	99.52%	-
GL_Poyser	HP1138	IN	Near Goshen	*M.* sp	23050	1	12.6	7.0	6.60	7.94	Enriched	SEP03 above		-	-	-	-	-
GL_Randolph	HP1134	NY	East Randolph Fish Hatchery	*M. primigenius*	10350	1	1156.6	1131.8	2207.00	2183.00	Enriched	EID27	15.95%	-	34456	94.9	100.00%	-
GP_Rawlins	UW6368	WY	Near Rawlins	*M. columbi*	11560	1	11.2	5.8	21.44	25.53	Enriched	EID26	5.10%	-	10205	28.6	99.99%	-
GL_Riley	HP1726	MI	Morrison Lake Country Club, Saranac	*M.* sp (intermediate)	12320	5	1.615	3.928733	39.63	48.42	Enriched	SEP50	1.94%	-	13416	32.9	99.28%	-
EC_Scarborough	HP1137	ME	Near Scarborough	*M.* sp (intermediate)	12200	1	50.4	38.3	26.98	16.80	Enriched	EID25	0.34%	-	13680	46.0	99.30%	-
WC_SanAnCr	UCMP04	CA	San Antonio Creek	*M.* sp	19620	8	0.0049	0.02	4.66	4.92	Enriched	SVED74	0.02%	2.24%	5460	13.8	97.76%	-
WC_BeakCr	UCMP17	OR	Beak (Bear?) Creek	*M.* sp	18510	5	17.59	25.71	31.01	52.66	Enriched	SEP44	1.15%	-	13674	40.0	99.82%	-
WC_SanAnCr	UCMP19	CA	San Antonio Creek	*M.* sp	-	5	4.40	6.92	10.59	18.54	Enriched	SSEP45	0.61%	-	17279	38.8	99.57%	-
GL_Hughes	UM22798	MI	North of Assyria	*M. jeffersonii*	11100	5	0.01	0.13	1.19	2.52	Enriched	SVEP49	0.18%	4.16%	11282	27.0	99.14%	-
GP_RWF	UNSM01	NE	Red Willow Fauna, Rw-102	*M. sp*	17070	3	366.40	451.30	398.10	438.90	Enriched	SEP11	8.03%	-	15690	45.6	99.78%	-
GP_RWF	UNSM02	NE	Red Willow Fauna, Rw-102	*M. sp*	12130	3	212.50	259.50	47.90	70.35	Enriched	SSEP12	0.82%	-	15832	44.5	99.54%	-
GP_RWF	UNSM08	NE	Red Willow Fauna, Rw-102	*M. columbi*	16160	3 & 8	3.03	4.36	0.65	4.98	Enriched	SVEP1478	1.19%	-	6196	15.2	97.79%	-
GP_RWF	UNSM09	NE	Little Sand Pit near McCook, Rw-110	*M. sp*	11585	3	14.21	17.98	0.65	3.58	Enriched	SVEP15	0.03%	0.95%	7552	21.0	95.71%	-
GP_Beverly	UNSM13	NE	Beverly Gravel Pits	*M. columbi*	10650	8	1.35	0.45	0.09	no	Enriched	SVEP72	0.06%	2.31%	4457	10.0	94.27%	-
GP_Trenton	UNSM14	NE	Trenton Reservoir	*M. sp*	33670	3 & 8	1.13	0.15	11.61	7.10	Enriched	SVEP1679	0.22%	-	4064	9.7	92.89%	-
GP_Rh	UNSM15	NE	Richardson Co, Big Nemaha R.	*M. columbi*	-	3	6.15	1.75	13.54	13.75	Enriched	SEP17	1.38%	-	10832	25.4	99.20%	-

Supplemental Table 3 (Continued): Quantitative PCR, enrichment and sequencing results

Tip Prefix	MADC#	St/Prv/Reg	Location/Locality	*Species*	AMS Date	Lib Set	Extract 49bp 0.1X rep1	Extract 49bp 0.1X rep2	IDLib 49bp 0.1X rep1	IDLib 49bp 0.1X rep2	Sequ. Strategy	Library Name	Round 1 on-target	Round 2 on-target	Unique Mito Reads	Mean X unique covg	Mitogen. @3X uniq	Mitogen. @2X uniq
GP_CrappieHl	UNSM16	NE	Crappie Hole	*M.* sp (intermediate)	23590	3	11.37	10.32	30.30	23.48	Enriched	SSEP18	0.84%	-	12229	42.0	99.85%	-
GP_BigNemaha	UNSM21	NE	South Fork Big Nemaha R.	*M. columbi*	13850	3	0.03	0.01	24.74	16.33	Enriched	SEP19	7.68%	-	14966	37.7	99.75%	-
GP_RWF	UNSM22	NE	Palisade Sand Pit	*M.* sp	-	3	2.20	2.12	0.20	0.44	Enriched	SVEP20	0.05%	2.44%	10276	26.5	98.58%	-
GL_BigBoneLk	UNSM27	KY	Big Bone Lick	*M. primigenius*	13985	4	0.37	0.14	0.32	1.39	Enriched	SSEP25	1.94%	-	6726	14.9	96.28%	-
GL_BigBoneLk	UNSM29	KY	Big Bone Lick	*M. primigenius*	12930	4	1.90	1.78	0.01	2.90	Enriched	SSEP26	2.96%	-	4732	10.5	91.52%	-
GL_BigBoneLk	UNSM30	KY	Big Bone Lick	*M. primigenius*	13215	5	8.2	6.5	267.80	226.00	Enriched	SEP53	44.21%	-	44328	121.1	99.79%	-
GL_BigBoneLk	UNSM31	KY	Big Bone Lick	*M. primigenius*	13950	4	0.23	1.44	0.00	5.77	Enriched	SSED27	52.10%	-	6413	13.5	95.28%	-
GL_BigBoneLk	UNSM32	KY	Big Bone Lick	*M. primigenius*	13860	4	0.25	1.39	0.33	4.48	Enriched	SSED28	31.81%	-	7490	16.5	96.77%	-
GP_CrappieHl	UNSM33	NE	Crappie Hole	*M.* sp (intermediate)	-	4	0.47	0.0004	0.07	2.26	Enriched	SVEP29	0.07%	4.07%	2747	8.4	94.12%	-
GP_CrappieHl	UNSM34	NE	Crappie Hole	*M. columbi*	23670	4	82.45	115.80	108.80	150.20	Enriched	SEP30	3.83%	-	7008	21.1	99.68%	-
GP_CrappieHl	UNSM35	NE	Crappie Hole	*M.* sp	23800	4	9.80	13.89	108.20	122.80	Enriched	SEP31	3.80%	-	7036	20.9	99.65%	-
GP_Biehl	UNSM42	NE	Biehl Farm	*M.* sp	15200	4	0.46	0.07	0.24	2.59	Enriched	SVEP32	0.09%	6.73%	5400	14.9	99.18%	-
GP_UW20579	UW20579	WY	nd	*M.* sp	38260	1	209.7	238.9	444.10	561.80	Enriched	EID15	25.20%	-	97481	240.8	100.00%	-
WC_WenasCr	WAST_01	WA	Wenas Creek	*M.* sp	13398	4	0.42	0.42	1.19	0.69	Enriched	SVEP37	0.12%	5.53%	5868	14.8	98.13%	-
YU_YPC130.0002	YPC130.0002	YU	Quartz Creek	*M.* sp	36690	6	57.80	100.30	67.18	38.50	Enriched	SEP64	40.90%	-	9577	35.9	100.00%	-
YU_YPC136.0005	YPC136.0005	YU	Sulphur Creek	*M. primigenius*	-	1	28.1	31.5	69.34	104.30	Enriched	EID03	2.71%	-	5052	16.6	97.17%	-
YU_YPC173.001	YPC173.001	YU	Ch'ijee's Bluff	*M. primigenius*	>45400	6	694.4	647.5	2265.00	2219.00	Enriched	SEP61	33.82%	-	90864	309.9	100.00%	-
YU_YPC3.0229	YPC3.0229	YU	Finning	*M. primigenius*	-	1	89.9	89.9	777.80	1254.00	Enriched	EID07	0.71%	-	1739	7.8	95.96%	-
YU_YPC3.0256	YPC3.0256	YU	Finning	*M. primigenius*	28960	1	9.1	10.3	409.20	621.30	Enriched	EID05	5.66%	-	10574	32.9	98.95%	-
YU_YPC5.0046	YPC5.0046	YU	Hunker Creek	*M.* sp	22430	1	208.2	611.9	10250.00	7960.00	Enriched	EID10	44.07%	-	4285	20.3	99.64%	-
MW_BadgerCr	DMNS08	CO	Badger Creek	*M.* sp	-	4	1.38	1.80	9.09	6.18	Enriched	SED35	6.14%	-	405	1.2	18.68%	34.55%
NS_IK-99-235	IK-99-235	AK	Upper Ikpikpuk R., North Slope	*M.* sp	40870	1	0.9	1.0	6.75	3.27	Enriched	EID09	0.40%	-	820	3.1	58.17%	77.94%
GL_Toledo	ISM09	IL	Near Toledo	*M.* cf *primigenius*	-	3	2.69	2.57	0.21	no	Enriched	SVEP08	0.29%	16.93%	3814	8.5	89.11%	-
EC_NewBedfordSm	HP2134	MA	New Bedford Sound	*M.* sp	-	12	3.89	7.87	nd	nd	Enriched	EEP42	0.07%	-	434	1.5	20.63%	44.16%
WC_Pygmy	SBMNH27	CA	Santa Rosa Island	*M. exilis*	-	12	0.0034	0.0153	nd	nd	Enriched	EEP43	0.12%	-	721	1.9	28.03%	51.28%
MW_Snowmass	SM3	CO	Snowmass Site	*M. columbi*	-	4	0.00	0.11	no	0.02	Enriched	SVEP36	0.26%	3.91%	3446	6.8	67.52%	-
WC_WhidbeyIs	UCMP09	WA	Whidbey Island	*M.* sp	19200	5 & 8	1.16	2.40	no	no	Enriched	SVEP7380	5.16%	-	2333	5.3	78.30%	-
RWF	UNSM07	NE	Red Willow Fauna, Rw-102	*M.* sp	-	3	3.50	3.88	1.40	0.36	Enriched	SVEP13	0.01%	0.32%	1320	3.2	47.80%	-
RWF	UNSM23	NE	Palisade Sand Pit	*M.* sp	-	3	6.02	2.21	6.86	15.56	Enriched	SEP22	0.25%	-	566	1.6	25.94%	43.02%
RWF	UNSM24	NE	Palisade Sand Pit	*M.* sp	-	4	6.37	3.31	0.38	1.88	Enriched	SEP23	0.24%	-	496	1.3	17.99%	34.65%
YU_YP180.40	YP180.40	YU	Old Crow; CRH-94	*M.* sp	-	1	3.9	2.6	18.76	13.12	Enriched	EID20	2.68%	-	1583	5.1	82.70%	-
YU_YP180.41	YP180.41	YU	Old Crow; CRH-94	*M.* sp	-	1	1.0	0.3	13.27	11.27	Enriched	EID18	1.14%	-	1934	5.3	79.40%	-
YU_YP221.2	YP221.2	YU	Old Crow; CRH-11A	*M.* sp	-	1	0.8	1.2	no	no	Enriched	VEP83	0.07%	-	1053	3.1	55.66%	-

Supplemental Table 4: Priors used in BEAST analyses and resulting posterior statistics

Setup		Posterior summary statistic	posterior	prior	likelihood	treeModel.rootHeight	tmrca(Mammoths)	constant.popSize
Alignment	All complete mitochondrial genomes	**mean**	-32519.90	271.24	-32791.14	0.0294	0.0064	0.0120
Temp. Calib.	None	**stderr of mean**	0.18	0.08	0.16	0.0000	0.0000	0.0000
Site model	GTR+I+G, 5 gamma categories	**stdev**	8.45	3.94	7.47	0.0014	0.0004	0.0014
Pop. model	Constant	**variance**	71.41	15.56	55.81	0.0000	0.0000	0.0000
Clock model	Strict, unestimated	**median**	-32519.62	271.24	-32790.77	0.0293	0.0063	0.0119
Tip dates?	No	**geometric mean**	n/a	271.21	n/a	0.0293	0.0063	0.0119
kappas	Uniform distribution, 0-1000	**95% HPD lower**	-32536.72	263.64	-32805.67	0.0266	0.0057	0.0094
alpha	Uniform distribution, 0-10	**95% HPD upper**	-32504.16	278.91	-32776.89	0.0321	0.0071	0.0148
Generations	10,000,000	**auto-correlation time (ACT)**	3859.19	3299.88	3896.42	1239.74	4411.88	1386.07
		effective sample size (ESS)	2332.35	2727.67	2310.07	7260.40	2040.17	6493.90

Supplemental Table 4 (Continued): Posterior statistics of BEAST analyses

ac	ag	at	cg	gt	frequencies1	frequencies2	frequencies3	frequencies4	alpha	pInv	clock.rate	treeLikelihood	coalescent
0.0180	0.8151	0.0149	0.0283	0.0067	0.3296	0.2507	0.1330	0.2867	0.7510	0.6391	1.0000	-32791.14	298.69
0.0001	0.0014	0.0001	0.0002	0.0001	0.0002	0.0001	0.0001	0.0002	0.0214	0.0042	n/a	0.16	0.07
0.0040	0.0568	0.0034	0.0079	0.0039	0.0037	0.0035	0.0026	0.0035	0.2453	0.0483	n/a	7.47	3.87
0.0000	0.0032	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0602	0.0023	n/a	55.81	15.01
0.0177	0.8131	0.0147	0.0275	0.0059	0.3296	0.2507	0.1329	0.2869	0.7110	0.6436	n/a	-32790.77	298.68
0.0176	0.8131	0.0146	0.0272	0.0055	0.3296	0.2507	0.1329	0.2867	0.7152	0.6372	n/a	n/a	298.67
0.0104	0.7012	0.0089	0.0143	0.0006	0.3224	0.2437	0.1277	0.2797	0.3593	0.5471	n/a	-32805.67	290.82
0.0259	0.9213	0.0215	0.0440	0.0142	0.3365	0.2574	0.1378	0.2935	1.2108	0.7306	n/a	-32776.89	305.84
5765.45	5777.79	5839.96	5685.51	13115.48	15593.51	12441.12	14649.60	17422.59	68736.04	67113.82	n/a	3896.42	2950.55
1561.20	1557.86	1541.28	1583.15	686.29	577.23	723.49	614.42	516.63	130.95	134.12	n/a	2310.07	3050.62

Supplemental Table 4 (Continued): Priors used in BEAST analyses and resulting posterior statistics

Setup		Posterior summary statistic	posterior	prior	likelihood	treeModel. rootHeight	tmrca (CladeI)	tmrca (HapB)	tmrca (Mammoths)	constant. popSize
Alignment	All complete finite-dated mitogenomes	**mean**	-32438.6	-1065.53	-31373.07	6.13E+06	3.70E+05	5.70E+05	9.82E+05	1.22E+06
Temp. Calib.	Root + tips	**stderr of mean**	1.0126	0.99	0.1136	9579.939	5163.019	10158.67	16045.68	18170.79
Site model	TN+I+G, 5 gamma categories	**stdev**	21.0854	18.81	8.1414	5.67E+05	1.03E+05	2.01E+05	2.99E+05	3.89E+05
Pop. model	Constant	**variance**	444.5937	353.8151	66.2819	3.22E+11	1.07E+10	4.03E+10	8.95E+10	1.51E+11
Clock model	Uncorrelated lognormal	**median**	-32438.29	-1065.24	-31372.84	6.13E+06	3.57E+05	5.40E+05	9.37E+05	1.15E+06
Tip dates?	Yes; radiocarbon years before present	**geometric mean**	n/a	n/a	n/a	6.11E+06	3.57E+05	5.37E+05	9.41E+05	1.16E+06
kappas	Uniform distribution, 0-1000	**95% HPD lower**	-32477.79	-1100.72	-31388.71	4.98E+06	1.87E+05	2.20E+05	4.81E+05	5.52E+05
alpha	Uniform distribution, 0-10	**95% HPD upper**	-32395.79	-1028.59	-31356.97	7.20E+06	5.66E+05	9.60E+05	1.59E+06	1.96E+06
ucldstdev	Uniform distribution, 0-10	**auto-corr. time (ACT)**	1.04E+05	1.25E+05	8766.5022	12828.74	1.12E+05	1.15E+05	1.29E+05	98269.47
ucldmean	Uniform distribution, 1e-12-0.1	**effective samp. size (ESS)**	433.5801	360.94	5133.7465	3508.14	400.8697	390.8118	347.628	457.9754
rootHeight	Normal distribution, 6.6e6, stdev 5.5									
Generations	50,000,000									

Supplemental Table 4 (Continued): Posterior statistics of BEAST analyses

kappa1	kappa2	frequencies1	frequencies2	frequencies3	frequencies4	alpha	pInv	ucld.mean	ucld.stdev	meanRate	coefficientOfVariation	covariance	treeLikelihood	coalescent
58.8666	75.8169	0.3295	0.2509	0.1333	0.2863	0.8878	0.6559	7.68E-09	0.4517	5.71E-09	0.4542	-0.0101	-31373.1	-1016.65
0.3535	0.4418	7.82E-05	7.09E-05	5.51E-05	7.53E-05	0.0179	2.67E-03	1.10E-10	1.72E-03	3.41E-11	1.67E-03	1.04E-03	0.1136	1.0019
9.1362	11.3373	3.66E-03	3.38E-03	2.62E-03	3.55E-03	0.3993	0.058	2.09E-09	0.0862	7.70E-10	0.0888	0.0843	8.1414	19.0918
83.4699	128.534	1.34E-05	1.14E-05	6.84E-06	1.26E-05	0.1595	3.36E-03	4.38E-18	7.44E-03	5.93E-19	7.89E-03	7.11E-03	66.2819	364.496
57.8385	74.5585	0.3295	0.2509	0.1332	0.2863	0.8025	0.663	7.38E-09	0.4469	5.68E-09	0.4487	-0.0113	-31372.8	-1016.38
58.1844	74.9997	0.3295	0.2508	0.1332	0.2863	0.8132	0.6532	7.41E-09	0.4434	5.66E-09	0.4455	n/a	n/a	n/a
42.3892	55.511	0.3224	0.2445	0.1281	0.2795	0.2788	0.5387	4.15E-09	0.2956	4.30E-09	0.2911	-0.1736	-31388.7	-1052.49
77.036	98.8818	0.3369	0.2574	0.1382	0.2933	1.6729	0.7573	1.21E-08	0.6313	7.32E-09	0.6358	0.1549	-31357	-978.845
67388.83	68355.02	20575.56	19825.53	19959.03	20270.38	90321.51	95337.16	1.23E+05	17989.53	88126.744	15958	6883.645	8766.502	1.24E+05
667.8407	658.4008	2187.304	2270.052	2254.869	2220.235	498.2756	472.0615	364.7443	2501.733	510.6849	2820.215	6537.961	5133.747	363.0675

Supplemental Table 4 (Continued): Priors used in BEAST analyses and resulting posterior statistics

Setup		Posterior summary statistic	posterior	prior	likelihood	treeModel. rootHeight	tmrca (CladeI)	tmrca (HapB)	tmrca (Mammoths)	constant. popSize
Alignment	All complete finite-dated mitogenomes	**mean**	-32490.44	-1117.78	-31372.67	6.56E+06	6.68E+05	9.95E+05	1.79E+06	2.65E+06
Temp. Calib.	Root only	**stderr of mean**	1.1162	1.1204	0.0915	4721.44	13728.85	21973.99	39311.97	49549.1
Site model	TN+I+G, 5 gamma categories	**stdev**	21.2288	19.8734	8.3635	5.49E+05	2.38E+05	4.13E+05	6.31E+05	8.89E+05
Pop. model	Constant	**variance**	450.6634	394.9502	69.9476	3.01E+11	5.67E+10	1.71E+11	3.99E+11	7.90E+11
Clock model	Uncorrelated lognormal	**median**	-32489.98	-1117.1	-31372.3	6.56E+06	6.23E+05	9.25E+05	1.68E+06	2.49E+06
Tip dates?	No	**geometric mean**	n/a	n/a	n/a	6.54E+06	6.32E+05	9.18E+05	1.69E+06	2.52E+06
kappas	Uniform distribution, 0-1000	**95% HPD lower**	-32531.16	-1158.98	-31389.11	5.47E+06	3.00E+05	3.06E+05	7.68E+05	1.23E+06
alpha	Uniform distribution, 0-10	**95% HPD upper**	-32448.85	-1081.31	-31356.85	7.63E+06	1.14E+06	1.78E+06	3.05E+06	4.42E+06
ucldstdev	Uniform distribution, 0-10	**auto-correlation time (ACT)**	1.87E+05	2.15E+05	8081.6985	5000	2.25E+05	1.91E+05	2.62E+05	2.10E+05
ucldmean	Uniform distribution, 1e-12-0.1	**effective sample size (ESS)**	361.7172	314.6097	8353.4421	13502	300.7171	353.425	257.9294	321.8088
rootHeight	Uniform distribution, 0-1e8									
Generations	75,000,000									

Supplemental Table 4 (Continued): Posterior statistics of BEAST analyses

kappa1	kappa2	frequencies1	frequencies2	frequencies3	frequencies4	alpha	pInv	ucld.mean	ucld.stdev	meanRate	coefficientOfVariation	covariance	treeLikelihood	coalescent
58.6305	75.4348	0.3295	0.2509	0.1332	0.2864	0.8703	0.6517	4.26E-09	0.4887	4.19E-09	0.4985	-0.0118	-31372.7	-1069.29
0.25	0.3119	6.92E-05	5.72E-05	4.27E-05	5.71E-05	0.0145	2.35E-03	6.48E-11	1.60E-03	3.57E-11	1.79E-03	8.51E-04	0.0915	1.1205
9.0051	11.2412	3.68E-03	3.24E-03	2.59E-03	3.40E-03	0.385	0.0579	1.27E-09	0.0921	7.12E-10	0.0992	0.0828	8.3635	19.9082
81.0911	126.365	1.35E-05	1.05E-05	6.68E-06	1.15E-05	0.1482	3.35E-03	1.61E-18	8.48E-03	5.06E-19	9.84E-03	6.85E-03	69.9476	396.3346
57.7444	74.233	0.3296	0.2509	0.1332	0.2864	0.7956	0.6602	4.11E-09	0.483	4.16E-09	0.4895	-0.0138	-31372.3	-1068.67
57.9625	74.626	0.3295	0.2509	0.1331	0.2864	0.8005	0.6489	4.08E-09	0.4801	4.12E-09	0.4888	n/a	n/a	n/a
42.6169	55.3508	0.3222	0.2446	0.1282	0.2796	0.2841	0.5378	2.03E-09	0.3175	2.80E-09	0.3164	-0.1674	-31389.1	-1108.32
76.7331	98.2064	0.3364	0.2573	0.1382	0.2928	1.6097	0.7548	6.81E-09	0.676	5.59E-09	0.7008	0.1529	-31356.9	-1030.6
52053.86	51987.79	23908.85	21023.46	18436.86	19081.2	96369.72	1.11E+05	1.76E+05	20294.9	1.69E+05	22035.82	7134.223	8081.699	2.14E+05
1296.926	1298.574	2823.64	3211.175	3661.686	3538.038	700.5313	607.8991	384.1385	3326.452	398.3105	3063.648	9462.838	8353.442	315.6559

Supplemental Table 4 (Continued): Priors used in BEAST analyses and resulting posterior statistics

Setup		Posterior summary statistic	posterior	prior	likelihood	treeModel. rootHeight	tmrca (CladeI)	tmrca (HapB)	tmrca (Mammoths)	constant.popSize
Alignment	All complete finite-dated mitogenomes	**mean**	-32381.75	-1006.76	-31374.99	1.66E+06	1.78E+05	2.70E+05	4.47E+05	4.83E+05
Temp. Calib.	Tips only	**stderr of mean**	0.6252	0.6389	0.0824	33930.2	1474.062	3062.056	4958.207	4870.584
Site model	TN+I+G, 5 gamma categories	**stdev**	17.2903	15.5607	8.3644	5.89E+05	38614.41	79559.23	1.16E+05	1.34E+05
Pop. model	Constant	**variance**	298.9532	242.1365	69.963	3.46E+11	1.49E+09	6.33E+09	1.34E+10	1.81E+10
Clock model	Uncorrelated lognormal	**median**	-32380.94	-1005.51	-31374.71	1.58E+06	1.72E+05	2.58E+05	4.29E+05	4.61E+05
Tip dates?	Yes; radiocarbon years before present	**geometric mean**	n/a	n/a	n/a	1.56E+06	1.75E+05	2.59E+05	4.33E+05	4.66E+05
kappas	Uniform distribution, 0-1000	**95% HPD lower**	-32415.17	-1037.4	-31391.33	6.08E+05	1.16E+05	1.32E+05	2.62E+05	2.70E+05
alpha	Uniform distribution, 0-10	**95% HPD upper**	-32347.55	-977.774	-31358.61	2.82E+06	2.54E+05	4.29E+05	6.78E+05	7.50E+05
ucldstdev	Uniform distribution, 0-10	**auto-correlation time (ACT)**	1.18E+05	1.52E+05	8744.9351	2.99E+05	1.31E+05	1.33E+05	1.65E+05	1.18E+05
ucldmean	Uniform distribution, 1e-12-0.1	**effective sample size (ESS)**	764.67	593.2164	10292.815	300.8954	686.1882	675.0422	545.8424	761.3393
rootHeight	Uniform distribution, 0-1e8									
Generations	100,000,000									

Supplemental Table 4 (Continued): Posterior statistics of BEAST analyses

kappa1	kappa2	frequencies1	frequencies2	frequencies3	frequencies4	alpha	pInv	ucld.mean	ucld.stdev	meanRate	coefficientOfVariation	covariance	treeLikelihood	coalescent
58.9889	75.9325	0.3297	0.2509	0.1332	0.2863	0.8886	0.6542	1.66E-08	0.4386	1.78E-08	0.4409	-0.0191	-31375	-954.495
0.234	0.2898	5.38E-05	5.14E-05	3.87E-05	5.31E-05	0.0127	2.02E-03	1.31E-10	1.49E-03	2.69E-10	1.73E-03	6.89E-04	0.0824	0.6389
9.178	11.4644	3.52E-03	3.32E-03	2.57E-03	3.44E-03	0.3853	0.0581	3.37E-09	0.0906	4.54E-09	0.095	0.0826	8.3644	15.5607
84.2363	131.4327	1.24E-05	1.10E-05	6.58E-06	1.19E-05	0.1484	3.38E-03	1.13E-17	8.21E-03	2.06E-17	9.03E-03	6.82E-03	69.963	242.1365
58.0911	74.79	0.3297	0.2509	0.1331	0.2862	0.8148	0.6626	1.64E-08	0.4321	1.74E-08	0.4318	-0.0202	-31374.7	-953.239
58.2976	75.0942	0.3296	0.2509	0.1331	0.2862	0.8173	0.6514	1.62E-08	0.4292	1.73E-08	0.4308	n/a	n/a	n/a
41.6237	54.48	0.3229	0.2444	0.1282	0.2798	0.2826	0.5411	1.02E-08	0.2583	9.58E-09	0.2664	-0.1725	-31391.3	-985.133
76.8261	98.6564	0.3366	0.2572	0.1381	0.293	1.6253	0.7578	2.34E-08	0.6108	2.71E-08	0.6313	0.1468	-31358.6	-925.508
58523.12	57525.37	20992.38	21538.6	20466.51	21439.02	97587.17	1.09E+05	1.37E+05	24459.43	3.17E+05	29936.35	6269.07	8744.935	1.52E+05
1538.025	1564.701	4287.746	4179.01	4397.916	4198.419	922.3548	825.3502	657.5871	3679.972	284.3957	3006.713	14357.79	10292.82	593.2164

Supplemental Table 4 (Continued): Priors used in BEAST analyses and resulting posterior statistics

Setup		Posterior summary statistic	posterior	prior	likelihood	treeModel. rootHeight	tmrca (Clade I)	tmrca (HapB)	tmrca (Mammoths)	constant.popSize
Alignment	As above but with CMNH40031@200kya	**mean**	-32537.49	-1087.95	-31449.54	6.26E+06	4.06E+05	6.44E+05	1.08E+06	1.35E+06
Temp. Calib.	Root + tips	**stderr of mean**	0.8282	0.8159	0.1115	6985.585	5062.897	7792.609	15732.34	16686.64
Site model	TN+I+G, 5 gamma categories	**stdev**	19.1833	17.0357	8.2117	5.04E+05	1.05E+05	1.74E+05	2.87E+05	3.89E+05
Pop. model	Constant	**variance**	367.9988	290.2151	67.4321	2.54E+11	1.10E+10	3.02E+10	8.25E+10	1.51E+11
Clock model	Uncorrelated lognormal	**median**	-32536.88	-1087.68	-31449.19	6.26E+06	3.89E+05	6.16E+05	1.04E+06	1.30E+06
Tip dates?	Yes; radiocarbon years before present	**geometric mean**	n/a	n/a	n/a	6.24E+06	3.94E+05	6.22E+05	1.05E+06	1.30E+06
kappas	Uniform distribution, 0-1000	**95% HPD lower**	-32574.57	-1122.09	-31466.22	5.29E+06	2.33E+05	3.66E+05	5.89E+05	7.19E+05
alpha	Uniform distribution, 0-10	**95% HPD upper**	-32499.9	-1055.85	-31433.92	7.24E+06	6.10E+05	1.01E+06	1.65E+06	2.14E+06
ucldstdev	Uniform distribution, 0-10	**auto-correlation time (ACT)**	83898.631	1.03E+05	8304.0155	8661.308	1.04E+05	90452.54	1.35E+05	82749.96
ucldmean	Uniform distribution, 1e-12-0.1	**effective sample size (ESS)**	536.4807	435.9141	5420.2693	5196.675	430.7802	497.609	333.2334	543.9278
rootHeight	Normal distribution, 6.6e6, stdev 5.5									
Generations	50,000,000									

Supplemental Table 4 (Continued): Posterior statistics of BEAST analyses

kappa1	kappa2	frequencies1	frequencies2	frequencies3	frequencies4	alpha	pInv	ucld.mean	ucld.stdev	meanRate	coefficientOfVariation	covariance	treeLikelihood	coalescent
59.422	75.9806	0.3296	0.2509	0.1331	0.2864	0.8952	0.6593	6.78E-09	0.4335	5.40E-09	0.4372	-0.0081	-31449.5	-1039.33
0.3562	0.4338	8.52E-05	7.33E-05	5.80E-05	8.16E-05	0.0183	2.66E-03	7.60E-11	1.80E-03	2.62E-11	1.80E-03	1.08E-03	0.1115	0.8217
9.0944	11.2499	3.64E-03	3.31E-03	2.52E-03	3.45E-03	0.3929	0.0532	1.60E-09	0.0816	6.46E-10	0.085	0.0835	8.2117	17.2217
82.7087	126.5597	1.32E-05	1.10E-05	6.35E-06	1.19E-05	0.1544	2.83E-03	2.57E-18	6.66E-03	4.17E-19	7.22E-03	6.97E-03	67.4321	296.5859
58.6139	75.0315	0.3296	0.2508	0.1332	0.2864	0.807	0.6647	6.61E-09	0.4294	5.39E-09	0.4309	-0.009	-31449.2	-1039.1
58.7469	75.1744	0.3296	0.2508	0.1331	0.2864	0.8253	0.6571	6.59E-09	0.4258	5.36E-09	0.429	n/a	n/a	n/a
42.5206	56.5558	0.322	0.2445	0.1284	0.2798	0.3337	0.5519	3.99E-09	0.2765	4.14E-09	0.2739	-0.1674	-31466.2	-1072.3
77.2493	99.5409	0.3362	0.2574	0.1382	0.2931	1.6569	0.7543	1.01E-08	0.5949	6.68E-09	0.6055	0.1615	-31433.9	-1005.16
69071.12	66926.84	24710.8	22030.84	23806.98	25232.38	97985.31	1.13E+05	1.01E+05	21810.89	74187.583	20125.8	7478.405	8304.016	1.02E+05
651.6472	672.5254	1821.471	2043.046	1890.622	1783.819	459.3546	399.9591	444.7083	2063.649	606.7053	2236.433	6018.663	5420.269	439.2002

Supplemental Table 4 (Continued): Priors used in BEAST analyses and resulting posterior statistics

Setup		Posterior summary statistic	posterior	prior	likelihood	treeModel. rootHeight	tmrca (Clade I)	tmrca (HapB)	tmrca (Mammoths)	constant.popSize
Alignment	As above but with CMNH40031@200kya	**mean**	-32486.49	-1035.92	-31450.57	2.07E+06	2.13E+05	3.85E+05	5.75E+05	6.02E+05
Temp. Calib.	Tips only	**stderr of mean**	0.9294	0.9296	0.11	45367.81	2558.519	4140.982	8290.487	8739.166
Site model	TN+I+G, 5 gamma categories	**stdev**	18.1876	16.273	8.2694	7.36E+05	48353.55	79372.67	1.32E+05	1.72E+05
Pop. model	Constant	**variance**	330.7884	264.8116	68.3837	5.41E+11	2.34E+09	6.30E+09	1.75E+10	2.94E+10
Clock model	Uncorrelated lognormal	**median**	-32485.62	-1034.66	-31450.37	1.95E+06	2.05E+05	3.71E+05	5.51E+05	5.72E+05
Tip dates?	Yes; radiocarbon years before present	**geometric mean**	n/a	n/a	n/a	1.95E+06	2.09E+05	3.78E+05	5.61E+05	5.80E+05
kappas	Uniform distribution, 0-1000	**95% HPD lower**	-32521.63	-1068.8	-31466.06	8.68E+05	1.33E+05	2.58E+05	3.67E+05	3.28E+05
alpha	Uniform distribution, 0-10	**95% HPD upper**	-32450.54	-1004.83	-31433.96	3.57E+06	3.10E+05	5.43E+05	8.49E+05	9.54E+05
ucldstdev	Uniform distribution, 0-10	**auto-correlation time (ACT)**	1.18E+05	1.47E+05	7972.1949	1.71E+05	1.26E+05	1.23E+05	1.77E+05	1.17E+05
ucldmean	Uniform distribution, 1e-12-0.1	**effective sample size (ESS)**	382.9266	306.4174	5645.873	262.9862	357.1339	367.3559	254.9746	385.5122
rootHeight	Uniform distribution, 0-1e8									
Generations	50,000,000									

Supplemental Table 4 (Continued): Posterior statistics of BEAST analyses

kappa1	kappa2	frequencies1	frequencies2	frequencies3	frequencies4	alpha	pInv	ucld.mean	ucld.stdev	meanRate	coefficientOfVariation	covariance	treeLikelihood	coalescent
59.1839	75.7118	0.3296	0.2509	0.1332	0.2864	0.8817	0.6511	1.35E-08	0.4438	1.42E-08	0.4449	-0.0182	-31450.6	-983.651
0.2976	0.3682	7.30E-05	6.64E-05	5.11E-05	6.68E-05	0.0256	4.02E-03	1.50E-10	1.80E-03	2.28E-10	1.86E-03	9.55E-04	0.11	0.9296
8.9348	11.1846	3.60E-03	3.24E-03	2.57E-03	3.32E-03	0.4203	0.0623	2.85E-09	0.0879	3.50E-09	0.0903	0.0831	8.2694	16.273
79.831	125.0959	1.30E-05	1.05E-05	6.60E-06	1.10E-05	0.1766	3.88E-03	8.10E-18	7.73E-03	1.22E-17	8.15E-03	6.91E-03	68.3837	264.8116
58.5033	74.8361	0.3297	0.2509	0.1332	0.2863	0.7899	0.6594	1.34E-08	0.4374	1.40E-08	0.4376	-0.0204	-31450.4	-982.396
58.5242	74.9047	0.3296	0.2508	0.1331	0.2863	0.8008	0.6479	1.32E-08	0.4351	1.38E-08	0.4358	n/a	n/a	n/a
42.121	55.0625	0.3227	0.2445	0.1284	0.2798	0.2668	0.5261	7.83E-09	0.2823	7.84E-09	0.283	-0.177	-31466.1	-1016.53
76.7229	97.8692	0.3366	0.2571	0.1384	0.2926	1.6755	0.757	1.91E-08	0.6247	2.17E-08	0.6311	0.151	-31434	-952.56
49956.85	48775.42	18507.4	18831.54	17791.78	18218.83	1.67E+05	1.88E+05	1.26E+05	18850.02	1.92E+05	19130.92	5941.302	7972.195	1.47E+05
900.9776	922.8008	2432	2390.14	2529.821	2470.521	268.9129	239.2073	357.8492	2387.796	234.7175	2352.736	7575.781	5645.873	306.4174

Supplemental Table 5: Priors used in Bayesian Skygrid analyses and resulting posterior statistics

Setup		Posterior summary statistic	post-erior	prior	likeli-hood	treeModel. rootHeight	skygrid. precision	skygrid. logPopSize1	skygrid. logPopSize2	skygrid. logPopSize3	skygrid. logPopSize4
Alignment	Finite-dated complete mammoth mitos	**mean**	-29594.7	-974.998	-28619.7	4.11E+05	4.14E+01	1.22E+01	1.28E+01	1.29E+01	1.32E+01
Temp. Calib.	Tips only	**stderr of mean**	1.5443	1.5741	0.2191	8.51E+03	3.66E+00	1.69E-02	1.43E-02	1.57E-02	1.58E-02
Site model	TN+I+G, 5 gamma categories	**stdev**	34.7962	34.4496	8.4257	1.09E+05	1.66E+02	3.94E-01	3.45E-01	3.63E-01	4.38E-01
Pop. model	Constant	**variance**	1210.778	1186.777	70.9927	1.19E+10	2.74E+04	1.55E-01	1.19E-01	1.32E-01	1.92E-01
Clock model	Uncorrelated lognormal	**median**	-29600.5	-981.243	-28619.4	4.02E+05	4.54E+00	1.22E+01	1.28E+01	1.29E+01	1.32E+01
Tip dates?	Yes; radiocarbon years before present	**geometric mean**	n/a	n/a	n/a	3.98E+05	6.11E+00	1.22E+01	1.28E+01	1.29E+01	1.32E+01
kappas	Uniform distribution, 0-500	**95% HPD Lower**	-29655.1	-1031.95	-28636.7	2.15E+05	1.37E-01	1.14E+01	1.22E+01	1.22E+01	1.25E+01
alpha	Uniform distribution, 0-10	**95% HPD Upper**	-29516.6	-894.402	-28604.3	6.20E+05	1.90E+02	1.29E+01	1.35E+01	1.36E+01	1.41E+01
ucldstdev	Uniform distribution, 0-10	**auto-correlation time (ACT)**	17731.09	18795.81	6089.53	5.46E+04	4.41E+03	1.66E+04	1.55E+04	1.68E+04	1.18E+04
ucldmean	Uniform distribution, 1e-12-0.1	**effective sample size (ESS)**	507.6393	478.8833	1478.111	1.65E+02	2.04E+03	5.43E+02	5.82E+02	5.36E+02	7.65E+02
rootHeight	Uniform distribution, 0-1e8										
Generations	10,000,000										

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid.logPopSize 5	skygrid.logPopSize 6	skygrid.logPopSize 7	skygrid.logPopSize 8	skygrid.logPopSize 9	skygrid.logPopSize 10	skygrid.logPopSize 11	skygrid.logPopSize 12	skygrid.logPopSize 13	skygrid.logPopSize 14	skygrid.logPopSize 15	skygrid.logPopSize 16	skygrid.logPopSize 17	skygrid.logPopSize 18	skygrid.logPopSize 19	skygrid.logPopSize 20	skygrid.logPopSize 21	skygrid.logPopSize 22
1.34E+01	1.32E+01	1.29E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.28E+01	1.28E+01	1.28E+01	1.29E+01	1.29E+01	1.29E+01	1.30E+01
2.78E-02	4.05E-02	4.00E-02	3.58E-02	2.97E-02	2.71E-02	3.01E-02	3.21E-02	3.28E-02	3.45E-02	3.86E-02	4.29E-02	4.98E-02	4.87E-02	5.03E-02	6.37E-02	6.96E-02	7.31E-02
5.32E-01	6.16E-01	6.15E-01	5.82E-01	5.71E-01	6.03E-01	6.78E-01	7.25E-01	7.78E-01	8.13E-01	8.51E-01	9.31E-01	1.00E+00	1.06E+00	1.09E+00	1.14E+00	1.22E+00	1.29E+00
2.83E-01	3.80E-01	3.79E-01	3.39E-01	3.26E-01	3.64E-01	4.60E-01	5.26E-01	6.05E-01	6.62E-01	7.24E-01	8.67E-01	1.01E+00	1.11E+00	1.20E+00	1.30E+00	1.49E+00	1.66E+00
1.33E+01	1.31E+01	1.28E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.26E+01	1.26E+01	1.27E+01	1.27E+01	1.27E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01
1.34E+01	1.31E+01	1.29E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.26E+01	1.27E+01	1.27E+01	1.28E+01	1.28E+01	1.28E+01	1.29E+01	1.29E+01	1.29E+01
1.24E+01	1.20E+01	1.16E+01	1.15E+01	1.16E+01	1.15E+01	1.12E+01	1.12E+01	1.11E+01	1.11E+01	1.10E+01	1.11E+01	1.08E+01	1.08E+01	1.09E+01	1.10E+01	1.08E+01	1.06E+01
1.43E+01	1.45E+01	1.41E+01	1.39E+01	1.39E+01	1.39E+01	1.39E+01	1.41E+01	1.42E+01	1.43E+01	1.44E+01	1.47E+01	1.48E+01	1.49E+01	1.53E+01	1.55E+01	1.57E+01	1.58E+01
2.47E+04	3.88E+04	3.81E+04	3.40E+04	2.44E+04	1.82E+04	1.77E+04	1.76E+04	1.60E+04	1.62E+04	1.86E+04	1.91E+04	2.22E+04	1.92E+04	1.91E+04	2.80E+04	2.92E+04	2.89E+04
3.64E+02	2.32E+02	2.36E+02	2.65E+02	3.69E+02	4.95E+02	5.09E+02	5.12E+02	5.63E+02	5.54E+02	4.85E+02	4.72E+02	4.06E+02	4.70E+02	4.72E+02	3.22E+02	3.08E+02	3.11E+02
23	**23**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**
1.29E+01	1.29E+01	1.29E+01	1.28E+01	1.28E+01	1.27E+01	1.26E+01	1.26E+01	1.25E+01	1.24E+01	1.23E+01	1.23E+01	1.22E+01	1.21E+01	1.20E+01	1.19E+01	1.19E+01	1.18E+01
7.78E-02	7.94E-02	8.28E-02	8.47E-02	8.88E-02	9.15E-02	9.40E-02	9.82E-02	9.71E-02	9.65E-02	9.81E-02	9.80E-02	9.61E-02	9.50E-02	9.10E-02	8.60E-02	8.47E-02	8.26E-02
1.32E+00	1.34E+00	1.37E+00	1.40E+00	1.47E+00	1.51E+00	1.57E+00	1.62E+00	1.67E+00	1.71E+00	1.77E+00	1.80E+00	1.83E+00	1.88E+00	1.91E+00	1.92E+00	1.98E+00	2.01E+00
1.73E+00	1.80E+00	1.86E+00	1.96E+00	2.16E+00	2.27E+00	2.45E+00	2.64E+00	2.79E+00	2.93E+00	3.14E+00	3.25E+00	3.35E+00	3.55E+00	3.64E+00	3.70E+00	3.90E+00	4.05E+00
1.28E+01	1.28E+01	1.28E+01	1.27E+01	1.27E+01	1.27E+01	1.26E+01	1.26E+01	1.25E+01	1.25E+01	1.24E+01	1.24E+01	1.23E+01	1.23E+01	1.22E+01	1.22E+01	1.21E+01	1.21E+01
1.29E+01	1.28E+01	n/a	1.27E+01	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
1.07E+01	1.05E+01	1.01E+01	1.00E+01	9.72E+00	9.45E+00	9.17E+00	9.19E+00	8.99E+00	9.14E+00	8.79E+00	8.40E+00	8.60E+00	8.14E+00	7.63E+00	8.06E+00	7.25E+00	7.34E+00
1.60E+01	1.60E+01	1.59E+01	1.58E+01	1.58E+01	1.58E+01	1.57E+01	1.59E+01	1.58E+01	1.60E+01	1.59E+01	1.57E+01	1.60E+01	1.57E+01	1.54E+01	1.57E+01	1.51E+01	1.52E+01
3.15E+04	3.16E+04	3.31E+04	3.29E+04	3.29E+04	3.32E+04	3.24E+04	3.29E+04	3.04E+04	2.86E+04	2.76E+04	2.66E+04	2.48E+04	2.29E+04	2.05E+04	1.80E+04	1.66E+04	1.51E+04
2.86E+02	2.85E+02	2.72E+02	2.74E+02	2.73E+02	2.71E+02	2.78E+02	2.73E+02	2.96E+02	3.15E+02	3.26E+02	3.38E+02	3.63E+02	3.93E+02	4.39E+02	5.00E+02	5.44E+02	5.94E+02

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid. logPopSize41	skygrid. logPopSize42	skygrid. logPopSize43	skygrid. logPopSize44	skygrid. logPopSize45	skygrid. logPopSize46	skygrid. logPopSize47	skygrid. logPopSize48	skygrid. logPopSize49	skygrid. logPopSize50	skygrid. cutOff	kappa1	kappa2
1.17E+01	1.16E+01	1.16E+01	1.15E+01	1.15E+01	1.14E+01	1.14E+01	1.13E+01	1.13E+01	1.13E+01	5.00E+05	7.95E+01	8.99E+01
7.95E-02	7.55E-02	7.19E-02	6.85E-02	6.69E-02	6.43E-02	6.19E-02	5.98E-02	5.65E-02	5.62E-02	n/a	1.56E+00	1.71E+00
2.07E+00	2.13E+00	2.18E+00	2.24E+00	2.29E+00	2.33E+00	2.38E+00	2.44E+00	2.49E+00	2.55E+00	n/a	1.71E+01	1.86E+01
4.28E+00	4.54E+00	4.75E+00	5.00E+00	5.23E+00	5.43E+00	5.68E+00	5.94E+00	6.19E+00	6.50E+00	n/a	2.92E+02	3.48E+02
1.20E+01	1.20E+01	1.19E+01	1.19E+01	1.18E+01	1.18E+01	1.18E+01	1.18E+01	1.17E+01	1.17E+01	n/a	7.77E+01	8.79E+01
n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	7.77E+01	8.80E+01
6.77E+00	7.01E+00	7.05E+00	6.83E+00	6.33E+00	6.40E+00	6.20E+00	6.08E+00	5.85E+00	5.51E+00	n/a	5.06E+01	5.52E+01
1.50E+01	1.53E+01	1.55E+01	1.54E+01	1.51E+01	1.53E+01	1.54E+01	1.55E+01	1.57E+01	1.56E+01	n/a	1.14E+02	1.27E+02
1.33E+04	1.13E+04	9.78E+03	8.45E+03	7.70E+03	6.86E+03	6.07E+03	5.42E+03	4.65E+03	4.37E+03	n/a	7.47E+04	7.58E+04
6.77E+02	7.95E+02	9.21E+02	1.07E+03	1.17E+03	1.31E+03	1.48E+03	1.66E+03	1.94E+03	2.06E+03	n/a	1.20E+02	1.19E+02

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

frequencies1	frequencies2	frequencies3	frequencies4	alpha	pInv	ucld.mean	ucld.stdev	meanRate	coefficientOfVariation	covariance	treeLikelihood	skygrid
3.29E-01	2.51E-01	1.33E-01	2.87E-01	1.10E+00	7.54E-01	1.83E-08	4.58E-01	1.78E-08	4.65E-01	-2.12E-02	-2.86E+04	-9.37E+02
1.80E-04	1.56E-04	1.43E-04	1.70E-04	6.12E-02	6.34E-03	2.70E-10	4.57E-03	3.09E-10	5.29E-03	1.63E-03	2.19E-01	1.63E+00
3.43E-03	3.26E-03	2.68E-03	3.30E-03	7.49E-01	5.45E-02	3.47E-09	9.31E-02	3.52E-09	1.01E-01	8.20E-02	8.43E+00	3.60E+01
1.18E-05	1.06E-05	7.20E-06	1.09E-05	5.62E-01	2.97E-03	1.20E-17	8.68E-03	1.24E-17	1.01E-02	6.72E-03	7.10E+01	1.30E+03
3.29E-01	2.51E-01	1.33E-01	2.87E-01	9.20E-01	7.64E-01	1.81E-08	4.53E-01	1.75E-08	4.56E-01	-2.33E-02	-2.86E+04	-9.44E+02
3.29E-01	2.51E-01	1.33E-01	2.87E-01	9.32E-01	7.52E-01	1.80E-08	4.49E-01	1.75E-08	4.54E-01	n/a	n/a	n/a
3.23E-01	2.44E-01	1.28E-01	2.81E-01	2.18E-01	6.43E-01	1.19E-08	2.88E-01	1.14E-08	2.78E-01	-1.77E-01	-2.86E+04	-9.96E+02
3.36E-01	2.57E-01	1.38E-01	2.93E-01	2.37E+00	8.42E-01	2.55E-08	6.49E-01	2.51E-08	6.65E-01	1.41E-01	-2.86E+04	-8.52E+02
2.47E+04	2.06E+04	2.57E+04	2.39E+04	6.01E+04	1.22E+05	5.46E+04	2.17E+04	6.91E+04	2.49E+04	3.55E+03	6.09E+03	1.84E+04
3.65E+02	4.36E+02	3.51E+02	3.76E+02	1.50E+02	7.37E+01	1.65E+02	4.15E+02	1.30E+02	3.62E+02	2.53E+03	1.48E+03	4.88E+02

Supplemental Table 5 (Continued): Priors used in Bayesian Skygrid analyses and resulting posterior statistics

Setup		Posterior summary statistic	post-erior	prior	likeli-hood	treeModel. rootHeight	skygrid. precision	skygrid. logPopSize1	skygrid. logPopSize2	skygrid. logPopSize3	skygrid. logPopSize4
Alignment	High latitude finite-dated complete mts	**mean**	-26656.5	-344.135	-26312.3	4.91E+05	2.10E+02	1.28E+01	1.28E+01	1.28E+01	1.29E+01
Temp. Calib.	Tips only	**stderr of mean**	0.8999	0.8783	0.1387	8.92E+03	5.70E+00	1.94E-02	2.02E-02	1.95E-02	1.95E-02
Site model	TN+G, 5 gamma categories	**stdev**	45.7187	45.2838	5.3054	1.93E+05	4.21E+02	5.99E-01	5.48E-01	5.01E-01	5.05E-01
Pop. model	Constant	**variance**	2090.203	2050.618	28.1475	3.72E+10	1.78E+05	3.58E-01	3.00E-01	2.51E-01	2.55E-01
Clock model	Uncorrelated lognormal	**median**	-26655	-342.529	-26312.1	4.51E+05	5.25E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01
Tip dates?	Yes; radiocarbon years before present	**geometric mean**	n/a	n/a	n/a	4.63E+05	4.69E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01
kappas	Uniform distribution, 0-500	**95% HPD Lower**	-26744.3	-430.618	-26322.5	2.15E+05	8.99E-02	1.17E+01	1.18E+01	1.19E+01	1.19E+01
alpha	Uniform distribution, 0-10	**95% HPD Upper**	-26576.4	-263.943	-26302.1	8.72E+05	9.54E+02	1.41E+01	1.40E+01	1.38E+01	1.39E+01
ucldstdev	Uniform distribution, 0-10	**auto-correlation time (ACT)**	3487.845	3386.162	6156.461	1.92E+04	1.65E+03	9.50E+03	1.23E+04	1.36E+04	1.35E+04
ucldmean	Uniform distribution, 1e-12-0.1	**effective sample size (ESS)**	2580.676	2658.171	1462.041	4.68E+02	5.47E+03	9.47E+02	7.34E+02	6.60E+02	6.69E+02
rootHeight	Uniform distribution, 0-1e8										
Generations	10,000,000										

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid.logPopSize 5	skygrid.logPopSize 6	skygrid.logPopSize 7	skygrid.logPopSize 8	skygrid.logPopSize 9	skygrid.logPopSize 10	skygrid.logPopSize 11	skygrid.logPopSize 12	skygrid.logPopSize 13	skygrid.logPopSize 14	skygrid.logPopSize 15	skygrid.logPopSize 16	skygrid.logPopSize 17	skygrid.logPopSize 18	skygrid.logPopSize 19	skygrid.logPopSize 20	skygrid.logPopSize 21	skygrid.logPopSize 22
1.28E+01	1.27E+01	1.27E+01	1.27E+01	1.28E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01
2.06E-02	1.95E-02	1.80E-02	1.59E-02	1.62E-02	1.71E-02	1.76E-02	1.82E-02	1.92E-02	2.01E-02	2.41E-02	2.58E-02	2.63E-02	2.63E-02	2.65E-02	2.73E-02	2.76E-02	2.86E-02
5.34E-01	5.47E-01	5.35E-01	5.41E-01	5.75E-01	6.30E-01	6.69E-01	6.81E-01	6.95E-01	7.16E-01	7.26E-01	7.64E-01	7.89E-01	8.10E-01	8.28E-01	8.54E-01	8.53E-01	8.75E-01
2.85E-01	2.99E-01	2.87E-01	2.92E-01	3.31E-01	3.97E-01	4.47E-01	4.64E-01	4.83E-01	5.12E-01	5.28E-01	5.84E-01	6.22E-01	6.56E-01	6.86E-01	7.30E-01	7.28E-01	7.65E-01
1.28E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01
1.28E+01	1.27E+01	1.27E+01	1.27E+01	1.28E+01	1.28E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.28E+01
1.17E+01	1.17E+01	1.17E+01	1.17E+01	1.18E+01	1.18E+01	1.17E+01	1.17E+01	1.17E+01	1.16E+01	1.17E+01	1.16E+01	1.15E+01	1.15E+01	1.13E+01	1.13E+01	1.12E+01	1.12E+01
1.38E+01	1.39E+01	1.38E+01	1.38E+01	1.39E+01	1.41E+01	1.42E+01	1.42E+01	1.44E+01	1.44E+01	1.45E+01	1.45E+01	1.46E+01	1.46E+01	1.45E+01	1.46E+01	1.45E+01	1.47E+01
1.34E+04	1.15E+04	1.02E+04	7.75E+03	7.11E+03	6.60E+03	6.27E+03	6.44E+03	6.86E+03	7.11E+03	9.93E+03	1.03E+04	1.00E+04	9.46E+03	9.19E+03	9.19E+03	9.42E+03	9.63E+03
6.74E+02	7.84E+02	8.87E+02	1.16E+03	1.27E+03	1.36E+03	1.44E+03	1.40E+03	1.31E+03	1.27E+03	9.07E+02	8.74E+02	9.00E+02	9.51E+02	9.80E+02	9.79E+02	9.56E+02	9.35E+02
23	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**
1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.27E+01	1.27E+01	1.27E+01	1.26E+01	1.26E+01	1.26E+01	1.25E+01	1.25E+01	1.25E+01	1.24E+01	1.24E+01	1.24E+01	1.24E+01	1.23E+01
2.92E-02	3.04E-02	3.16E-02	3.19E-02	3.31E-02	3.26E-02	3.34E-02	3.22E-02	3.30E-02	3.32E-02	3.35E-02	3.32E-02	3.33E-02	3.33E-02	3.34E-02	3.37E-02	3.39E-02	3.47E-02
8.85E-01	9.04E-01	9.18E-01	9.40E-01	9.82E-01	1.02E+00	1.07E+00	1.07E+00	1.12E+00	1.16E+00	1.21E+00	1.25E+00	1.29E+00	1.33E+00	1.37E+00	1.41E+00	1.44E+00	1.49E+00
7.83E-01	8.17E-01	8.43E-01	8.83E-01	9.64E-01	1.03E+00	1.14E+00	1.14E+00	1.26E+00	1.34E+00	1.46E+00	1.57E+00	1.67E+00	1.76E+00	1.88E+00	1.98E+00	2.08E+00	2.22E+00
1.28E+01	1.28E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.26E+01	1.26E+01	1.26E+01	1.26E+01	1.26E+01	1.26E+01	1.26E+01	1.25E+01	1.25E+01	1.25E+01	1.25E+01
1.28E+01	n/a	1.28E+01	1.27E+01	1.27E+01	1.26E+01	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
1.11E+01	1.11E+01	1.10E+01	1.09E+01	1.08E+01	1.06E+01	1.05E+01	1.05E+01	1.05E+01	1.04E+01	1.03E+01	1.00E+01	9.79E+00	9.74E+00	9.39E+00	9.71E+00	9.25E+00	9.33E+00
1.46E+01	1.47E+01	1.46E+01	1.46E+01	1.47E+01	1.46E+01	1.46E+01	1.46E+01	1.48E+01	1.47E+01	1.48E+01	1.46E+01	1.46E+01	1.46E+01	1.45E+01	1.50E+01	1.46E+01	1.48E+01
9.81E+03	1.02E+04	1.06E+04	1.04E+04	1.02E+04	9.31E+03	8.80E+03	8.16E+03	7.77E+03	7.41E+03	6.92E+03	6.33E+03	5.99E+03	5.65E+03	5.32E+03	5.17E+03	4.97E+03	4.87E+03
9.17E+02	8.85E+02	8.46E+02	8.66E+02	8.82E+02	9.67E+02	1.02E+03	1.10E+03	1.16E+03	1.22E+03	1.30E+03	1.42E+03	1.50E+03	1.59E+03	1.69E+03	1.74E+03	1.81E+03	1.85E+03

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid. logPopSize41	skygrid. logPopSize42	skygrid. logPopSize43	skygrid. logPopSize44	skygrid. logPopSize45	skygrid. logPopSize46	skygrid. logPopSize47	skygrid. logPopSize48	skygrid. logPopSize49	skygrid. logPopSize50	skygrid .cutOff	kappa1	kappa2
1.23E+01	1.23E+01	1.23E+01	1.23E+01	1.22E+01	1.22E+01	1.22E+01	1.22E+01	1.22E+01	1.22E+01	6.00E+05	8.41E+01	1.01E+02
3.52E-02	3.38E-02	3.43E-02	3.39E-02	3.44E-02	3.37E-02	3.28E-02	3.17E-02	3.13E-02	3.11E-02	n/a	2.67E+00	2.97E+00
1.52E+00	1.55E+00	1.59E+00	1.61E+00	1.66E+00	1.67E+00	1.67E+00	1.70E+00	1.74E+00	1.75E+00	n/a	2.19E+01	2.58E+01
2.30E+00	2.40E+00	2.54E+00	2.61E+00	2.76E+00	2.79E+00	2.79E+00	2.90E+00	3.03E+00	3.06E+00	n/a	4.80E+02	6.67E+02
1.25E+01	1.25E+01	1.25E+01	1.25E+01	1.25E+01	1.25E+01	1.25E+01	1.25E+01	1.25E+01	1.25E+01	n/a	8.13E+01	9.68E+01
n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	8.13E+01	9.75E+01
9.34E+00	9.31E+00	9.21E+00	9.13E+00	9.15E+00	9.17E+00	8.77E+00	9.02E+00	8.55E+00	8.54E+00	n/a	4.70E+01	5.71E+01
1.47E+01	1.47E+01	1.47E+01	1.48E+01	1.48E+01	1.48E+01	1.45E+01	1.50E+01	1.47E+01	1.47E+01	n/a	1.23E+02	1.51E+02
4.84E+03	4.30E+03	4.16E+03	3.97E+03	3.85E+03	3.66E+03	3.47E+03	3.11E+03	2.92E+03	2.84E+03	n/a	1.34E+05	1.19E+05
1.86E+03	2.10E+03	2.16E+03	2.27E+03	2.34E+03	2.46E+03	2.59E+03	2.89E+03	3.08E+03	3.17E+03	n/a	6.74E+01	7.56E+01

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

frequencies1	frequencies2	frequencies3	frequencies4	alpha	ucld.mean	ucld.stdev	meanRate	coefficientOfVariation	covariance	treeLikelihood	skygrid
3.29E-01	2.50E-01	1.33E-01	2.88E-01	4.65E-02	1.66E-08	3.07E-01	1.62E-08	3.01E-01	-1.77E-02	-2.63E+04	-3.04E+02
2.01E-04	1.87E-04	1.19E-04	1.67E-04	1.44E-03	2.48E-10	4.31E-03	2.45E-10	4.18E-03	1.66E-03	1.39E-01	9.15E-01
3.67E-03	3.44E-03	2.53E-03	3.39E-03	3.07E-02	5.16E-09	1.35E-01	4.91E-09	1.32E-01	1.38E-01	5.31E+00	4.75E+01
1.34E-05	1.18E-05	6.40E-06	1.15E-05	9.41E-04	2.66E-17	1.83E-02	2.41E-17	1.74E-02	1.91E-02	2.81E+01	2.25E+03
3.29E-01	2.50E-01	1.33E-01	2.88E-01	4.21E-02	1.62E-08	2.96E-01	1.58E-08	2.90E-01	-2.05E-02	-2.63E+04	-3.02E+02
3.29E-01	2.50E-01	1.33E-01	2.88E-01	3.41E-02	1.58E-08	2.71E-01	1.54E-08	2.66E-01	n/a	n/a	n/a
3.23E-01	2.43E-01	1.29E-01	2.81E-01	1.00E-03	6.41E-09	5.25E-02	6.38E-09	4.38E-02	-2.95E-01	-2.63E+04	-3.91E+02
3.37E-01	2.56E-01	1.38E-01	2.95E-01	1.01E-01	2.64E-08	5.83E-01	2.59E-08	5.64E-01	2.49E-01	-2.63E+04	-2.16E+02
2.70E+04	2.66E+04	1.98E+04	2.20E+04	1.99E+04	2.08E+04	9.15E+03	2.25E+04	9.08E+03	1.30E+03	6.16E+03	3.35E+03
3.34E+02	3.38E+02	4.54E+02	4.09E+02	4.52E+02	4.33E+02	9.84E+02	4.00E+02	9.91E+02	6.90E+03	1.46E+03	2.69E+03

Supplemental Table 5 (Continued): Priors used in Bayesian Skygrid analyses and resulting posterior statistics

Setup		Posterior summary statistic	post-erior	prior	likeli-hood	treeModel.rootHeight	skygrid.precision	skygrid.logPopSize1	skygrid.logPopSize2	skygrid.logPopSize3
Alignment	Low latitude finite-dated complete mts	**mean**	-25169	-587.47	-24581.5	1.32E+05	6.66E+01	1.17E+01	1.19E+01	1.21E+01
Temp. Calib.	Tips only	**stderr of mean**	1.0152	0.9556	0.0927	2.01E+03	2.68E+00	1.49E-02	1.19E-02	1.25E-02
Site model	HKY+G, 5 gamma categories	**stdev**	39.6305	39.2264	6.0573	4.38E+04	2.31E+02	5.56E-01	4.30E-01	4.45E-01
Pop. model	Constant	**variance**	1570.58	1538.711	36.6915	1.92E+09	5.34E+04	3.09E-01	1.85E-01	1.98E-01
Clock model	Uncorrelated lognormal	**median**	-25176.8	-595.399	-24581.3	1.23E+05	6.41E+00	1.17E+01	1.19E+01	1.21E+01
Tip dates?	Yes; radiocarbon years before present	**geometric mean**	n/a	n/a	n/a	1.26E+05	8.86E+00	1.17E+01	1.19E+01	1.21E+01
kappas	Uniform distribution, 0-1000	**95% HPD Lower**	-25234.6	-652.291	-24593.4	6.50E+04	5.02E-02	1.06E+01	1.10E+01	1.13E+01
alpha	Uniform distribution, 0-10	**95% HPD Upper**	-25080.3	-499.15	-24569.9	2.15E+05	3.48E+02	1.28E+01	1.27E+01	1.30E+01
ucldstdev	Uniform distribution, 0-10	**auto-correlation time (ACT)**	11814.82	10683.93	4215.377	3.78E+04	2.42E+03	1.30E+04	1.38E+04	1.42E+04
ucldmean	Uniform distribution, 1e-12-0.1	**effective sample size (ESS)**	1523.68	1684.961	4270.555	4.76E+02	7.42E+03	1.39E+03	1.30E+03	1.27E+03
rootHeight	Uniform distribution, 0-1e8									
Generations	20,000,000									

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid.logPopSize4	skygrid.logPopSize5	skygrid.logPopSize6	skygrid.logPopSize7	skygrid.logPopSize8	skygrid.logPopSize9	skygrid.logPopSize10	skygrid.logPopSize11	skygrid.logPopSize12	skygrid.logPopSize13	skygrid.logPopSize14	skygrid.logPopSize15	skygrid.logPopSize16	skygrid.logPopSize17	skygrid.logPopSize18	skygrid.logPopSize19	skygrid.logPopSize20	skygrid.logPopSize21
1.22E+01	1.22E+01	1.22E+01	1.22E+01	1.23E+01	1.24E+01	1.25E+01	1.26E+01	1.26E+01	1.26E+01	1.26E+01	1.25E+01	1.25E+01	1.24E+01	1.23E+01	1.22E+01	1.20E+01	1.19E+01
1.54E-02	1.66E-02	1.62E-02	1.80E-02	1.44E-02	1.23E-02	1.48E-02	2.12E-02	2.44E-02	3.19E-02	3.70E-02	4.57E-02	5.04E-02	5.20E-02	6.02E-02	6.57E-02	6.48E-02	6.81E-02
5.14E-01	5.43E-01	5.54E-01	5.82E-01	6.03E-01	6.38E-01	6.99E-01	8.14E-01	8.90E-01	9.52E-01	9.99E-01	1.07E+00	1.12E+00	1.16E+00	1.23E+00	1.31E+00	1.39E+00	1.45E+00
2.64E-01	2.94E-01	3.07E-01	3.38E-01	3.64E-01	4.07E-01	4.89E-01	6.62E-01	7.91E-01	9.07E-01	9.98E-01	1.14E+00	1.25E+00	1.36E+00	1.51E+00	1.73E+00	1.92E+00	2.11E+00
1.22E+01	1.22E+01	1.22E+01	1.22E+01	1.22E+01	1.23E+01	1.24E+01	1.24E+01	1.25E+01	1.25E+01	1.25E+01	1.24E+01	1.24E+01	1.23E+01	1.22E+01	1.21E+01	1.20E+01	1.19E+01
1.22E+01	1.22E+01	1.22E+01	1.22E+01	1.23E+01	1.24E+01	1.25E+01	1.25E+01	1.26E+01	1.26E+01	1.25E+01	1.25E+01	1.24E+01	1.23E+01	1.22E+01	n/a	n/a	n/a
1.14E+01	1.12E+01	1.11E+01	1.11E+01	1.12E+01	1.12E+01	1.12E+01	1.12E+01	1.11E+01	1.10E+01	1.08E+01	1.06E+01	1.03E+01	1.01E+01	9.73E+00	9.45E+00	9.28E+00	8.84E+00
1.33E+01	1.33E+01	1.33E+01	1.34E+01	1.35E+01	1.36E+01	1.39E+01	1.42E+01	1.44E+01	1.46E+01	1.47E+01	1.47E+01	1.48E+01	1.49E+01	1.48E+01	1.48E+01	1.49E+01	1.47E+01
1.63E+04	1.70E+04	1.54E+04	1.72E+04	1.03E+04	6.66E+03	8.05E+03	1.22E+04	1.36E+04	2.03E+04	2.47E+04	3.29E+04	3.66E+04	3.59E+04	4.30E+04	4.49E+04	3.95E+04	3.96E+04
1.11E+03	1.06E+03	1.17E+03	1.05E+03	1.75E+03	2.70E+03	2.24E+03	1.47E+03	1.33E+03	8.89E+02	7.28E+02	5.47E+02	4.91E+02	5.01E+02	4.18E+02	4.01E+02	4.56E+02	4.55E+02
22	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**
1.18E+01	1.16E+01	1.15E+01	1.13E+01	1.12E+01	1.11E+01	1.10E+01	1.09E+01	1.09E+01	1.08E+01	1.07E+01	1.07E+01	1.06E+01	1.06E+01	1.06E+01	1.05E+01	1.05E+01	1.04E+01
6.97E-02	7.09E-02	6.83E-02	6.70E-02	6.67E-02	6.54E-02	6.21E-02	5.90E-02	5.63E-02	5.45E-02	5.22E-02	5.13E-02	5.10E-02	5.26E-02	5.26E-02	4.70E-02	4.46E-02	4.36E-02
1.49E+00	1.55E+00	1.61E+00	1.67E+00	1.72E+00	1.79E+00	1.82E+00	1.85E+00	1.89E+00	1.94E+00	1.99E+00	2.05E+00	2.12E+00	2.20E+00	2.28E+00	2.32E+00	2.36E+00	2.41E+00
2.21E+00	2.41E+00	2.60E+00	2.79E+00	2.96E+00	3.22E+00	3.32E+00	3.43E+00	3.57E+00	3.77E+00	3.96E+00	4.19E+00	4.50E+00	4.82E+00	5.18E+00	5.36E+00	5.58E+00	5.83E+00
1.18E+01	1.17E+01	1.16E+01	1.15E+01	1.14E+01	1.14E+01	1.13E+01	1.12E+01	1.12E+01	1.11E+01	1.11E+01	1.10E+01	1.10E+01	1.10E+01	1.10E+01	1.10E+01	1.09E+01	1.09E+01
n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
8.74E+00	8.45E+00	8.00E+00	7.73E+00	7.73E+00	7.30E+00	6.96E+00	6.81E+00	6.68E+00	6.68E+00	6.44E+00	6.01E+00	5.85E+00	5.60E+00	5.64E+00	5.29E+00	5.20E+00	4.94E+00
1.47E+01	1.47E+01	1.44E+01	1.43E+01	1.45E+01	1.44E+01	1.41E+01	1.41E+01	1.40E+01	1.42E+01	1.41E+01	1.40E+01	1.41E+01	1.40E+01	1.44E+01	1.42E+01	1.43E+01	1.42E+01
3.97E+04	3.75E+04	3.23E+04	2.90E+04	2.70E+04	2.39E+04	2.09E+04	1.83E+04	1.60E+04	1.42E+04	1.24E+04	1.13E+04	1.04E+04	1.03E+04	9.61E+03	7.42E+03	6.43E+03	5.88E+03
4.54E+02	4.81E+02	5.58E+02	6.20E+02	6.66E+02	7.54E+02	8.61E+02	9.86E+02	1.13E+03	1.27E+03	1.45E+03	1.59E+03	1.73E+03	1.74E+03	1.87E+03	2.43E+03	2.80E+03	3.06E+03

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid. logPopSize40	skygrid. logPopSize41	skygrid. logPopSize42	skygrid. logPopSize43	skygrid. logPopSize44	skygrid. logPopSize45	skygrid. logPopSize46	skygrid. logPopSize47	skygrid. logPopSize48	skygrid. logPopSize49	skygrid. logPopSize50	skygrid.cutOff	kappa
1.04E+01	1.04E+01	1.04E+01	1.03E+01	1.03E+01	1.03E+01	1.03E+01	1.03E+01	1.03E+01	1.03E+01	1.03E+01	2.00E+05	8.75E+01
4.21E-02	3.92E-02	3.57E-02	3.45E-02	3.49E-02	3.43E-02	3.44E-02	3.47E-02	3.48E-02	3.48E-02	3.49E-02	n/a	1.11E+00
2.45E+00	2.50E+00	2.57E+00	2.62E+00	2.68E+00	2.73E+00	2.80E+00	2.87E+00	2.92E+00	2.97E+00	3.03E+00	n/a	3.53E+01
6.02E+00	6.26E+00	6.58E+00	6.88E+00	7.20E+00	7.46E+00	7.83E+00	8.22E+00	8.54E+00	8.84E+00	9.18E+00	n/a	1.25E+03
1.09E+01	1.09E+01	1.09E+01	1.09E+01	1.08E+01	1.08E+01	1.08E+01	1.08E+01	1.08E+01	1.08E+01	1.08E+01	n/a	7.97E+01
n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	8.17E+01
4.76E+00	4.28E+00	4.46E+00	4.19E+00	4.42E+00	4.19E+00	3.55E+00	3.72E+00	3.41E+00	3.57E+00	3.17E+00	n/a	3.56E+01
1.43E+01	1.40E+01	1.43E+01	1.43E+01	1.48E+01	1.48E+01	1.46E+01	1.51E+01	1.50E+01	1.56E+01	1.53E+01	n/a	1.57E+02
5.31E+03	4.42E+03	3.49E+03	3.12E+03	3.05E+03	2.83E+03	2.73E+03	2.64E+03	2.55E+03	2.47E+03	2.39E+03	n/a	1.76E+04
3.39E+03	4.07E+03	5.16E+03	5.77E+03	5.90E+03	6.35E+03	6.60E+03	6.83E+03	7.07E+03	7.29E+03	7.52E+03	n/a	1.02E+03

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

frequencies1	frequencies2	frequencies3	frequencies4	alpha	ucld.mean	ucld.stdev	meanRate	coefficientOfVariation	covariance	treeLikelihood	skygrid
3.28E-01	2.51E-01	1.34E-01	2.87E-01	3.87E-02	1.68E-08	4.99E-01	1.63E-08	5.14E-01	-1.02E-02	-2.46E+04	-5.56E+02
1.20E-04	1.09E-04	8.17E-05	1.10E-04	9.41E-04	2.23E-10	2.90E-03	2.53E-10	3.14E-03	8.80E-04	9.27E-02	9.90E-01
3.55E-03	3.42E-03	2.63E-03	3.43E-03	2.64E-02	4.61E-09	1.37E-01	4.38E-09	1.51E-01	1.03E-01	6.06E+00	4.11E+01
1.26E-05	1.17E-05	6.90E-06	1.18E-05	6.96E-04	2.13E-17	1.87E-02	1.92E-17	2.29E-02	1.05E-02	3.67E+01	1.69E+03
3.28E-01	2.51E-01	1.34E-01	2.87E-01	3.46E-02	1.65E-08	4.85E-01	1.61E-08	4.96E-01	-1.22E-02	-2.46E+04	-5.64E+02
3.28E-01	2.51E-01	1.34E-01	2.87E-01	2.82E-02	1.62E-08	4.81E-01	1.57E-08	4.93E-01	n/a	n/a	n/a
3.21E-01	2.44E-01	1.29E-01	2.81E-01	1.05E-03	8.05E-09	2.51E-01	8.15E-09	2.47E-01	-2.07E-01	-2.46E+04	-6.23E+02
3.35E-01	2.57E-01	1.39E-01	2.94E-01	8.63E-02	2.56E-08	7.75E-01	2.48E-08	8.23E-01	1.94E-01	-2.46E+04	-4.63E+02
2.06E+04	1.82E+04	1.74E+04	1.84E+04	2.29E+04	4.22E+04	8.07E+03	6.02E+04	7.76E+03	1.33E+03	4.22E+03	1.04E+04
8.75E+02	9.88E+02	1.03E+03	9.79E+02	7.87E+02	4.27E+02	2.23E+03	2.99E+02	2.32E+03	1.36E+04	4.27E+03	1.72E+03

Supplemental Table 5 (Continued): Priors used in Bayesian Skygrid analyses and resulting posterior statistics

Setup		Posterior summary statistic	post-erior	prior	likeli-hood	treeModel.rootHeight	skygrid.precision	skygrid.logPopSize1	skygrid.logPopSize2	skygrid.logPopSize3
Alignment	Periglacial (northern + GL) complete mts	**mean**	-27797.7	-571.968	-27225.7	4.39E+05	2.15E+02	1.25E+01	1.26E+01	1.27E+01
Temp. Calib.	Tips only	**stderr of mean**	0.5662	0.5691	0.0913	3.18E+03	3.84E+00	8.37E-03	8.24E-03	7.88E-03
Site model	TN+G, 5 gamma categories	**stdev**	44.0929	43.8829	6.0857	9.30E+04	4.23E+02	3.74E-01	3.26E-01	3.25E-01
Pop. model	Constant	**variance**	1944.183	1925.711	37.0358	8.65E+09	1.79E+05	1.40E-01	1.07E-01	1.05E-01
Clock model	Uncorrelated lognormal	**median**	-27797.2	-571.676	-27225.5	4.27E+05	5.23E+01	1.25E+01	1.26E+01	1.27E+01
Tip dates?	Yes; radiocarbon years before present	**geometric mean**	n/a	n/a	n/a	4.30E+05	4.87E+01	1.25E+01	1.26E+01	1.27E+01
kappas	Uniform distribution, 0-1000	**95% HPD Lower**	-27878.8	-652.98	-27237.4	2.69E+05	1.44E-01	1.17E+01	1.20E+01	1.21E+01
alpha	Uniform distribution, 0-10	**95% HPD Upper**	-27716.9	-492.196	-27213.7	6.16E+05	9.94E+02	1.33E+01	1.33E+01	1.33E+01
ucldstdev	Uniform distribution, 0-10	**auto-correlation time (ACT)**	2968.234	3027.549	4051.228	2.10E+04	1.48E+03	9.02E+03	1.15E+04	1.06E+04
ucldmean	Uniform distribution, 1e-12-0.1	**effective sample size (ESS)**	6064.886	5946.063	4443.591	8.55E+02	1.21E+04	1.99E+03	1.57E+03	1.70E+03
rootHeight	Uniform distribution, 0-1e8									
Generations	20,000,000									

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid.logPopSize4	skygrid.logPopSize5	skygrid.logPopSize6	skygrid.logPopSize7	skygrid.logPopSize8	skygrid.logPopSize9	skygrid.logPopSize10	skygrid.logPopSize11	skygrid.logPopSize12	skygrid.logPopSize13	skygrid.logPopSize14	skygrid.logPopSize15	skygrid.logPopSize16	skygrid.logPopSize17	skygrid.logPopSize18	skygrid.logPopSize19	skygrid.logPopSize20	skygrid.logPopSize21
1.28E+01	1.27E+01	1.26E+01	1.26E+01	1.26E+01	1.27E+01	1.27E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01
8.15E-03	1.01E-02	9.99E-03	7.38E-03	6.37E-03	6.84E-03	8.01E-03	8.78E-03	9.70E-03	1.00E-02	1.05E-02	1.15E-02	1.13E-02	1.19E-02	1.24E-02	1.29E-02	1.36E-02	1.38E-02
3.65E-01	4.00E-01	4.21E-01	3.94E-01	4.01E-01	4.42E-01	4.72E-01	5.07E-01	5.48E-01	5.63E-01	5.88E-01	6.30E-01	6.41E-01	6.62E-01	6.71E-01	6.98E-01	7.16E-01	7.18E-01
1.33E-01	1.60E-01	1.77E-01	1.55E-01	1.61E-01	1.95E-01	2.22E-01	2.57E-01	3.00E-01	3.17E-01	3.45E-01	3.96E-01	4.10E-01	4.39E-01	4.50E-01	4.88E-01	5.12E-01	5.15E-01
1.27E+01	1.27E+01	1.26E+01	1.26E+01	1.26E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01	1.27E+01
1.28E+01	1.27E+01	1.26E+01	1.26E+01	1.26E+01	1.27E+01	1.27E+01	1.27E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01
1.21E+01	1.20E+01	1.18E+01	1.18E+01	1.18E+01	1.18E+01	1.18E+01	1.18E+01	1.18E+01	1.18E+01	1.18E+01	1.18E+01	1.17E+01	1.16E+01	1.16E+01	1.16E+01	1.15E+01	1.15E+01
1.35E+01	1.35E+01	1.35E+01	1.34E+01	1.34E+01	1.35E+01	1.37E+01	1.37E+01	1.39E+01	1.40E+01	1.40E+01	1.41E+01	1.41E+01	1.41E+01	1.42E+01	1.43E+01	1.43E+01	1.43E+01
8.97E+03	1.14E+04	1.01E+04	6.32E+03	4.55E+03	4.32E+03	5.19E+03	5.40E+03	5.65E+03	5.68E+03	5.77E+03	6.02E+03	5.56E+03	5.84E+03	6.14E+03	6.13E+03	6.50E+03	6.65E+03
2.01E+03	1.57E+03	1.78E+03	2.85E+03	3.96E+03	4.17E+03	3.47E+03	3.33E+03	3.19E+03	3.17E+03	3.12E+03	2.99E+03	3.24E+03	3.08E+03	2.93E+03	2.94E+03	2.77E+03	2.71E+03
22	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**
1.28E+01	1.27E+01	1.27E+01	1.27E+01	1.26E+01	1.26E+01	1.26E+01	1.25E+01	1.25E+01	1.24E+01	1.24E+01	1.24E+01	1.23E+01	1.23E+01	1.23E+01	1.23E+01	1.22E+01	1.22E+01
1.51E-02	1.53E-02	1.54E-02	1.58E-02	1.62E-02	1.64E-02	1.64E-02	1.63E-02	1.60E-02	1.62E-02	1.63E-02	1.61E-02	1.59E-02	1.57E-02	1.54E-02	1.54E-02	1.56E-02	1.55E-02
7.31E-01	7.44E-01	7.58E-01	7.65E-01	7.76E-01	7.94E-01	8.09E-01	8.38E-01	8.62E-01	9.03E-01	9.38E-01	9.74E-01	1.00E+00	1.03E+00	1.07E+00	1.10E+00	1.14E+00	1.17E+00
5.34E-01	5.54E-01	5.74E-01	5.85E-01	6.02E-01	6.31E-01	6.55E-01	7.02E-01	7.42E-01	8.15E-01	8.79E-01	9.49E-01	1.01E+00	1.07E+00	1.14E+00	1.20E+00	1.30E+00	1.37E+00
1.27E+01	1.27E+01	1.27E+01	1.26E+01	1.26E+01	1.26E+01	1.26E+01	1.26E+01	1.25E+01	1.25E+01	1.25E+01	1.25E+01	1.25E+01	1.24E+01	1.24E+01	1.24E+01	1.24E+01	1.24E+01
1.27E+01	1.27E+01	1.27E+01	1.26E+01	1.26E+01	1.26E+01	1.25E+01	1.25E+01	1.24E+01	1.24E+01	1.24E+01	1.23E+01	1.23E+01	1.23E+01	1.22E+01	1.22E+01	1.22E+01	1.22E+01
1.14E+01	1.13E+01	1.12E+01	1.11E+01	1.11E+01	1.08E+01	1.08E+01	1.08E+01	1.07E+01	1.06E+01	1.05E+01	1.01E+01	1.01E+01	1.01E+01	9.97E+00	9.84E+00	9.74E+00	9.62E+00
1.43E+01	1.43E+01	1.42E+01	1.42E+01	1.43E+01	1.41E+01	1.42E+01	1.42E+01	1.43E+01	1.43E+01	1.42E+01	1.41E+01	1.41E+01	1.41E+01	1.41E+01	1.41E+01	1.42E+01	1.42E+01
7.65E+03	7.61E+03	7.44E+03	7.69E+03	7.84E+03	7.70E+03	7.36E+03	6.79E+03	6.23E+03	5.81E+03	5.43E+03	4.93E+03	4.52E+03	4.13E+03	3.75E+03	3.54E+03	3.37E+03	3.14E+03
2.35E+03	2.37E+03	2.42E+03	2.34E+03	2.30E+03	2.34E+03	2.45E+03	2.65E+03	2.89E+03	3.10E+03	3.31E+03	3.65E+03	3.98E+03	4.36E+03	4.80E+03	5.08E+03	5.34E+03	5.73E+03

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid. logPopSize40	skygrid. logPopSize41	skygrid. logPopSize42	skygrid. logPopSize43	skygrid. logPopSize44	skygrid. logPopSize45	skygrid. logPopSize46	skygrid. logPopSize47	skygrid. logPopSize48	skygrid. logPopSize49	skygrid. logPopSize50	skygrid. cutOff	kappa1
1.22E+01	1.22E+01	1.22E+01	1.22E+01	1.22E+01	1.22E+01	1.22E+01	1.21E+01	1.21E+01	1.21E+01	1.21E+01	6.00E+05	8.19E+01
1.52E-02	1.41E-02	1.42E-02	1.44E-02	1.43E-02	1.44E-02	1.45E-02	1.44E-02	1.39E-02	1.44E-02	1.47E-02	n/a	1.77E+00
1.21E+00	1.24E+00	1.28E+00	1.32E+00	1.35E+00	1.39E+00	1.42E+00	1.45E+00	1.48E+00	1.52E+00	1.55E+00	n/a	2.16E+01
1.46E+00	1.54E+00	1.63E+00	1.74E+00	1.82E+00	1.94E+00	2.00E+00	2.09E+00	2.20E+00	2.30E+00	2.39E+00	n/a	4.66E+02
1.24E+01	1.24E+01	1.24E+01	1.24E+01	1.24E+01	1.24E+01	1.24E+01	1.24E+01	1.24E+01	1.24E+01	1.24E+01	n/a	7.89E+01
n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	7.93E+01
9.49E+00	9.33E+00	9.30E+00	9.18E+00	9.09E+00	8.99E+00	8.86E+00	8.81E+00	8.75E+00	8.61E+00	8.55E+00	n/a	4.21E+01
1.42E+01	1.42E+01	1.44E+01	1.43E+01	1.44E+01	1.44E+01	1.44E+01	1.44E+01	1.45E+01	1.46E+01	1.46E+01	n/a	1.23E+02
2.84E+03	2.34E+03	2.21E+03	2.13E+03	2.03E+03	1.92E+03	1.89E+03	1.79E+03	1.59E+03	1.62E+03	1.63E+03	n/a	1.21E+05
6.33E+03	7.68E+03	8.15E+03	8.45E+03	8.86E+03	9.36E+03	9.52E+03	1.01E+04	1.14E+04	1.11E+04	1.11E+04	n/a	1.49E+02

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

kappa2	frequencies1	frequencies2	frequencies3	frequencies4	alpha	ucld.mean	ucld.stdev	meanRate	coefficientOfVariation	covariance	treeLikelihood	skygrid
9.72E+01	3.29E-01	2.50E-01	1.33E-01	2.88E-01	3.61E-02	1.71E-08	2.15E-01	1.69E-08	2.12E-01	-1.22E-02	-2.72E+04	-5.32E+02
2.07E+00	1.24E-04	1.05E-04	9.02E-05	1.22E-04	7.46E-04	1.14E-10	2.14E-03	1.15E-10	2.12E-03	8.89E-04	9.13E-02	5.92E-01
2.53E+01	3.67E-03	3.37E-03	2.61E-03	3.58E-03	2.33E-02	3.27E-09	1.01E-01	3.23E-09	9.94E-02	1.09E-01	6.09E+00	4.60E+01
6.41E+02	1.35E-05	1.14E-05	6.82E-06	1.28E-05	5.44E-04	1.07E-17	1.01E-02	1.04E-17	9.88E-03	1.18E-02	3.70E+01	2.12E+03
9.35E+01	3.29E-01	2.50E-01	1.33E-01	2.88E-01	3.34E-02	1.69E-08	2.09E-01	1.67E-08	2.08E-01	-1.30E-02	-2.72E+04	-5.31E+02
9.42E+01	3.29E-01	2.50E-01	1.33E-01	2.88E-01	2.68E-02	1.67E-08	1.83E-01	1.66E-08	1.81E-01	n/a	n/a	n/a
5.30E+01	3.22E-01	2.44E-01	1.28E-01	2.80E-01	1.05E-03	1.08E-08	3.32E-03	1.07E-08	6.48E-03	-2.19E-01	-2.72E+04	-6.18E+02
1.47E+02	3.36E-01	2.57E-01	1.39E-01	2.94E-01	7.94E-02	2.37E-08	3.92E-01	2.33E-08	3.91E-01	2.03E-01	-2.72E+04	-4.50E+02
1.21E+05	2.06E+04	1.73E+04	2.15E+04	2.10E+04	1.84E+04	2.18E+04	8.11E+03	2.30E+04	8.19E+03	1.21E+03	4.05E+03	2.98E+03
1.49E+02	8.75E+02	1.04E+03	8.38E+02	8.57E+02	9.76E+02	8.24E+02	2.22E+03	7.83E+02	2.20E+03	1.49E+04	4.44E+03	6.05E+03

Supplemental Table 5 (Continued): Priors used in Bayesian Skygrid analyses and resulting posterior statistics

Setup		Posterior summary statistic	post-erior	prior	likeli-hood	treeModel. rootHeight	skygrid. precision	skygrid. logPopSize1	skygrid. logPopSize2	skygrid. logPopSize3
Alignment	Temperate (GP, WC, MW) completes	**mean**	-24241	-378.667	-23862.4	1.38E+05	5.33E+01	1.10E+01	1.15E+01	1.18E+01
Temp. Calib.	Tips only	**stderr of mean**	1.8154	1.8176	0.0723	3.87E+03	2.83E+00	1.73E-02	4.08E-02	5.82E-02
Site model	HKY+G, 5 gamma categories	**stdev**	49.6827	49.184	5.1711	9.24E+04	2.15E+02	7.11E-01	9.51E-01	1.22E+00
Pop. model	Constant	**variance**	2468.368	2419.066	26.7408	8.53E+09	4.62E+04	5.05E-01	9.04E-01	1.50E+00
Clock model	Uncorrelated lognormal	**median**	-24251.5	-389.217	-23862.1	1.16E+05	2.05E+00	1.09E+01	1.14E+01	1.17E+01
Tip dates?	Yes; radiocarbon years before present	**geometric mean**	n/a	n/a	n/a	1.20E+05	3.00E+00	1.09E+01	1.15E+01	1.18E+01
kappas	Uniform distribution, 0-1000	**95% HPD Lower**	-24320.8	-456.449	-23873	3.77E+04	3.36E-03	9.65E+00	1.02E+01	1.03E+01
alpha	Uniform distribution, 0-10	**95% HPD Upper**	-24129.4	-267.417	-23852.8	2.67E+05	2.76E+02	1.24E+01	1.28E+01	1.36E+01
ucldstdev	Uniform distribution, 0-10	**auto-correlation time (ACT)**	60093.05	61463.97	8808.44	7.90E+04	7.80E+03	2.66E+04	8.28E+04	1.02E+05
ucldmean	Uniform distribution, 1e-12-0.1	**effective sample size (ESS)**	748.9218	732.2175	5109.304	5.69E+02	5.77E+03	1.69E+03	5.44E+02	4.42E+02
rootHeight	Uniform distribution, 0-1e8									
Generations	50,000,000									

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid.logPopSize 4	skygrid.logPopSize 5	skygrid.logPopSize 6	skygrid.logPopSize 7	skygrid.logPopSize 8	skygrid.logPopSize 9	skygrid.logPopSize 10	skygrid.logPopSize 11	skygrid.logPopSize 12	skygrid.logPopSize 13	skygrid.logPopSize 14	skygrid.logPopSize 15	skygrid.logPopSize 16	skygrid.logPopSize 17	skygrid.logPopSize 18	skygrid.logPopSize 19	skygrid.logPopSize 20	skygrid.logPopSize 21
1.18E+01	1.21E+01	1.25E+01	1.27E+01	1.28E+01	1.27E+01	1.27E+01	1.26E+01	1.26E+01	1.24E+01	1.23E+01	1.21E+01	1.19E+01	1.18E+01	1.16E+01	1.15E+01	1.14E+01	1.12E+01
2.75E-02	4.48E-02	7.14E-02	8.65E-02	8.84E-02	9.39E-02	1.10E-01	1.34E-01	1.56E-01	1.66E-01	1.90E-01	2.11E-01	2.06E-01	1.98E-01	2.08E-01	1.85E-01	1.83E-01	1.99E-01
8.99E-01	1.19E+00	1.66E+00	1.82E+00	1.94E+00	2.10E+00	2.26E+00	2.52E+00	2.76E+00	2.95E+00	3.13E+00	3.48E+00	3.48E+00	3.43E+00	3.68E+00	3.53E+00	3.67E+00	3.90E+00
8.08E-01	1.42E+00	2.77E+00	3.33E+00	3.77E+00	4.41E+00	5.11E+00	6.33E+00	7.62E+00	8.73E+00	9.77E+00	1.21E+01	1.21E+01	1.18E+01	1.36E+01	1.25E+01	1.35E+01	1.52E+01
1.17E+01	1.20E+01	1.23E+01	1.24E+01	1.25E+01	1.24E+01	1.23E+01	1.22E+01	1.21E+01	1.20E+01	1.18E+01	1.17E+01	1.16E+01	1.15E+01	1.14E+01	1.13E+01	1.13E+01	1.12E+01
1.17E+01	1.21E+01	1.24E+01	1.26E+01	1.26E+01	1.26E+01	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
1.01E+01	1.03E+01	1.02E+01	9.88E+00	9.83E+00	8.97E+00	8.92E+00	8.30E+00	8.03E+00	7.18E+00	7.00E+00	6.65E+00	5.99E+00	5.71E+00	5.04E+00	4.75E+00	4.52E+00	4.13E+00
1.34E+01	1.42E+01	1.53E+01	1.61E+01	1.70E+01	1.73E+01	1.76E+01	1.78E+01	1.85E+01	1.84E+01	1.87E+01	1.91E+01	1.88E+01	1.88E+01	1.85E+01	1.83E+01	1.87E+01	1.88E+01
4.22E+04	6.36E+04	8.29E+04	1.01E+05	9.32E+04	9.00E+04	1.06E+05	1.27E+05	1.44E+05	1.41E+05	1.66E+05	1.65E+05	1.58E+05	1.49E+05	1.43E+05	1.24E+05	1.12E+05	1.17E+05
1.07E+03	7.08E+02	5.43E+02	4.45E+02	4.83E+02	5.00E+02	4.26E+02	3.54E+02	3.13E+02	3.19E+02	2.71E+02	2.73E+02	2.85E+02	3.01E+02	3.14E+02	3.64E+02	4.01E+02	3.85E+02
22	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**
1.10E+01	1.10E+01	1.09E+01	1.07E+01	1.06E+01	1.05E+01	1.04E+01	1.03E+01	1.02E+01	1.01E+01	1.00E+01	9.98E+00	9.87E+00	9.77E+00	9.69E+00	9.65E+00	9.60E+00	9.59E+00
2.03E-01	2.24E-01	2.30E-01	2.66E-01	2.49E-01	2.37E-01	2.15E-01	1.99E-01	2.20E-01	2.09E-01	2.65E-01	2.65E-01	2.72E-01	3.07E-01	3.10E-01	2.97E-01	2.85E-01	2.60E-01
4.19E+00	4.44E+00	4.53E+00	5.21E+00	5.18E+00	5.21E+00	5.19E+00	5.13E+00	5.55E+00	5.67E+00	6.59E+00	6.65E+00	7.00E+00	7.72E+00	7.84E+00	7.78E+00	7.73E+00	7.47E+00
1.76E+01	1.97E+01	2.05E+01	2.71E+01	2.68E+01	2.72E+01	2.69E+01	2.63E+01	3.08E+01	3.21E+01	4.35E+01	4.43E+01	4.90E+01	5.96E+01	6.15E+01	6.05E+01	5.97E+01	5.58E+01
1.11E+01	1.11E+01	1.10E+01	1.10E+01	1.10E+01	1.09E+01	1.09E+01	1.09E+01	1.08E+01	1.08E+01	1.08E+01	1.08E+01	1.07E+01	1.07E+01	1.07E+01	1.07E+01	1.06E+01	1.06E+01
n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
3.64E+00	3.43E+00	2.65E+00	2.50E+00	2.74E+00	1.69E+00	1.89E+00	1.53E+00	1.60E+00	1.34E+00	7.93E-02	-4.63E-01	-1.63E-01	-3.67E-01	-9.63E-01	-6.94E-01	-1.88E+00	-2.45E+00
1.82E+01	1.82E+01	1.80E+01	1.81E+01	1.88E+01	1.80E+01	1.78E+01	1.80E+01	1.83E+01	1.81E+01	1.86E+01	1.85E+01	1.86E+01	1.90E+01	1.88E+01	1.98E+01	1.95E+01	1.92E+01
1.06E+05	1.15E+05	1.16E+05	1.17E+05	1.04E+05	9.30E+04	7.69E+04	6.77E+04	7.09E+04	6.12E+04	7.29E+04	7.13E+04	6.78E+04	7.09E+04	7.01E+04	6.56E+04	6.10E+04	5.45E+04
4.25E+02	3.93E+02	3.88E+02	3.84E+02	4.33E+02	4.84E+02	5.85E+02	6.65E+02	6.34E+02	7.35E+02	6.18E+02	6.31E+02	6.64E+02	6.35E+02	6.42E+02	6.87E+02	7.38E+02	8.25E+02

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid. logPopSize40	skygrid. logPopSize41	skygrid. logPopSize42	skygrid. logPopSize43	skygrid. logPopSize44	skygrid. logPopSize45	skygrid. logPopSize46	skygrid. logPopSize47	skygrid. logPopSize48	skygrid. logPopSize49	skygrid. logPopSize50	skygrid. cutOff	kappa
9.52E+00	9.50E+00	9.46E+00	9.47E+00	9.44E+00	9.42E+00	9.39E+00	9.39E+00	9.38E+00	9.34E+00	9.34E+00	2.70E+05	9.15E+01
2.76E-01	2.90E-01	2.86E-01	2.72E-01	2.72E-01	2.76E-01	2.89E-01	2.64E-01	2.52E-01	2.48E-01	2.32E-01	n/a	8.32E-01
7.70E+00	8.01E+00	8.38E+00	8.38E+00	8.52E+00	8.68E+00	9.01E+00	8.73E+00	8.66E+00	8.56E+00	8.50E+00	n/a	4.52E+01
5.93E+01	6.42E+01	7.03E+01	7.02E+01	7.25E+01	7.53E+01	8.11E+01	7.62E+01	7.50E+01	7.34E+01	7.22E+01	n/a	2.04E+03
1.06E+01	1.06E+01	1.06E+01	1.06E+01	1.06E+01	1.06E+01	1.06E+01	1.06E+01	1.06E+01	1.06E+01	1.06E+01	n/a	8.05E+01
n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	8.31E+01
-2.13E+00	-2.57E+00	-2.88E+00	-3.74E+00	-3.50E+00	-4.33E+00	-4.34E+00	-4.50E+00	-4.57E+00	-5.95E+00	-5.75E+00	n/a	2.95E+01
2.02E+01	2.03E+01	2.08E+01	2.06E+01	2.14E+01	2.12E+01	2.13E+01	2.18E+01	2.22E+01	2.17E+01	2.24E+01	n/a	1.80E+02
5.77E+04	5.89E+04	5.23E+04	4.73E+04	4.60E+04	4.54E+04	4.62E+04	4.13E+04	3.82E+04	3.77E+04	3.36E+04	n/a	1.52E+04
7.80E+02	7.64E+02	8.61E+02	9.51E+02	9.79E+02	9.92E+02	9.73E+02	1.09E+03	1.18E+03	1.19E+03	1.34E+03	n/a	2.95E+03

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

frequencies1	frequencies2	frequencies3	frequencies4	alpha	ucld.mean	ucld.stdev	meanRate	coefficientOfVariation	covariance	treeLikelihood	skygrid
3.28E-01	2.51E-01	1.34E-01	2.87E-01	5.27E-02	2.84E-08	9.78E-01	1.94E-08	1.05E+00	1.72E-03	-2.39E+04	-3.48E+02
8.00E-05	6.92E-05	5.21E-05	7.21E-05	1.04E-03	1.72E-09	1.19E-02	4.65E-10	1.39E-02	1.60E-03	7.23E-02	1.90E+00
3.63E-03	3.33E-03	2.59E-03	3.39E-03	4.31E-02	4.79E-08	3.61E-01	9.43E-09	4.43E-01	1.30E-01	5.17E+00	5.15E+01
1.32E-05	1.11E-05	6.71E-06	1.15E-05	1.86E-03	2.29E-15	1.30E-01	8.90E-17	1.96E-01	1.69E-02	2.67E+01	2.65E+03
3.28E-01	2.51E-01	1.34E-01	2.87E-01	4.38E-02	2.15E-08	9.18E-01	1.79E-08	9.70E-01	-1.07E-02	-2.39E+04	-3.59E+02
3.28E-01	2.51E-01	1.34E-01	2.87E-01	3.69E-02	2.22E-08	9.16E-01	1.72E-08	9.69E-01	n/a	n/a	n/a
3.21E-01	2.44E-01	1.29E-01	2.81E-01	1.01E-03	4.39E-09	3.67E-01	4.40E-09	3.59E-01	-2.38E-01	-2.39E+04	-4.27E+02
3.35E-01	2.57E-01	1.39E-01	2.94E-01	1.30E-01	6.29E-08	1.68E+00	3.81E-08	1.89E+00	2.67E-01	-2.39E+04	-2.30E+02
2.19E+04	1.94E+04	1.82E+04	2.03E+04	2.60E+04	5.79E+04	4.90E+04	1.09E+05	4.46E+04	6.83E+03	8.81E+03	6.10E+04
2.06E+03	2.33E+03	2.47E+03	2.21E+03	1.73E+03	7.77E+02	9.19E+02	4.11E+02	1.01E+03	6.59E+03	5.11E+03	7.37E+02

Supplemental Table 5 (Continued): Priors used in Bayesian Skygrid analyses and resulting posterior statistics

Setup		Posterior summary statistic	post-erior	prior	likeli-hood	treeModel. rootHeight	skygrid. precision	skygrid. logPopSize1	skygrid. logPopSize2	skygrid. logPopSize3
Alignment	Old World complete mitogenomes	**mean**	-25214.9	-225.832	-24989.1	6.39E+05	1.91E+02	1.25E+01	1.26E+01	1.25E+01
Temp. Calib.	Tips only	**stderr of mean**	1.5016	1.4606	0.1622	1.40E+04	5.25E+00	2.24E-02	2.27E-02	2.25E-02
Site model	TN+G, 5 gamma categories	**stdev**	49.9057	49.5535	4.2782	3.03E+05	3.98E+02	7.20E-01	6.65E-01	6.27E-01
Pop. model	Constant	**variance**	2490.576	2455.546	18.3028	9.16E+10	1.58E+05	5.18E-01	4.42E-01	3.93E-01
Clock model	Uncorrelated lognormal	**median**	-25212.5	-223.24	-24988.8	5.69E+05	3.88E+01	1.25E+01	1.25E+01	1.25E+01
Tip dates?	Yes; radiocarbon years before present	**geometric mean**	n/a	n/a	n/a	5.80E+05	3.22E+01	1.25E+01	1.26E+01	1.25E+01
kappas	Uniform distribution, 0-500	**95% HPD Lower**	-25307.4	-318.554	-24997.4	1.75E+05	7.70E-02	1.11E+01	1.13E+01	1.14E+01
alpha	Uniform distribution, 0-10	**95% HPD Upper**	-25125.7	-138.943	-24981.1	1.29E+06	9.37E+02	1.39E+01	1.39E+01	1.38E+01
ucldstdev	Uniform distribution, 0-10	**auto-correlation time (ACT)**	8149.905	7820.416	12938.08	1.92E+04	1.57E+03	8.70E+03	1.05E+04	1.16E+04
ucldmean	Uniform distribution, 1e-12-0.1	**effective sample size (ESS)**	1104.43	1150.962	695.6984	4.70E+02	5.74E+03	1.03E+03	8.55E+02	7.73E+02
rootHeight	Uniform distribution, 0-1e8									
Generations	10,000,000									

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid.logPopSize 4	skygrid.logPopSize 5	skygrid.logPopSize 6	skygrid.logPopSize 7	skygrid.logPopSize 8	skygrid.logPopSize 9	skygrid.logPopSize 10	skygrid.logPopSize 11	skygrid.logPopSize 12	skygrid.logPopSize 13	skygrid.logPopSize 14	skygrid.logPopSize 15	skygrid.logPopSize 16	skygrid.logPopSize 17	skygrid.logPopSize 18	skygrid.logPopSize 19	skygrid.logPopSize 20	skygrid.logPopSize 21
1.26E+01	1.26E+01	1.25E+01	1.25E+01	1.25E+01	1.26E+01	1.27E+01	1.28E+01	1.29E+01	1.30E+01	1.31E+01	1.31E+01	1.32E+01	1.32E+01	1.32E+01	1.33E+01	1.33E+01	1.33E+01
2.23E-02	2.52E-02	2.54E-02	2.14E-02	2.26E-02	3.71E-02	3.50E-02	3.83E-02	5.79E-02	5.37E-02	6.39E-02	6.74E-02	6.65E-02	7.24E-02	7.43E-02	7.48E-02	7.33E-02	7.77E-02
6.25E-01	6.61E-01	6.97E-01	7.19E-01	7.38E-01	8.53E-01	9.17E-01	1.02E+00	1.17E+00	1.27E+00	1.34E+00	1.39E+00	1.49E+00	1.57E+00	1.62E+00	1.66E+00	1.67E+00	1.70E+00
3.91E-01	4.37E-01	4.86E-01	5.17E-01	5.44E-01	7.28E-01	8.41E-01	1.04E+00	1.37E+00	1.62E+00	1.80E+00	1.94E+00	2.23E+00	2.46E+00	2.63E+00	2.74E+00	2.80E+00	2.87E+00
1.26E+01	1.25E+01	1.25E+01	1.25E+01	1.25E+01	1.25E+01	1.26E+01	1.27E+01	1.27E+01	1.28E+01	1.28E+01	1.28E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01
1.26E+01	1.25E+01	1.25E+01	1.24E+01	1.25E+01	1.26E+01	1.27E+01	1.28E+01	1.29E+01	1.29E+01	1.30E+01	1.31E+01	1.31E+01	1.31E+01	1.32E+01	1.32E+01	1.32E+01	1.32E+01
1.15E+01	1.14E+01	1.11E+01	1.12E+01	1.12E+01	1.12E+01	1.11E+01	1.12E+01	1.11E+01	1.11E+01	1.11E+01	1.12E+01	1.13E+01	1.10E+01	1.09E+01	1.10E+01	1.11E+01	1.09E+01
1.39E+01	1.39E+01	1.39E+01	1.40E+01	1.41E+01	1.41E+01	1.41E+01	1.46E+01	1.50E+01	1.53E+01	1.56E+01	1.59E+01	1.64E+01	1.62E+01	1.63E+01	1.65E+01	1.67E+01	1.68E+01
1.15E+04	1.30E+04	1.20E+04	8.00E+03	8.47E+03	1.70E+04	1.31E+04	1.27E+04	2.21E+04	1.60E+04	2.04E+04	2.11E+04	1.78E+04	1.92E+04	1.89E+04	1.83E+04	1.72E+04	1.89E+04
7.84E+02	6.91E+02	7.51E+02	1.13E+03	1.06E+03	5.28E+02	6.86E+02	7.08E+02	4.07E+02	5.61E+02	4.41E+02	4.27E+02	5.05E+02	4.69E+02	4.75E+02	4.91E+02	5.22E+02	4.76E+02

22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
1.33E+01	1.33E+01	1.33E+01	1.33E+01	1.33E+01	1.33E+01	1.32E+01	1.32E+01	1.32E+01	1.31E+01	1.31E+01	1.31E+01	1.30E+01	1.30E+01	1.29E+01	1.28E+01	1.28E+01	1.27E+01
7.64E-02	8.35E-02	8.79E-02	8.53E-02	8.40E-02	8.84E-02	8.18E-02	8.07E-02	8.07E-02	8.12E-02	8.24E-02	7.66E-02	7.31E-02	7.04E-02	6.58E-02	6.09E-02	6.44E-02	6.45E-02
1.71E+00	1.77E+00	1.81E+00	1.77E+00	1.75E+00	1.80E+00	1.79E+00	1.79E+00	1.82E+00	1.83E+00	1.84E+00	1.79E+00	1.76E+00	1.78E+00	1.74E+00	1.69E+00	1.72E+00	1.74E+00
2.91E+00	3.15E+00	3.29E+00	3.13E+00	3.06E+00	3.24E+00	3.20E+00	3.22E+00	3.30E+00	3.34E+00	3.37E+00	3.19E+00	3.09E+00	3.18E+00	3.04E+00	2.86E+00	2.97E+00	3.01E+00
1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.29E+01	1.28E+01	1.28E+01	1.28E+01	1.28E+01	1.27E+01	1.27E+01
n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
1.08E+01	1.08E+01	1.05E+01	1.09E+01	1.06E+01	1.05E+01	1.08E+01	1.05E+01	1.03E+01	9.91E+00	9.78E+00	1.01E+01	9.84E+00	9.88E+00	9.50E+00	9.59E+00	9.16E+00	9.27E+00
1.69E+01	1.71E+01	1.70E+01	1.76E+01	1.73E+01	1.73E+01	1.76E+01	1.74E+01	1.73E+01	1.70E+01	1.68E+01	1.72E+01	1.70E+01	1.70E+01	1.65E+01	1.65E+01	1.61E+01	1.62E+01
1.80E+04	1.99E+04	2.12E+04	2.09E+04	2.08E+04	2.17E+04	1.88E+04	1.82E+04	1.78E+04	1.78E+04	1.81E+04	1.66E+04	1.56E+04	1.40E+04	1.28E+04	1.17E+04	1.25E+04	1.24E+04
4.99E+02	4.52E+02	4.25E+02	4.30E+02	4.33E+02	4.15E+02	4.79E+02	4.94E+02	5.07E+02	5.07E+02	4.96E+02	5.43E+02	5.78E+02	6.41E+02	7.03E+02	7.72E+02	7.17E+02	7.25E+02

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid. logPopSize40	skygrid. logPopSize41	skygrid. logPopSize42	skygrid. logPopSize43	skygrid. logPopSize44	skygrid. logPopSize45	skygrid. logPopSize46	skygrid. logPopSize47	skygrid. logPopSize48	skygrid. logPopSize49	skygrid. logPopSize50	skygrid. cutOff	kappa1
1.27E+01	1.26E+01	1.26E+01	1.25E+01	1.25E+01	1.25E+01	1.24E+01	1.24E+01	1.23E+01	1.23E+01	1.22E+01	6.00E+05	9.81E+01
6.36E-02	6.04E-02	5.86E-02	5.35E-02	5.01E-02	4.61E-02	4.27E-02	3.84E-02	3.98E-02	4.07E-02	4.01E-02	n/a	4.55E+00
1.75E+00	1.78E+00	1.81E+00	1.84E+00	1.86E+00	1.93E+00	1.91E+00	1.93E+00	2.01E+00	2.01E+00	2.06E+00	n/a	3.79E+01
3.06E+00	3.17E+00	3.29E+00	3.38E+00	3.46E+00	3.71E+00	3.66E+00	3.71E+00	4.04E+00	4.04E+00	4.23E+00	n/a	1.44E+03
1.27E+01	1.27E+01	1.27E+01	1.26E+01	1.26E+01	1.26E+01	1.26E+01	1.26E+01	1.25E+01	1.25E+01	1.25E+01	n/a	8.89E+01
n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	9.18E+01
9.30E+00	8.93E+00	8.70E+00	8.70E+00	8.42E+00	8.58E+00	8.93E+00	8.46E+00	8.06E+00	8.03E+00	8.17E+00	n/a	4.18E+01
1.62E+01	1.59E+01	1.57E+01	1.57E+01	1.55E+01	1.57E+01	1.60E+01	1.55E+01	1.53E+01	1.52E+01	1.52E+01	n/a	1.80E+02
1.19E+04	1.04E+04	9.41E+03	7.61E+03	6.53E+03	5.17E+03	4.48E+03	3.57E+03	3.53E+03	3.69E+03	3.42E+03	n/a	1.30E+05
7.57E+02	8.69E+02	9.57E+02	1.18E+03	1.38E+03	1.74E+03	2.01E+03	2.52E+03	2.55E+03	2.44E+03	2.63E+03	n/a	6.94E+01

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

kappa2	frequencies1	frequencies2	frequencies3	frequencies4	alpha	ucld.mean	ucld.stdev	meanRate	coefficientOfVariation	covariance	treeLikelihood	skygrid
1.17E+02	3.29E-01	2.50E-01	1.33E-01	2.88E-01	4.14E-02	1.39E-08	3.50E-01	1.41E-08	3.34E-01	-4.36E-02	-2.50E+04	-1.86E+02
5.34E+00	1.86E-04	1.47E-04	1.20E-04	1.56E-04	1.41E-03	2.61E-10	8.07E-03	3.27E-10	7.76E-03	2.05E-03	1.62E-01	1.53E+00
4.48E+01	3.51E-03	3.41E-03	2.54E-03	3.39E-03	3.11E-02	5.62E-09	2.08E-01	5.89E-09	1.97E-01	1.78E-01	4.28E+00	5.20E+01
2.00E+03	1.23E-05	1.16E-05	6.44E-06	1.15E-05	9.66E-04	3.16E-17	4.31E-02	3.47E-17	3.88E-02	3.15E-02	1.83E+01	2.70E+03
1.07E+02	3.29E-01	2.50E-01	1.33E-01	2.88E-01	3.47E-02	1.31E-08	3.27E-01	1.32E-08	3.15E-01	-4.99E-02	-2.50E+04	-1.83E+02
1.10E+02	3.29E-01	2.50E-01	1.33E-01	2.87E-01	2.83E-02	1.28E-08	2.71E-01	1.29E-08	2.59E-01	n/a	n/a	n/a
4.58E+01	3.22E-01	2.43E-01	1.28E-01	2.80E-01	1.03E-03	3.93E-09	3.28E-04	3.92E-09	2.74E-04	-3.86E-01	-2.50E+04	-2.83E+02
2.05E+02	3.35E-01	2.57E-01	1.38E-01	2.94E-01	1.02E-01	2.42E-08	7.24E-01	2.47E-08	6.82E-01	2.92E-01	-2.50E+04	-9.46E+01
1.28E+05	2.53E+04	1.68E+04	2.02E+04	1.90E+04	1.86E+04	1.94E+04	1.36E+04	2.78E+04	1.40E+04	1.20E+03	1.29E+04	7.80E+03
7.04E+01	3.56E+02	5.35E+02	4.46E+02	4.73E+02	4.85E+02	4.65E+02	6.62E+02	3.24E+02	6.43E+02	7.50E+03	6.96E+02	1.15E+03

Supplemental Table 5 (Continued): Priors used in Bayesian Skygrid analyses and resulting posterior statistics

Setup		Posterior summary statistic	post-erior	prior	likeli-hood	treeModel. rootHeight	skygrid. precision	skygrid. logPopSize1	skygrid. logPopSize2	skygrid. logPopSize3
Alignment	New World complete mitogenomes	**mean**	-27257.1	-718.902	-26538.2	3.18E+05	2.87E+01	1.18E+01	1.23E+01	1.26E+01
Temp. Calib.	Tips only	**stderr of mean**	1.8092	1.8144	0.0959	7.96E+03	1.44E+00	8.60E-03	8.84E-03	1.09E-02
Site model	GTR+G, 5 gamma categories	**stdev**	36.2041	35.919	7.2198	1.22E+05	1.44E+02	3.82E-01	3.94E-01	4.58E-01
Pop. model	Constant	**variance**	1310.738	1290.171	52.1249	1.48E+10	2.07E+04	1.46E-01	1.55E-01	2.10E-01
Clock model	Uncorrelated lognormal	**median**	-27260.8	-722.765	-26538	3.04E+05	3.12E+00	1.17E+01	1.23E+01	1.26E+01
Tip dates?	Yes; radiocarbon years before present	**geometric mean**	n/a	n/a	n/a	2.96E+05	3.66E+00	1.17E+01	1.23E+01	1.26E+01
kappas	Uniform distribution, 0-500	**95% HPD Lower**	-27324.5	-785.28	-26552.9	1.05E+05	1.66E-02	1.10E+01	1.16E+01	1.18E+01
alpha	Uniform distribution, 0-10	**95% HPD Upper**	-27174.7	-636.36	-26524.6	5.32E+05	9.25E+01	1.25E+01	1.31E+01	1.35E+01
ucldstdev	Uniform distribution, 0-10	**auto-correlation time (ACT)**	89913.36	91872.56	6358.66	1.54E+05	3.62E+03	1.83E+04	1.81E+04	2.02E+04
ucldmean	Uniform distribution, 1e-12-0.1	**effective sample size (ESS)**	400.4299	391.8907	5662.199	2.34E+02	9.95E+03	1.97E+03	1.99E+03	1.78E+03
rootHeight	Uniform distribution, 0-1e8									
Generations	10,000,000									

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid.logPopSize 4	skygrid.logPopSize 5	skygrid.logPopSize 6	skygrid.logPopSize 7	skygrid.logPopSize 8	skygrid.logPopSize 9	skygrid.logPopSize 10	skygrid.logPopSize 11	skygrid.logPopSize 12	skygrid.logPopSize 13	skygrid.logPopSize 14	skygrid.logPopSize 15	skygrid.logPopSize 16	skygrid.logPopSize 17	skygrid.logPopSize 18	skygrid.logPopSize 19	skygrid.logPopSize 20	skygrid.logPopSize 21
1.31E+01	1.31E+01	1.28E+01	1.26E+01	1.25E+01	1.23E+01	1.22E+01	1.21E+01	1.20E+01	1.19E+01	1.19E+01	1.20E+01	1.20E+01	1.20E+01	1.21E+01	1.21E+01	1.21E+01	1.21E+01
2.38E-02	3.06E-02	3.47E-02	3.21E-02	3.75E-02	4.17E-02	3.77E-02	3.92E-02	3.79E-02	4.20E-02	4.47E-02	5.76E-02	8.05E-02	9.91E-02	1.15E-01	1.27E-01	1.35E-01	1.37E-01
7.18E-01	8.10E-01	8.58E-01	8.36E-01	8.91E-01	9.55E-01	9.85E-01	1.07E+00	1.10E+00	1.20E+00	1.29E+00	1.45E+00	1.58E+00	1.76E+00	1.92E+00	2.09E+00	2.19E+00	2.23E+00
5.15E-01	6.56E-01	7.36E-01	6.99E-01	7.94E-01	9.11E-01	9.71E-01	1.14E+00	1.22E+00	1.43E+00	1.67E+00	2.10E+00	2.49E+00	3.09E+00	3.70E+00	4.36E+00	4.80E+00	4.99E+00
1.30E+01	1.30E+01	1.27E+01	1.25E+01	1.24E+01	1.23E+01	1.22E+01	1.21E+01	1.20E+01	1.19E+01	1.19E+01	1.19E+01	1.20E+01	1.20E+01	1.20E+01	1.20E+01	1.21E+01	1.21E+01
1.31E+01	1.31E+01	1.28E+01	1.26E+01	1.24E+01	1.23E+01	1.22E+01	1.20E+01	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
1.19E+01	1.18E+01	1.13E+01	1.11E+01	1.08E+01	1.06E+01	1.03E+01	1.00E+01	9.77E+00	9.41E+00	9.46E+00	9.30E+00	9.06E+00	8.46E+00	8.54E+00	7.97E+00	8.01E+00	7.42E+00
1.45E+01	1.47E+01	1.45E+01	1.43E+01	1.41E+01	1.42E+01	1.41E+01	1.41E+01	1.40E+01	1.39E+01	1.44E+01	1.47E+01	1.49E+01	1.52E+01	1.59E+01	1.60E+01	1.66E+01	1.64E+01
3.96E+04	5.13E+04	5.88E+04	5.31E+04	6.37E+04	6.88E+04	5.28E+04	4.87E+04	4.26E+04	4.44E+04	4.30E+04	5.69E+04	9.37E+04	1.15E+05	1.29E+05	1.33E+05	1.37E+05	1.34E+05
9.10E+02	7.01E+02	6.12E+02	6.78E+02	5.65E+02	5.23E+02	6.82E+02	7.40E+02	8.46E+02	8.12E+02	8.38E+02	6.33E+02	3.84E+02	3.14E+02	2.79E+02	2.71E+02	2.63E+02	2.68E+02
22	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**
1.21E+01	1.21E+01	1.21E+01	1.20E+01	1.20E+01	1.19E+01	1.18E+01	1.18E+01	1.17E+01	1.17E+01	1.16E+01	1.16E+01	1.15E+01	1.15E+01	1.14E+01	1.14E+01	1.13E+01	1.13E+01
1.38E-01	1.35E-01	1.37E-01	1.30E-01	1.23E-01	1.18E-01	1.12E-01	1.11E-01	1.09E-01	1.03E-01	9.57E-02	8.72E-02	8.47E-02	7.98E-02	4.66E-02	4.60E-02	4.71E-02	4.13E-02
2.30E+00	2.41E+00	2.49E+00	2.50E+00	2.66E+00	2.68E+00	2.72E+00	2.79E+00	2.89E+00	2.95E+00	3.03E+00	3.16E+00	3.23E+00	3.40E+00	3.58E+00	3.66E+00	3.75E+00	3.76E+00
5.31E+00	5.79E+00	6.19E+00	6.24E+00	7.05E+00	7.17E+00	7.42E+00	7.78E+00	8.34E+00	8.72E+00	9.17E+00	9.98E+00	1.04E+01	1.15E+01	1.28E+01	1.34E+01	1.40E+01	1.41E+01
1.21E+01	1.21E+01	1.21E+01	1.21E+01	1.20E+01	1.20E+01	1.20E+01	1.20E+01	1.19E+01	1.19E+01	1.19E+01	1.18E+01	1.18E+01	1.18E+01	1.17E+01	1.17E+01	1.17E+01	1.17E+01
n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
7.21E+00	7.27E+00	7.12E+00	6.65E+00	6.32E+00	6.60E+00	5.97E+00	5.69E+00	5.64E+00	5.66E+00	5.21E+00	4.87E+00	5.30E+00	4.67E+00	5.16E+00	4.26E+00	4.65E+00	3.72E+00
1.66E+01	1.70E+01	1.71E+01	1.68E+01	1.69E+01	1.76E+01	1.70E+01	1.70E+01	1.72E+01	1.76E+01	1.73E+01	1.73E+01	1.79E+01	1.75E+01	1.81E+01	1.76E+01	1.83E+01	1.75E+01
1.29E+05	1.13E+05	1.09E+05	9.68E+04	7.74E+04	7.00E+04	6.07E+04	5.73E+04	5.16E+04	4.39E+04	3.59E+04	2.74E+04	2.48E+04	1.98E+04	6.09E+03	5.70E+03	5.68E+03	4.35E+03
2.80E+02	3.20E+02	3.30E+02	3.72E+02	4.65E+02	5.14E+02	5.93E+02	6.28E+02	6.98E+02	8.19E+02	1.00E+03	1.31E+03	1.45E+03	1.81E+03	5.91E+03	6.32E+03	6.34E+03	8.28E+03

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

skygrid. logPopSize40	skygrid. logPopSize41	skygrid. logPopSize42	skygrid. logPopSize43	skygrid. logPopSize44	skygrid. logPopSize45	skygrid. logPopSize46	skygrid. logPopSize47	skygrid. logPopSize48	skygrid. logPopSize49	skygrid. logPopSize50	skygrid. cutOff	ac	ag	at
1.13E+01	1.12E+01	1.12E+01	1.12E+01	1.11E+01	1.11E+01	1.11E+01	1.11E+01	1.11E+01	1.11E+01	1.11E+01	5.00E+05	2.06E-02	8.62E-01	1.40E-02
4.49E-02	4.27E-02	4.07E-02	4.07E-02	4.31E-02	4.21E-02	4.09E-02	4.37E-02	4.36E-02	4.61E-02	4.81E-02	n/a	1.23E-04	1.24E-03	1.01E-04
3.83E+00	3.91E+00	3.97E+00	4.06E+00	4.16E+00	4.29E+00	4.37E+00	4.52E+00	4.62E+00	4.76E+00	4.89E+00	n/a	6.91E-03	8.40E-02	5.31E-03
1.47E+01	1.53E+01	1.58E+01	1.65E+01	1.73E+01	1.84E+01	1.91E+01	2.04E+01	2.13E+01	2.26E+01	2.40E+01	n/a	4.77E-05	7.06E-03	2.82E-05
1.17E+01	1.16E+01	1.16E+01	1.16E+01	1.16E+01	1.16E+01	1.16E+01	1.16E+01	1.16E+01	1.15E+01	1.15E+01	n/a	1.98E-02	8.58E-01	1.33E-02
n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	1.94E-02	8.58E-01	1.30E-02
3.26E+00	3.23E+00	2.93E+00	3.20E+00	2.53E+00	2.11E+00	2.44E+00	1.90E+00	1.93E+00	1.41E+00	1.27E+00	n/a	8.78E-03	7.02E-01	5.10E-03
1.76E+01	1.77E+01	1.79E+01	1.84E+01	1.80E+01	1.79E+01	1.85E+01	1.85E+01	1.90E+01	1.89E+01	1.90E+01	n/a	3.48E-02	1.03E+00	2.52E-02
4.95E+03	4.30E+03	3.78E+03	3.61E+03	3.86E+03	3.47E+03	3.16E+03	3.37E+03	3.20E+03	3.37E+03	3.48E+03	n/a	1.14E+04	7.90E+03	1.30E+04
7.27E+03	8.37E+03	9.54E+03	9.96E+03	9.33E+03	1.04E+04	1.14E+04	1.07E+04	1.12E+04	1.07E+04	1.03E+04	n/a	3.16E+03	4.56E+03	2.78E+03

Supplemental Table 5 (Continued): Bayesian Skygrid analyses posterior statistics

cg	gt	frequencies1	frequencies2	frequencies3	frequencies4	alpha	ucld.mean	ucld.stdev	meanRate	coefficientOfVariation	covariance	treeLikelihood	skygrid
5.40E-04	9.49E-03	3.29E-01	2.51E-01	1.34E-01	2.87E-01	3.53E-02	1.98E-08	5.04E-01	1.93E-08	5.13E-01	-2.97E-02	-2.65E+04	-7.06E+02
1.33E-04	1.76E-04	8.10E-05	7.67E-05	6.11E-05	8.40E-05	5.40E-04	1.59E-10	8.09E-03	1.95E-10	1.05E-02	6.62E-04	9.59E-02	1.46E+00
1.81E-03	6.98E-03	3.55E-03	3.25E-03	2.60E-03	3.46E-03	2.36E-02	4.32E-09	1.30E-01	4.31E-09	1.47E-01	9.64E-02	7.22E+00	3.70E+01
3.28E-06	4.88E-05	1.26E-05	1.06E-05	6.77E-06	1.20E-05	5.55E-04	1.87E-17	1.68E-02	1.85E-17	2.15E-02	9.30E-03	5.21E+01	1.37E+03
1.50E-06	7.84E-03	3.29E-01	2.51E-01	1.34E-01	2.87E-01	3.20E-02	1.95E-08	4.87E-01	1.89E-08	4.87E-01	-3.39E-02	-2.65E+04	-7.11E+02
9.41E-08	7.03E-03	3.29E-01	2.51E-01	1.34E-01	2.87E-01	2.60E-02	1.93E-08	4.88E-01	1.88E-08	4.94E-01	n/a	n/a	n/a
2.27E-19	9.89E-06	3.22E-01	2.44E-01	1.29E-01	2.80E-01	1.08E-03	1.18E-08	2.77E-01	1.14E-08	2.72E-01	-2.07E-01	-2.66E+04	-7.76E+02
3.26E-03	2.30E-02	3.36E-01	2.57E-01	1.39E-01	2.94E-01	7.87E-02	2.84E-08	7.76E-01	2.81E-08	8.27E-01	1.68E-01	-2.65E+04	-6.23E+02
1.96E+05	2.28E+04	1.87E+04	2.00E+04	1.99E+04	2.12E+04	1.89E+04	4.90E+04	1.40E+05	7.42E+04	1.85E+05	1.70E+03	6.36E+03	5.57E+04
1.84E+02	1.58E+03	1.92E+03	1.80E+03	1.81E+03	1.70E+03	1.91E+03	7.35E+02	2.56E+02	4.85E+02	1.95E+02	2.12E+04	5.66E+03	6.46E+02

3.12 REFERENCES

Agenbroad LD (2005) North American Proboscideans: Mammoths: The state of Knowledge, 2003. *Quaternary International* **126**, 73-92.

Bandelt HJ (2008) Clock debate: when times are a-changin': time dependency of molecular rate estimates: tempest in a teacup. *Heredity (Edinb)* **100**, 1-2.

Barnes I, Shapiro B, Lister A*, et al.* (2007) Genetic structure and extinction of the woolly mammoth, Mammuthus primigenius. *Curr Biol* **17**, 1072-1075.

Bos KI, Schuenemann VJ, Golding GB*, et al.* (2011) A draft genome of Yersinia pestis from victims of the Black Death. *Nature* **478**, 506-510.

Carpenter ML, Buenrostro JD, Valdiosera C*, et al.* (2013) Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *Am J Hum Genet* **93**, 852-864.

Corner RG (1977) A late Pleistocene-Holocene vertebrate fauna from Red Willow County, Nebraska. *Transactions of the Nebraska Academy of Science* **4**, 77-93.

Dabney J, Knapp M, Glocke I*, et al.* (2013) 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.

Daley T, Smith AD (2013) Predicting the molecular complexity of sequencing libraries. *Nat Methods* **10**, 325-327.

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

Debruyne R, Chu G, King CE*, et al.* (2008) Out of America: Ancient DNA evidence for a new world origin of late quaternary woolly mammoths. *Current Biology* **18**, 1320-1326.

Debruyne R, Poinar HN (2009) Time Dependency of Molecular Rates in Ancient DNA Data Sets, A Sampling Artifact? *Syst Biol* **58**, 348-359.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969-1973.

Emerson BC (2007) Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. *Syst Biol* **56**, 337-345.

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

Enk J, Rouillard JM, Poinar H (2013) Quantitative PCR as a predictor of aligned ancient DNA read counts following targeted enrichment. *Biotechniques* **55**, 300-309.

Feranec RS, Kozlowski AL (2012) New Ams Radiocarbon Dates from Late Pleistocene Mastodons and Mammoths in New York State, USA. *Radiocarbon* **54**, 275-279.

Fisher DC (2009) Paleobiology and extinction of proboscideans in the Great Lakes region of North America. In: *American megafaunal extinctions at the end of the Pleistocene*, pp. 55-75. Springer, Netherlands.

Fisher DC, Tikhonov AN, Kosintsev PA*, et al.* (2012) Anatomy, death, and preservation of a woolly mammoth (Mammuthus primigenius) calf, Yamal Peninsula, northwest Siberia. *Quaternary International* **255**, 94-105.

Gilbert MTP, Drautz DI, Lesk AM*, et al.* (2008) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proc Natl Acad Sci U S A* **105**, 8327-8332.

Gill MS, Lemey P, Faria NR*, et al.* (2013) Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Molecular Biology and Evolution* **30**, 713-724.

Gillette DD, Madsen DB (1993) The Columbian Mammoth, Mammuthus-Columbi, from the Wasatch Mountains of Central Utah. *Journal of Paleontology* **67**, 669-680.

Hagelberg E, Thomas MG, Cook CE*, et al.* (1994) DNA from Ancient Mammoth Bones. *Nature* **370**, 333-334.

Harington C (1984) Mammoths, bison and time in North America. *Developments in Palaeontology and Stratigraphy* **7**, 299-309.

Haynes CV, Surovell TA, Hodgins GWL (2013) The UP Mammoth Site, Carbon County, Wyoming, USA: More Questions than Answers. *Geoarchaeology-an International Journal* **28**, 99-111.

Hills LV, Harington CR (2003) New radiocarbon dates for Columbian mammoth and Mexican horse from southern Alberta and the Lateglacial regional fauna. *Quaternary Science Reviews* **22**, 1521-1523.

Ho SY, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* **22**, 1561-1568.

Ho SY, Shapiro B, Phillips MJ, Cooper A, Drummond AJ (2007) Evidence for time dependency of molecular rate estimates. *Syst Biol* **56**, 515-522.

Ho SYW, Saarma U, Barnett R, Haile J, Shapiro B (2008) The Effect of Inappropriate Calibration: Three Case Studies in Molecular Ecology. *PLoS One* **3**.

Holen SR (2006) Taphonomy of two last glacial maximum mammoth sites in the central Great Plains of North America: A preliminary report on La Sena and Lovewell. *Quaternary International* **142**, 30-43.

Höss M, Paabo S, Vereshchagin NK (1994) Mammoth DNA-Sequences. *Nature* **370**, 333-333.

Hoyle BG, Fisher DC, Borns HW*, et al.* (2004) Late Pleistocene mammoth remains from Coastal Maine, USA. *Quaternary Research* **61**, 277-288.

Hurvich CM, Tsai CL (1989) Regression and Time-Series Model Selection in Small Samples. *Biometrika* **76**, 297-307.

Kearse M, Moir R, Wilson A*, et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

Kircher M (2012) Analysis of high-throughput ancient DNA sequencing data. *Methods Mol Biol* **840**, 197-228.

Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* **40**, e3.

Krause J, Dear PH, Pollack JL*, et al.* (2006) Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* **439**, 724-727.

Kurtén Br, Anderson E (1980) *Pleistocene mammals of North America* Columbia University Press, New York.

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

Lister A (1996) Evolution and taxonomy of Eurasian mammoths. *The Proboscidea*, 203-213.

Lister A, Bahn PG (2007) *Mammoths : giants of the ice age*, Rev. edn. University of California Press, Berkeley, Calif.

Lister AM, Sher AV (2001) The origin and evolution of the woolly mammoth. *Science* **294**, 1094-1097.

Lubinski PM, McCutcheon PT, Lillquist K*, et al.* (2009) Possible Lithic Artifacts from 2005–07 Excavations at the Wenas Creek Mammoth Site. *Current Research in the Pleistocene* **26**, 85-86.

MacFadden BJ, Hulbert RC (2009) Calibration of mammoth (Mammuthus) dispersal into North America using rare earth elements of Plio-Pleistocene mammals from Florida. *Quaternary Research* **71**, 41-48.

Maglio VJ (1973) Origin and evolution of the Elephantidae. *Transactions of the American Philosophical Society*, 1-149.

Mandel RD (2004) Investigations at the Chalk Rock Site, Brookings County, South Dakota. In: *Odyssey Archaeological Research Fund Report of Investigations, Summer and Fall, 2014*, pp. 18-23. Kansas Geological Survey, University of Kansas, Lawrence, Kansas.

McDaniel G, Jefferson GT (2003) Mammuthus meridionalis (Proboscidea: Elephantidae) from the Borrego Badlands of Anza-Borrego Desert State Park, California: phylogenetic and biochronologic implications. *Deinsea* **9**, 239-252.

Meyer M, Fu QM, Aximu-Petri A*, et al.* (2014) A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403-+.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**, pdb prot5448.

Minh BQ, Nguyen MA, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* **30**, 1188-1195.

Mitchell D, Willerslev E, Hansen A (2005) Damage and repair of ancient DNA. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **571**, 265-276.

Mokry M, Feitsma H, Nijman IJ*, et al.* (2010) Accurate SNP and mutation detection by targeted custom microarray-based genomic enrichment of short-fragment sequencing libraries. *Nucleic Acids Res* **38**, e116.

Neph S, Kuehn MS, Reynolds AP*, et al.* (2012) BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919-1920.

Nystrom V, Dalen L, Vartanyan S*, et al.* (2010) Temporal genetic change in the last remaining population of woolly mammoth. *Proceedings of the Royal Society B-Biological Sciences* **277**, 2331-2337.

Orlando L, Ginolhac A, Zhang GJ*, et al.* (2013) Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74-+.

Ozawa T, Hayashi S, Mikhelson VM (1997) Phylogenetic position of Mammoth and Steller's sea cow within Tethytheria demonstrated by mitochondrial DNA sequences. *J Mol Evol* **44**, 406-413.

Palkopoulou E, Dalen L, Lister AM*, et al.* (2013) Holarctic genetic structure and range dynamics in the woolly mammoth. *Proc Biol Sci* **280**, 20131910.

Parmalee P (1967) Castoroides and Cervalces from central Illinois. Illinois State Academy of Science. *Transactions* **60**, 127-130.

Pasenko M, Schubert B (2004) Mammuthus jeffersonii (Proboscidea, Mammalia) from Northern Illinois. *PaleoBios* **24**, 19-24.

Pigati JS, Miler IM, Johnson KR*, et al.* (2014) Geologic setting and stratigraphy of the Ziegler Reservoir fossil site, Snowmass Village, Colorado. *Quaternary Research* **In Press**.

Ramsey CB, Lee S (2013) Recent and Planned Developments of the Program Oxcal. *Radiocarbon* **55**, 720-730.

Reimer PJ, Bard E, Bayliss A*, et al.* (2013) Intcal13 and Marine13 Radiocarbon Age Calibration Curves 0-50,000 Years Cal Bp. *Radiocarbon* **55**, 1869-1887.

Roca AL, Georgiadis N, O'Brien SJ (2005) Cytonuclear genomic dissociation in African elephant species. *Nat Genet* **37**, 96-100.

Rogaev EI, Moliaka YK, Malyarchuk BA*, et al.* (2006) Complete mitochondrial genome and phylogeny of Pleistocene mammoth Mammuthus primigenius. *Plos Biology* **4**, e73.

Rohland N, Malaspinas AS, Pollack JL*, et al.* (2007) Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup. *Plos Biology* **5**, 1663-1671.

Sanders WJ (2010) *Cenozoic mammals of Africa* University of California Press, Berkeley.

Saunders JJ, Grimm EC, Widga CC*, et al.* (2010) Paradigms and proboscideans in the southern Great Lakes region, USA. *Quaternary International* **217**, 175-187.

Schubert M, Ginolhac A, Lindgreen S*, et al.* (2012) Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178.

Skeels MA (1962) *The mastodons and mammoths of Michigan* Michigan Academy of Science, Arts, and Letters.

Smith CI, Chamberlain AT, Riley MS, Stringer C, Collins MJ (2003) The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of Human Evolution* **45**, 203-217.

Stuart AJ, Sulerzhitsky LD, Orlova LA, Kuzmin YV, Lister AM (2002) The latest woolly mammoths (Mammuthus primigenius Blumenbach) in Europe and Asia: a review of the current evidence. *Quaternary Science Reviews* **21**, 1559-1569.

Surovell TA, Waguespack NM (2008) How many elephant kills are 14? Clovis mammoth and mastodon kills in context. *Quaternary International* **191**, 82-97.

Tamura K, Nei M (1993) Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial-DNA in Humans and Chimpanzees. *Molecular Biology and Evolution* **10**, 512-526.

Todd N, Roth V (1996) Origin and radiation of the Elephantidae. *The Proboscidea: Evolution and palaeoecology of elephants and their relatives*, 193-202.

Voorhies MR, Corner RG (1984) The Crappie Hole Site: a concentration of spirally-fractured Rancholabrean mammal bones in western Nebraska. *Current Research* **1**, 53-54.

Wagner DM, Klunk J, Harbeck M*, et al.* (2014) Yersinia pestis and the Plague of Justinian 541-543 AD: a genomic analysis. *Lancet Infect Dis*.

Waters MR, Stafford TW (2007) Redefining the age of Clovis: implications for the peopling of the Americas. *Science* **315**, 1122-1126.

Woodhams M (2006) Can deleterious mutations explain the time dependency of molecular rate estimates? *Molecular Biology and Evolution* **23**, 2271-2273.

Chapter 4

Ancient whole genome enrichment using baits built from modern DNA

JACOB ENK, ALISON DEVAULT, MELANIE KUCH, YUSUF MURGHA, JEAN-MARIE ROUILLARD, and HENDRIK POINAR
Molecular Biology and Evolution 2014, doi: 10.1093/molbev/msu074

4.1 PREFACE

Nuclear DNA sequences, and especially complete nuclear genomes, from ancient remains have proven incredibly powerful in elucidating evolutionary and population processes in extinct and extant organisms. Indeed several of the implications of the mitogenomic phylogeography we reconstruct in previous chapters are best explored with mammoth nuclear gene sequences. Our initial attempts at capturing and sequencing nuclear loci from mammoths proved only marginally successful, with only a few regions from a very small number attempted specimens characterized successfully. These results made it clear that not only are DNA extraction and library preparation scales major issues for ancient nuclear capture, but targeting only a small number of loci is problematic especially for mammoths because it is not known which regions might vary intraspecifically. Here we explore the use of complete nuclear genome capture, which if successful could drastically reduce the cost of sequencing complete ancient genomes from poorly-preserved remains with high exogenous DNA content. We do this with a cost-effective method of generating baits directly from modern elephant DNA rather than by computationally-directed oligonucleotides synthesis. After enriching several well- and poorly-preserved mammoth genomes, we demonstrate excellent potential in the approach and outline clear avenues for improving it.

4.2 ABSTRACT

We report metrics from complete genome capture of nuclear DNA from extinct mammoths using biotinylated RNAs transcribed from an Asian elephant DNA extract. Enrichment of the nuclear genome ranged from 1.06- to 18.65-fold, to an apparent maximum threshold of about 80% on-target. This projects an order of magnitude less costly complete genome sequencing from long-dead organisms, even when a reference genome is unavailable for bait design.

4.3 MAIN TEXT

Targeted enrichment through hybridization capture with short oligonucleotide 'baits' has proven to be an effective way to select key endogenous constituents embedded within a complex background of undesired DNA, like that commonly found in environmental and ancient specimens (Briggs et al. 2009; Burbano et al. 2010; Bos et al. 2011). However, capturing whole genomes using sequence-directed bait synthesis (Gnirke et al. 2009; Hodges et al. 2009) is impossible from organisms for which no closely-related complete genome sequence is available, and presently very expensive even when a reference genome is available. Recently, Carpenter et al. (2013) demonstrated whole genome capture using modern DNA as bait synthesis template, but this technique has yet to be tested with extinct organisms. In an effort to unlock the unique information only available in highly damaged but nonetheless recoverable genomic DNA from paleontological remains of extinct species, here we fill this gap by using reference-free bait synthesis for capturing the whole genome of an extinct species. As proof of concept, we targeted mammoths (*Mammuthus* sp.) because the genome of the closely related African elephant (*Loxodonta africana*), has been completely sequenced to appreciable depth, allowing thorough characterization of post-enrichment metrics.

MYcroarray (Ann Arbor, MI, USA) generated a genome-wide biotinylated RNA bait set using an Asian elephant (*Elephas maximus*, the closest extant relative of mammoths (Rohland et al. 2007; Rohland et al. 2010)) DNA extract

as template. We then used these baits in two rounds of in-solution hybridization capture of Illumina sequencing libraries generated from four Pleistocene-era mammoths (*M. primigenius* and *M. columbi*) and an extinct ground sloth (*Mylodon darwinii*). We then sequenced the enriched and non-enriched libraries on an Illumina® HiSeq™ 1500 flowcell. Following sequence curation, we aligned the reads to repeats-masked and non-masked versions of the African elephant nuclear genome, a mammoth (*M. columbi*) mitochondrial genome, the two-toed sloth (*Choloepus hoffmanni*) nuclear genome supercontigs, and a mylodon mitochondrial genome. Mapping rates are summarized in Figure 1 and tabulated in Table S5.

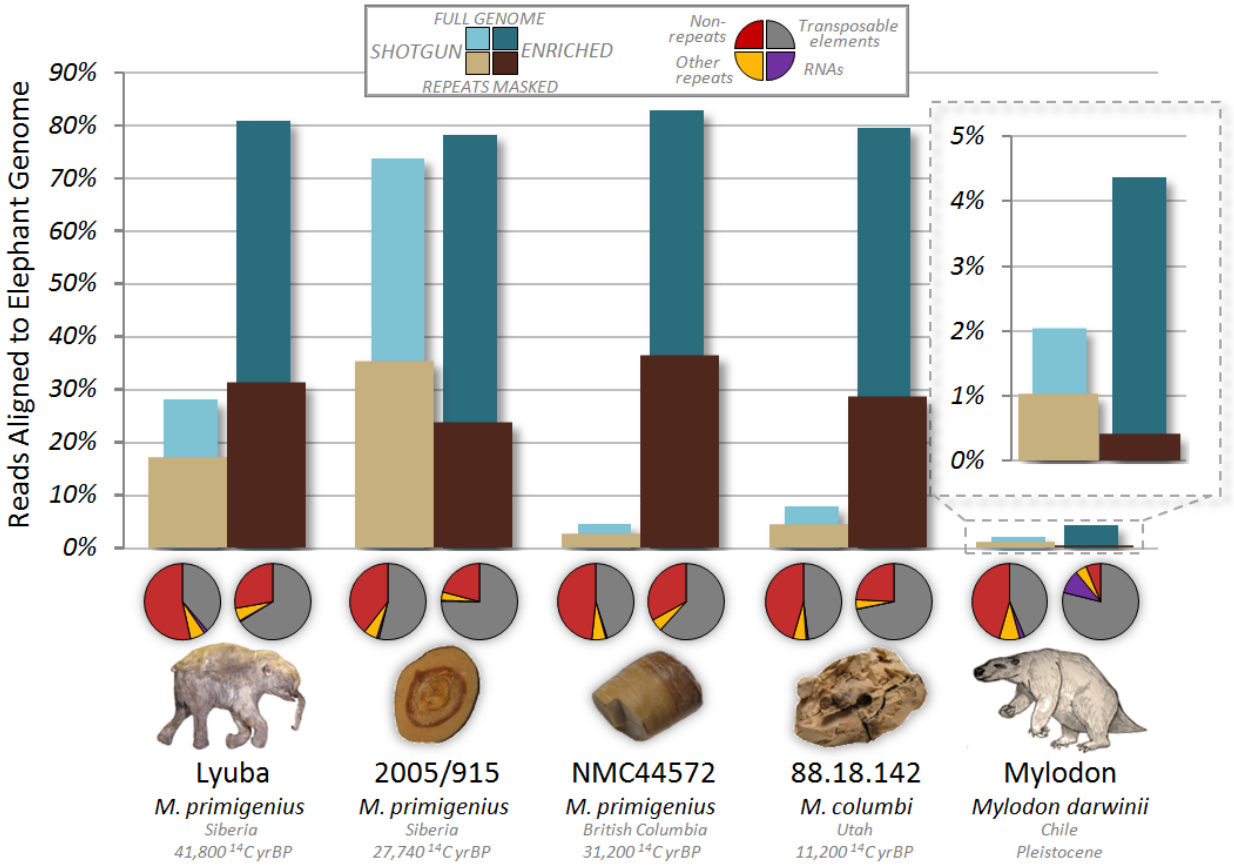


Figure 1: Raw read percentages alignable to the *Loxodonta africana* genome. Percentages are averages calculated from both replicates. Pie graphs represent the locus composition of unmasked alignments before (left pies) and after (right pies) enrichment. Transposable elements include short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), and long terminal repeats (LTRs). RNAs include ribosomal RNAs (rRNAs), transfer RNAs (rRNAs), small cytoplasmic RNAs (scRNAs), small nuclear RNAs (snRNAs), and small recognition particle RNAs (srpRNAs). Photograph of Lyuba reprinted with permission from Elsevier.

Prior to capture, the mammoth libraries yielded 4.4 to 73.8% of total raw reads alignable to the elephant genome. After capture, regardless of these starting proportions, between 78.3 and 82.8% of reads from each mammoth could be aligned, representing enrichment rates of 1.06 to 18.65-fold in raw read counts. Some biases in the capture and/or sequencing were clearly apparent. First, we observed a significant increase in average insert lengths of alignable reads (Student's t, $P = 0.001$) of 19.7bp on average (35.5%, Figure S1). The most apparent bias was an increase in repeat elements: enrichment rates of non-repeat regions of the genome were 1.84 to 14.32-fold, though the most endogenous-rich mammoth actually decreased in non-repeat content by 33%. This over-enrichment of repeats was particularly high for LINEs (2.1 – 36.5-fold enrichment, replicates averaged), which comprise about 28% of the elephant genome but 53.4 to 63.6% of all elephant-alignable reads following capture. Simple repeats also enriched substantially (1.2 to 26.0-fold) as did SINEs in some cases (0.6 to 11.8-fold). The

mylodon library, on the other hand, enriched on average 9.9-fold for rRNAs using this elephant bait set, compared to a 2.1-fold enrichment for elephant-alignable reads overall (but with a 61% drop in reads aligning to non-repeat regions). This is not unexpected since rRNAs are highly conserved within vertebrates (Mindell, Honeycutt 1990), and despite 100 million years of separation between Xenarthrans and Afrotherians (Hallstrom et al. 2007) such regions were still captured under our hybridization conditions. In addition, GC content of elephant-alignable reads dropped significantly following enrichment (Student's t, $P = 0.002$) from an average of 42.8 to 40.5%, but in masked reference-aligned reads it did not shift ($P = 0.84$). This suggests that it is the repeat enrichment driving this signal, and that our hybridization conditions were not stringent enough to generate a significant GC bias overall.

The clear over-representation of transposable elements and other repeats is probably driven by the more rapid reassociation rate of repeat sequences, the same property often exploited to prepare repeat-enriched $C_o t$-1 DNA from whole genomic DNA (Britten, Kohne 1968). Another possibility is that post-enrichment amplification inflates repeat representation through a higher PCR jumping rate. As such, we suspect that this repeat bias can be reduced by using amplification-free techniques following hybridization capture (Aird et al. 2011; Kozarewa, Turner 2011; Sawyer et al. 2012), by removing rapidly-renaturing DNA prior to bait manufacture, and/or by using target organism-specific blocking DNAs (e.g., elephant $C_o t$-1 in this case) during capture itself. For example, capture experiments of human genomes reported by Carpenter et al. (2013) included human $C_o t$-1 as a blocker and show a much lower rate of repeat-region enrichment than observed here.

Our observed average 5.9-fold enrichment of non-repeat sections of the mammoth nuclear genome from proportions as low as 2.5% endogenous is promising, with clear potential for improvement. Duplication rates and derived complexity projections (Daley, Smith 2013) (Figure S2) from the enriched libraries show that even without preventing capture or sequencing of repeats, 2.2 to 19.0% of the non-repeat nuclear genome of these specimens could be sequenced in ~1 Hiseq lane from about 10% of a single enriched library prepared from ~50mg of fossil tissue. Though other biases in post-capture coverage still need to be characterized, this versatile, fast, and inexpensive bait manufacturing technique shows great potential for resolving complete ancient genomes when DNA from a closely-related organism is available.

4.4 COMPETING FINANCIAL INTERESTS

YM and JMR are employees of MYcroarray, the company providing baits for these experiments.

4.5 ACKNOWLEDGMENTS

Specimens were obtained from the Bowmanville Zoo in Ontario, Canada; D. Fisher of the University of Michigan; B. Buigues of Cerpoplex; D. Guthrie and the National Museums of Canada; the Utah State University Eastern Prehistoric Museum; and G. MacDonald of the National Park Service, USA. C. Schwarz extracted DNA from the elephant specimen used in this study. R. Debruyne provided photographs of mammoths NMC44572 and 88.18.142, and D. Poinar provided the photograph of mammoth 2002/472, standing in for 2005/915. Funding was provided by an NSERC grant and Canada Research Chair to HP and generous contributions from MYcroarray. Sequence data are available at NCBI SRA BioProject Accession #SRP036634.

4.6 AUTHOR CONTRIBUTIONS

JMR and HP conceived of the study. JE, AD, MK, and YM performed wet laboratory work. JE analysed the data and wrote the manuscript with assistance from all authors.

4.7 SUPPLEMENTARY INFORMATION

4.7.1 Elephant DNA extraction

Four milliliters of blood from *Elephas maximus* (Asian elephant) was combined with 4mL of PBS buffer (Table S1), vortexed and centrifuged at 3500×g for 15m at room temperature. The supernatant was removed and the pellet was resuspended in 3mL of lysis buffer (Table S1) with 20µg of DNase-free RNase and 690pg of proteinase K. We

incubated this for 3h at 50°C, centrifuged to pellet undigested remnants, and then extracted the supernatant using an organic extraction procedure described elsewhere.(Poinar et al. 2006) The DNA was then precipitated from the resulting aqueous after adding 4mL of ETOH and 800µL of 5M sodium acetate, with mixing and subsequent centrifugation at 5000×g for 15 min at room temperature. The supernatant was decanted and the pellet was washed twice with 70% ETOH with centrifugation for 10m at 5000×g. The final supernatant was decanted and the pellet was dried and resuspended in 500µL TE and incubated overnight at 4°C. We quantified the DNA concentration with a GeneQuant Pro UV spectrophotometer (Amersham) to 11ng/ µL.

4.7.2 Bait synthesis

We sent roughly 1µg of elephant DNA extract to MYcroarray for global reverse transcription (both strands) with biotinylated rUTPs using their proprietary procedure. This provided an aqueous suspension of 71µg RNA baits.

4.7.3 Pleistocene fossil specimens

The four mammoths used in this evaluation were of Pleistocene age dated with AMS radiocarbon to less than 45 thousand ^{14}C years before present.

Lyuba: *Mammuthus primigenius* (woolly mammoth) subcutaneous abdominal muscle specimen from the Yamal Peninsula in Siberia, with AMS dates averaging 41,800 radiocarbon years before present (^{14}CyBP) (Fisher et al. 2012).

2005/915: *M. primigenius* bone specimen from the Taimyr Peninsula in Siberia, AMS dated to 27,740 ±220 ^{14}CyBP (Debruyne et al. 2008).

NMC-44572: *Mammuthus primigenius* tusk specimen from the Dominion Creek locality, British Columbia, AMS dated to 31,200 ±1200 ^{14}CyBP (Guthrie 2004).

USUEPM 88.18.142: *Mammuthus columbi* (Columbian mammoth) tusk specimen from Huntington Reservoir, Fairview, Utah, AMS dated to 11,220 ±110 ^{14}CyBP (Gillette, Madsen 1993).

4.7.4 Mammoth and mylodon DNA extraction

All ancient specimen manipulation took place in the dedicated ancient DNA facilities of the McMaster Ancient DNA Centre of McMaster University. Following subsampling, between 30 and 90mg of material was demineralised overnight at room temperature with 1mL 0.5M and shaking. Afterwards, samples were treated in the following ways:

Specimens Lyuba and mylodon

The EDTA-treated pellet was digested with 750µL of a Sarcosyl-based proteinase K solution according to recipe A in Table S2 at room temperature for 21h, frozen overnight, and then thawed and pH-adjusted with 1M Tris-HCl pH 9.0 solution and incubated for an additional 21h at room temperature. We purified 370µL of the digest using the QIAGEN MinElute PCR Purification kit with two washes of 700µL Buffer PE and eluted in 40µL (2 x 20µL) buffer EB at 0.05% Tween-20.

Specimens 2005/915 and NMC44572

Pellets were digested with 900µL of digestion solution (A in Table S2) for 19h at 50°C with shaking. This was removed and the pellet was demineralized again overnight with 1mL 0.5M EDTA at room temperature with

shaking. We then purified this second demineralization using an organic extraction procedure described elsewhere (Poinar et al. 2006) and concentrated the resulting aqueous to 60µL with Amicon Ultra 0.5mL 30kDa centrifugation filters (Millipore).

Specimen 88.18.142

The pellet was demineralized again with 1mL 0.5M EDTA overnight with shaking at room temperature. Then we added 980µL of digestion solution (B in Table S2) to the pellets and incubated with rotation at 50°C for 6h. This digestion was purified using the organic extraction protocol and concentrated with 30kDa filters to 60µL.

4.7.5 Library preparation

We used 25µL of each DNA extract in duplicate Illumina® library preparations with uracil DNA glycosylase treatment as described elsewhere (Kircher, Sawyer, Meyer 2012) replacing all SPRI bead cleanups with MinElute purification to 20µL Buffer EB. We did not heat-deactivate the B_{st} polymerase following the fill-in step and instead purified the reaction with MinElute to 50µL (2 x 15µL + 30µL) Buffer EB with 1mM EDTA ('TEB').

We then index-amplified each library using unique P5 and P7 indexing primers (Meyer, Kircher 2010) in 40µL reactions using 10µL of each library according to the recipe and thermal cycling conditions in Table S3. Amplifications were performed in real-time with a CFX96 Real-time PCR platform (BioRad). Indexed libraries were purified with MinElute to 15µL TEB.

4.7.6 Enrichment and re-amplification

We used the MYbaits (MYcroarray) kit protocol for enrichment except replaced buffer Hyb #4 with 10% SDS (at one tenth the volume) and Block #3 with blocking oligonucleotides matching one strand of all four universal adapter sections (BO2 split to match either side of the index, BO4, and BO6) (Meyer, Kircher 2010) at 2µM each per capture. Each capture reaction used 1µg of elephant baits and 9µL indexed library, which ranged from 0.5 to 5.3ng/µL as estimated with total library quantification by technicians at the sequencing facility. Hybridizations were done at 48°C for 37.5h. Following bead cleanup and MinElute purification to 15µL TEB, enriched eluates were amplified for 10 cycles according to the conditions in Table S4 and then purified with MinElute to 13µL TEB. Then 9µL of these re-amplified enriched eluates were used in another round of capture using identical conditions as the first round except incubated at 55°C for 39h. These were cleaned and then purified with MinElute to 13µL TEB, which we then re-amplified for 5 cycles using identical conditions as the first re-amplification (Table S4) except with 75nM each primer. These final re-amplified doubly-enriched libraries were then purified to 13µL TEB.

4.7.7 Sequencing

Non-enriched and enriched libraries were combined in two separate pools and sequenced on two separate lanes (with dozens of other libraries) of an Illumina HiSeq® 1500 flowcell at the Farncombe Metagenomics Facility at McMaster University. Both cBot cluster generation and sequencing employed the v3 chemistry and a 2x80bp dual 7bp indexing protocol, using the alternative primer mixes from the TruSeq Dual Index Sequencing Primer Kit (Paired End). We included a dedicated control lane with the PhiX control kit v3 and a 1% PhiX spike in each lane. Raw data was processed with HCS version 1.5.15.1 and RTA version 1.13.48.0. File conversion and demultiplexing using each 7bp reverse index (requiring a 100% match) was performed with CASAVA version 1.8.2.

4.7.8 Read processing and alignment

Reads 1 and 2 were trimmed and merged with SeqPrep (https://github.com/jstjohn/SeqPrep), searching for at least 5bp (-O 5) of the universal 13bp adapter linker (-A AGATCGGAAGAGC), and merging pairs when overlapping by at least 11bp (-o 11), with a minimum 84% identity in the overlapping region (-N 0.84). Trimmed and/or merged reads less than 24bp were simultaneously discarded (-L 24). The merged reads and non-mergable R1s were subsequently combined into a single read file and then with Burrows-Wheeler Aligner (BWA) version 0.6.1-r104 (Li, Durbin 2009) aligned to repeats-masked non-masked versions of the *Loxodonta africana* genome (NCBI RefSeq Assembly ID GCF_000001905.1), a mammoth mitochondrial genome (NCBI Accession NC_015529.1), the *Choloepus hoffmanni* nuclear supercontigs (NCBI BioProject PRJNA30809) and a mylodon mitogenome sequence previously constructed in our laboratory. Raw read alignment rates are available in Table S5. For *Loxodonta* repeat coverage analysis, resulting SAM files were converted to BED files using BEDOPs version 2.2.0 (Neph et al. 2012) and then alignment locations were extracted using the repeats annotations associated with the *Loxodonta* genome reference. For insert length distribution analysis, the merged reads alone were aligned to the unmasked elephant genome and the mapped and unmapped reads were analyzed separately (Figure S1).

4.7.9 Complexity measurement and projection

To generate complexity curves we randomly sampled the BED files in 1% increments and then collapsed them to reads with unique 5' and 3' coordinates and strand (direction). These complexity curves are the 'observed' datapoints presented in Figure S2. We also depict this information as % unique in Figure S3. From the same original BED files we also counted the numbers of copies of each coordinate-identical group of sequences and then used preseq version 0.0.3 (Daley, Smith 2013) to predict the theoretical unique read yield obtainable with deeper sequencing of the libraries, depicted as 'projected' datapoints in Figure S2.

4.7.10 Supplementary Tables

Table S1: PBS buffer and lysis buffer recipes used in elephant blood DNA extraction

PBS Buffer		**Lysis Buffer**	
Reagent	*Concentration*	*Reagent*	*Concentration*
NaCl	137mM	Tris-Cl (pH 8.0)	10mM
KCl	2.7mM	EDTA (pH 8.0)	0.1mM
Na_2HPO_4	10mM	SDS	0.5%
KH_2PO_4	2mM		

Table S2: Recipes for digestion solutions used for ancient DNA extraction

Digestion Solution A		**Digestion Solution B**	
Reagent	*Concentration*	*Reagent*	*Concentration*
Tris-Cl (pH 9.0)	0.01M	Tris-Cl (pH 9.0)	0.01M
Sarcosyl	0.5%	Sarcosyl	0.5%
Proteinase K	0.25mg/mL	Proteinase K	0.25mg/mL
$CaCl_2$	5mM	$CaCl_2$	5mM
DTT	50mM	DTT	50mM
PVP	1%	PVP	1%
PTB	2.5mM	PTB	0mM

Table S3: Indexing amplification conditions

Reaction composition		**Thermal Cycling**		
Reagent	*Concentration*	*Step*	*Temperature*	*Time*
Herculase II Fusion Buffer (Agilent)	1X	1	98°C	3 m
P5 indexing primer	100nM	2	98°C	20 s
P7 indexing primer	100nM	3	60°C	20 s
dNTPs	0.25mM	4	72°C	35 s
EvaGreen (Biotium)	0.5X	5	To Step 2 ×9	
Herculase II Fusion Polymerase (Agilent)	0.025U/µL	6	72°C	3 m

Table S4: Re-amplification conditions following the first round

Reaction composition		**Thermal Cycling**		
Reagent	*Concentration*	*Step*	*Temperature*	*Time*
Herculase II Fusion Buffer (Agilent)	1X	1	98°C	3 m
IS5_reamp.P5(Meyer, Kircher 2010)	100nM	2	98°C	20 s
IS6_reamp.P7(Meyer, Kircher 2010)	100nM	3	60°C	20 s
dNTPs	0.25mM	4	72°C	35 s
EvaGreen (Biotium)	0.5X	5	To Step 2 ×9	
Herculase II Fusion Polymerase (Agilent)	0.025U/µL	6	72°C	3 m

Table S5: Raw read alignment rates of both replicates (Rep1 and Rep2) to various references both before (tan) and after (brown) enrichment.

COUNTS:	R1s		*L. africana* Genome		*L.a.* Repeats Masked		*M. columbi* mt		*Choloepus hoffmanni*		*M. darwinii* mt	
Specimen	Rep1	Rep2	Rep1	Rep2	Rep1	Rep2	Rep1	Rep2	Rep1	Rep2	Rep1	Rep2
Lyuba	58,931	51,745	16,313	14,693	9,798	8,868	4,122	3,717	482	426	72	53
2005/915	223,858	225,127	165,070	166,501	78,955	79,641	277	251	2,566	2,591	4	6
NMC-44572	57,862	42,522	2,702	1,789	1,553	1,023	29	21	178	105	2	2
88.18.142	112,765	278,853	8,771	22,666	4,761	12,312	14	23	416	1,044	1	1
Mylodon	277,203	298,363	5,632	6,099	2,795	3,028	1	1	113,788	121,683	63	61
% of R1s												
Lyuba			27.68%	28.40%	16.63%	17.14%	6.995%	7.183%	0.82%	0.82%	0.1222%	0.1024%
2005/915			73.74%	73.96%	35.27%	35.38%	0.124%	0.111%	1.15%	1.15%	0.0018%	0.0027%
NMC-44572			4.67%	4.21%	2.68%	2.41%	0.050%	0.049%	0.31%	0.25%	0.0035%	0.0047%
88.18.142			7.78%	8.13%	4.22%	4.42%	0.012%	0.008%	0.37%	0.37%	0.0009%	0.0004%
Mylodon			2.03%	2.04%	1.01%	1.01%	0.000%	0.000%	41.05%	40.78%	0.0227%	0.0204%
COUNTS:	**R1s**		***L. africana* Genome**		***L.a.* Repeats Masked**		***M. columbi* mt**		***Choloepus hoffmanni***		***M. darwinii* mt**	
Specimen	Rep1	Rep2	Rep1	Rep2	Rep1	Rep2	Rep1	Rep2	Rep1	Rep2	Rep1	Rep2
Lyuba	1,209,265	894,586	988,507	717,349	382,066	274,478	73,633	64,187	3,711	3,332	333	264
2005/915	1,704,997	1,829,002	1,334,751	1,430,828	405,204	433,893	417	485	3,050	3,675	9	2
NMC-44572	1,127,311	985,281	928,345	820,221	413,448	356,707	1,201	1,078	7,800	7,446	8	11
88.18.142	1,388,923	1,532,251	1,112,235	1,212,129	405,373	426,355	224	241	8,859	9,714	54	3
Mylodon	1,001,919	946,097	42,527	42,388	3,857	3,851	10	13	508,502	485,441	194	187
% of R1s												
Lyuba			81.74%	80.19%	31.59%	30.68%	6.089%	7.175%	0.31%	0.37%	0.0275%	0.0295%
2005/915			78.28%	78.23%	23.77%	23.72%	0.024%	0.027%	0.18%	0.20%	0.0005%	0.0001%
NMC-44572			82.35%	83.25%	36.68%	36.20%	0.107%	0.109%	0.69%	0.76%	0.0007%	0.0011%
88.18.142			80.08%	79.11%	29.19%	27.83%	0.016%	0.016%	0.64%	0.63%	0.0039%	0.0002%
Mylodon			4.24%	4.48%	0.38%	0.41%	0.001%	0.001%	50.75%	51.31%	0.0194%	0.0198%

4.7.11 Supplementary Figures

Figure S1: Insert length distributions for unmapped and mapped merged reads. Tan distributions are from non-enriched libraries, brown are from enriched libraries. Data from first replicates only. Average insert lengths of non-enriched libraries depicted in dashed tan lines, enriched libraries dashed brown lines.

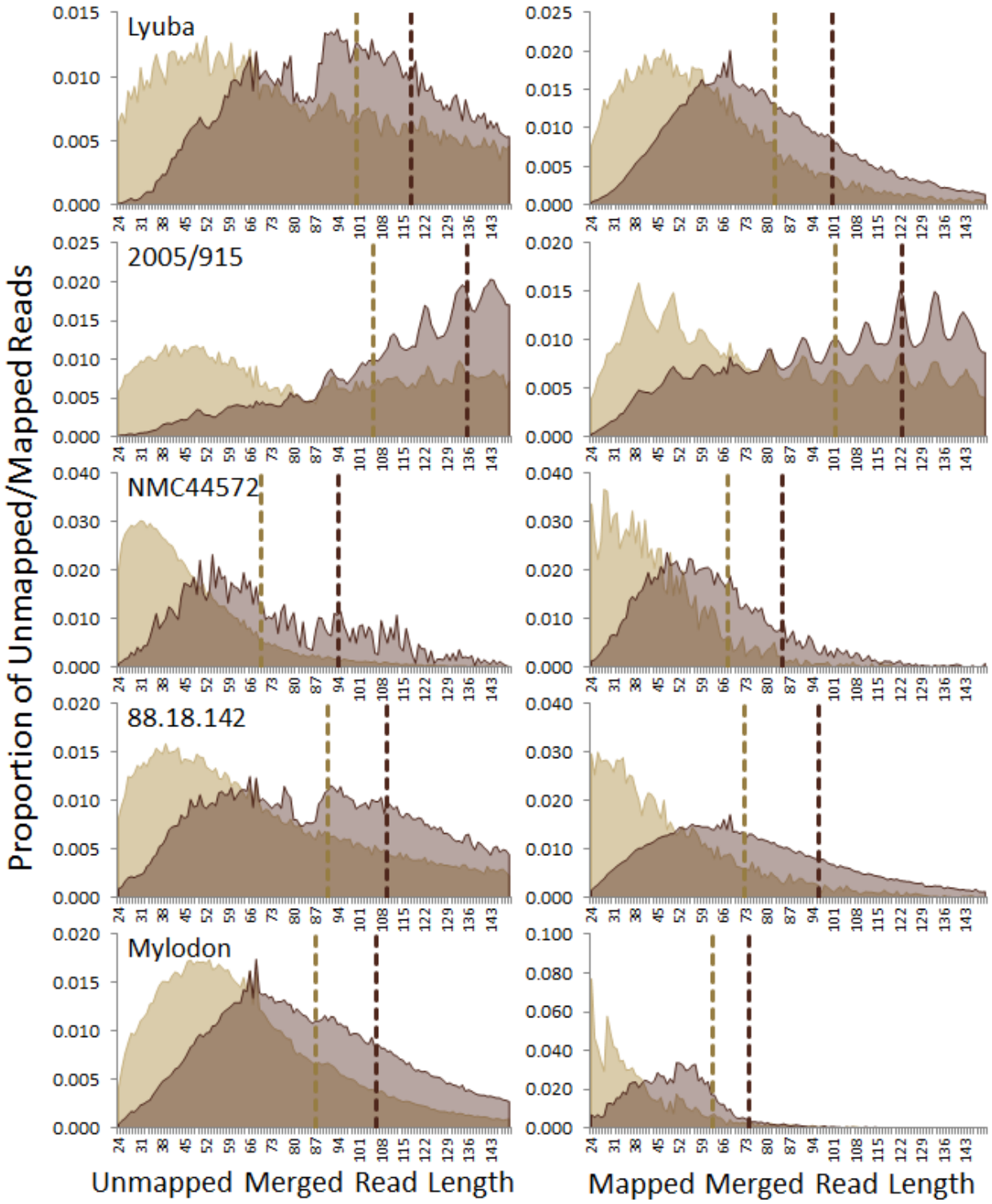


Figure S2. Enriched library complexity curves and projected unique read yield with deeper sequencing. Both replicates depicted as dark and light lines of the same hue. Observed data are depicted with solid lines, while projections (not including 95% HPDs) from that observed data are depicted in dashed lines.

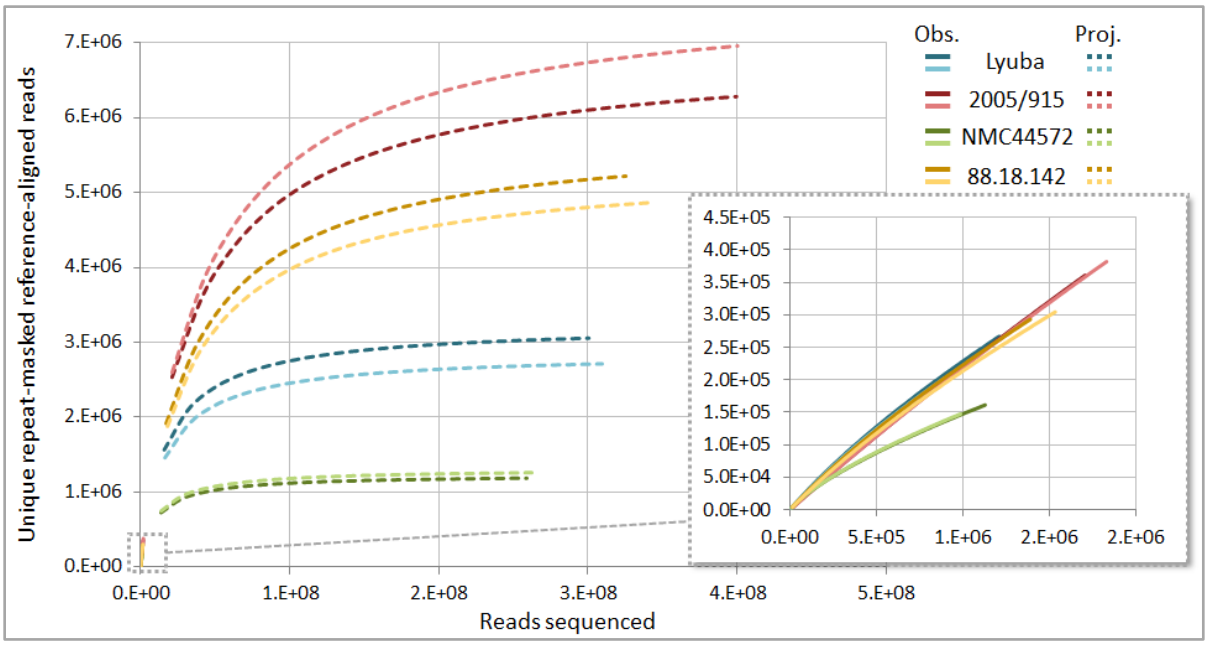
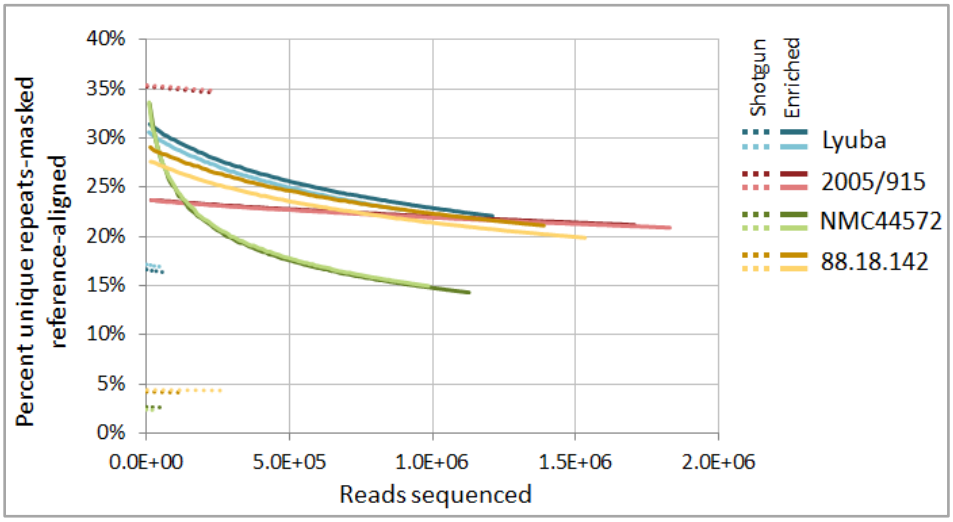


Figure S3. Percent unique alignable reads with increasing sequencing depth. Both replicates depicted as dark and light lines of the same hue. Shotgun data shown in dotted lines, enriched data in solid lines.



4.8 REFERENCES

Aird, D, MG Ross, WS Chen, M Danielsson, T Fennell, C Russ, DB Jaffe, C Nusbaum, A Gnirke. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol 12:R18.

Bos, KI, VJ Schuenemann, GB Golding, et al. 2011. A draft genome of Yersinia pestis from victims of the Black Death. Nature 478:506-510.

Briggs, AW, JM Good, RE Green, et al. 2009. Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. Science 325:318-321.

Britten, RJ, DE Kohne. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. Science 161:529-540.

Burbano, HA, E Hodges, RE Green, et al. 2010. Targeted investigation of the Neandertal genome by array-based sequence capture. Science 328:723-725.

Carpenter, ML, JD Buenrostro, C Valdiosera, et al. 2013. Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. Am J Hum Genet 93:852-864.

Daley, T, AD Smith. 2013. Predicting the molecular complexity of sequencing libraries. Nat Methods 10:325-327.

Debruyne, R, G Chu, CE King, et al. 2008. Out of America: Ancient DNA evidence for a new world origin of late quaternary woolly mammoths. Current Biology 18:1320-1326.

Fisher, DC, AN Tikhonov, PA Kosintsev, AN Rountrey, B Buigues, J van der Plicht. 2012. Anatomy, death, and preservation of a woolly mammoth (Mammuthus primigenius) calf, Yamal Peninsula, northwest Siberia. Quaternary International 255:94-105.

Gillette, DD, DB Madsen. 1993. The Columbian Mammoth, Mammuthus-Columbi, from the Wasatch Mountains of Central Utah. Journal of Paleontology 67:669-680.

Gnirke, A, A Melnikov, J Maguire, et al. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 27:182-189.

Guthrie, RD. 2004. Radiocarbon evidence of mid-Holocene mammoths stranded on an Alaskan Bering Sea island. Nature 429:746-749.

Hallstrom, BM, M Kullberg, MA Nilsson, A Janke. 2007. Phylogenomic data analyses provide evidence that Xenarthra and Afrotheria are sister groups. Molecular Biology and Evolution 24:2059-2068.

Hodges, E, M Rooks, Z Xuan, A Bhattacharjee, D Benjamin Gordon, L Brizuela, W Richard McCombie, GJ Hannon. 2009. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nat Protoc 4:960-974.

Kircher, M, S Sawyer, M Meyer. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res 40:e3.

Kozarewa, I, DJ Turner. 2011. Amplification-free library preparation for paired-end Illumina sequencing. Methods Mol Biol 733:257-266.

Li, H, R Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.

Meyer, M, M Kircher. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc 2010:pdb prot5448.

Mindell, DP, RL Honeycutt. 1990. Ribosomal-Rna in Vertebrates - Evolution and Phylogenetic Applications. Annual Review of Ecology and Systematics 21:541-566.

Neph, S, MS Kuehn, AP Reynolds, et al. 2012. BEDOPS: high-performance genomic feature operations. Bioinformatics 28:1919-1920.

Poinar, HN, C Schwarz, J Qi, et al. 2006. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science 311:392-394.

Rohland, N, AS Malaspinas, JL Pollack, M Slatkin, P Matheus, M Hofreiter. 2007. Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup. Plos Biology 5:1663-1671.

Rohland, N, D Reich, S Mallick, M Meyer, RE Green, NJ Georgiadis, AL Roca, M Hofreiter. 2010. Genomic DNA Sequences from Mastodon and Woolly Mammoth Reveal Deep Speciation of Forest and Savanna Elephants. Plos Biology 8.

Sawyer, S, J Krause, K Guschanski, V Savolainen, S Paabo. 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS One 7:e34131.

PART II

CONCLUSION

Prior to this thesis, nothing was known, or at least published, about the genetic identity or phylogeographic structure of Columbian, Jeffersonian, or pygmy mammoths. Here I and my coauthors filled this gap by exploiting recent technological and methodological advances to break the 'permafrost barrier' in ancient phylogeography research. First, in Chapter 1, we used high-throughput sequencing to demonstrate that two Columbian mammoths found in temperate, non-speleological burial contexts were at the mitochondrial level closely related to woolly mammoths in North America – more closely related than woolly mammoths on either side of the Bering Land Bridge were to each other. Then in Chapter 3, we demonstrate that this was apparently the rule rather than the exception for Columbian and other temperate-adapted North American mammoths. This is necessarily at odds with the conventional paleontological model of woolly and North American non-woolly mammoths having evolved separately for a million years. It instead suggests that mammoth evolution in North America, much as has been suspected in Eurasia [55], was characterized by multiple moments of interbreeding between nominal taxa. We further uncovered that despite this likely fluidity of gene interchange, mitogenomic phylogeographic structure still existed at a finer level in North American mammoths, with morphological correspondence in turn. In general, the ecozone in which a mammoth died and presumably lived (e.g., the periglacial steppe along the Laurentide ice sheet, or the savanna-parklands of the central and western regions of the continent), as well as their apparent morphological taxon, predicts their mitogenomic affinity. Where there are exceptions to this phylogeographic structure, individuals often exhibit morphological intermediacy, which we posit is likely the result of local genetic interchange at ecotones.

In order to discern the origin of this apparent 'convergence' of morphologies between distantly-related mammoth matrilines, but divergence of morphologies between closely-related matrilines, ancient DNA science must move towards population-scale nuclear genomic assessment. The matriarchal herd structure of extant and presumably extinct elephants can obscure the relationship between mitochondrial phylogenies and species trees. But in order to acquire nuclear DNA sequences from the very poorly-preserved remains like those studied mitogenomically here, more work in refining targeted enrichment is required. Our initial attempts at capturing nuclear targets from Columbian mammoths were relatively unsuccessful, with just a few of roughly three dozen attempted specimens yielding multi-fold coverage of just small portion of the targeted nuclear genome regions. Furthermore, despite efforts to choose targets that might vary intraspecifically, the majority of the regions that we successfully sequenced either did not differ from the modern elephant genome, or did not vary between the individuals sequenced where that could be assessed.

In theory, successful capture and sequencing of nuclear DNA from very poorly-preserved remains is merely a matter of the scale of raw material input. As we demonstrate in Chapter 2, and refine and repeat on a larger scale to acquire the data in Chapter 3, even mammoth DNA extracts with extremely low frequency and highly fragmented endogenous DNA could be enriched for entire mitochondrial genomes. As such, for routine nuclear capture, the field could benefit from techniques that parallelize DNA extraction from large subsamples (many grams of material), to provide enough absolute target template for capture and multi-fold consensus resolution. But in order to identify and then capture regions that will be informative for intraspecific phylogenetics across large sample sets, it will be necessary to first sequence complete genomes of several individuals. In Chapter 4 we test one promising method for achieving this, demonstrating that a relatively inexpensive and versatile approach to bait manufacture can greatly increase the frequency of highly fragmented nuclear DNA in complex ancient DNA extracts. I would not be surprised if after a few methodological refinements of this approach, a technique like this becomes a standard tool in paleogenetics, until sequencing is so inexpensive that the DNA content of very complex extracts can be routinely sequenced to saturation.

Many of the biological research questions related to the taxonomic and thus chronologic context of mammoth matriline phylogeny grew from my Master's degree research [56]. My original intention was to use the same techniques that had been applied to permafrost-preserved remains to sequence short regions of the mitochondrial genome from temperate-preserved remains; one that would have likely resulted in a very low success rate. But coincident with my arrival at McMaster, ancient DNA science was in the midst of a tremendous technological shift, which necessarily changed the breadth of research questions that could be approached and answered. In many ways this thesis tracks this shift, and thus is as much (if not more) a molecular biological endeavor as it is an evolutionary biological one. As such it sets the stage for continuing research in both areas. Many questions remain

regarding the sensitivity and behavior of the enrichment techniques explored here, and research should move towards a refined understanding of their nuances in order to reduce biases and increase predictability. For mammoths, a large portion of both their temporal and geographic range remain unsampled, and so fully contextualizing the genetic patterns we uncovered requires expanding research to even lower latitudes and more ancient samples. This latter item is particularly salient, as even sophisticated relaxed molecular clock models are likely erroneously estimating the chronology of phylogeographic cladogenesis in mammoths and other animals, which confounds placing it in the broader scheme of Pleistocene climatic and biogeographic history. A better understanding of targeted enrichment and acquiring nuclear DNA from even more poorly-preserved remains than the ones studied here will require work. But I am confident that with a pace of technological change seen in the period in which this thesis was written, these and other goals will be achieved in the near future.

PART III

REFERENCES

1. Grayson DK, Meltzer DJ: **A requiem for North American overkill**. *J Archaeol Sci* 2003, **30**(5):585-593.
2. Martin PS, Klein RG: **Quaternary extinctions: a prehistoric revolution**: University of Arizona Press; 1989.
3. Fisher DC, Tikhonov AN, Kosintsev PA, Rountrey AN, Buigues B, van der Plicht J: **Anatomy, death, and preservation of a woolly mammoth (Mammuthus primigenius) calf, Yamal Peninsula, northwest Siberia**. *Quatern Int* 2012, **255**:94-105.
4. Ludin R: **" Baby Mammoth Dima"; a new discovery**. *J Paleontol* 1978, **52**(4):941-942.
5. Mol D, Coppens Y, Tikhonov A, Agenbroad L, MacPhee R, Flemming C, Greenwood A, Buigues B, De Marliave C, Van Geel B: **The Jarkov mammoth: 20,000-year-old carcass of a Siberian woolly mammoth Mammuthus primigenius (Blumenbach, 1799)**. In: *Proceedings of the 1st International Congress' The World of Elephants '(Roma): 2001*. 305-309.
6. Campos-Arceiz A, Blake S: **Megagardeners of the forest–the role of elephants in seed dispersal**. *Acta Oecologica* 2011, **37**(6):542-553.
7. Fritz H, Duncan P, Gordon IJ, Illius AW: **Megaherbivores influence trophic guilds structure in African ungulate communities**. *Oecologia* 2002, **131**(4):620-625.
8. Owen-Smith RN: **Megaherbivores: the influence of very large body size on ecology**: Cambridge University Press; 1992.
9. Owen-Smith N: **Pleistocene extinctions: the pivotal role of megaherbivores**. *Paleobiology* 1987:351-362.
10. Haynes G: **Mammoths, mastodonts, and elephants: biology, behavior and the fossil record**: Cambridge University Press; 1993.
11. Johnson CN: **Ecological consequences of Late Quaternary extinctions of megafauna**. *Proceedings of the Royal Society B: Biological Sciences* 2009, **276**(1667):2509-2519.
12. Guthrie RD: **Mammals of the mammoth steppe as paleoenvironmental indicators**: Academic Press; 1982.
13. Johnson PH, Olson CB, Goodman M: **Isolation and characterization of deoxyribonucleic acid from tissue of the woolly mammoth, *Mammuthus primigenius***. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* 1985, **81**(4):1045-1051.
14. Hagelberg E, Thomas MG, Cook CE, Sher AV, Baryshnikov GF, Lister AM: **DNA from Ancient Mammoth Bones**. *Nature* 1994, **370**(6488):333-334.
15. Höss M, Paabo S, Vereshchagin NK: **Mammoth DNA-Sequences**. *Nature* 1994, **370**(6488):333-333.
16. Greenwood AD, Capelli C, Possnert G, Pääbo S: **Nuclear DNA sequences from late Pleistocene megafauna**. *Mol Biol Evol* 1999, **16**(11):1466-1473.
17. Greenwood AD, Lee F, Capelli C, DeSalle R, Tikhonov A, Marx PA, MacPhee RD: **Evolution of endogenous retrovirus-like elements of the woolly mammoth (Mammuthus primigenius) and its relatives**. *Mol Biol Evol* 2001, **18**(5):840-847.
18. Campbell KL, Roberts JE, Watson LN, Stetefeld J, Sloan AM, Signore AV, Howatt JW, Tame JR, Rohland N, Shen TJ *et al*: **Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance**. *Nature genetics* 2010, **42**(6):536-540.
19. Yang H, Golenberg EM, Shoshani J: **Phylogenetic resolution within the Elephantidae using fossil DNA sequence from the American mastodon (Mammut americanum) as an outgroup**. *Proceedings of the National academy of Sciences* 1996, **93**(3):1190-1194.
20. Thomas MG, Hagelberg E, Jones HB, Yang Z, Lister AM: **Molecular and morphological evidence on the phylogeny of the Elephantidae**. *Proceedings of the Royal Society of London Series B: Biological Sciences* 2000, **267**(1461):2493-2500.
21. Hauf J, Joger U, Zimmermann F, Lazarev P, Vartanyan S: **Protein and Nucleic acid sequences of Woolly Mammoth cytochrome b and the phylogenetic position of Mammuthus within the Elephantidae**. In: *Mammoths and the Mammoths Fauna: Proceedings of the First International Mammoth Conference, St Petersburg, Russia Deinsea (suppl): 1999*. 211-217.
22. Rohland N, Malaspinas AS, Pollack JL, Slatkin M, Matheus P, Hofreiter M: **Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup**. *Plos Biol* 2007, **5**(8):1663-1671.
23. Barnes I, Shapiro B, Lister A, Kuznetsova T, Sher A, Guthrie D, Thomas MG: **Genetic structure and extinction of the woolly mammoth, Mammuthus primigenius**. *Current biology : CB* 2007, **17**(12):1072-1075.

24. Debruyne R, Chu G, King CE, Bos K, Kuch M, Schwarz C, Szpak P, Grocke DR, Matheus P, Zazula G *et al*: **Out of America: Ancient DNA evidence for a new world origin of late quaternary woolly mammoths**. *Current Biology* 2008, **18**(17):1320-1326.
25. Gilbert MTP, Drautz DI, Lesk AM, Ho SYW, Qi J, Ratan A, Hsu CH, Sher A, Dalen L, Gotherstrom A *et al*: **Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes**. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(24):8327-8332.
26. Palkopoulou E, Dalen L, Lister AM, Vartanyan S, Sablin M, Sher A, Edmark VN, Brandstrom MD, Germonpre M, Barnes I *et al*: **Holarctic genetic structure and range dynamics in the woolly mammoth**. *Proceedings Biological sciences / The Royal Society* 2013, **280**(1770):20131910.
27. Campos PF, Willerslev E, Sher A, Orlando L, Axelsson E, Tikhonov A, Aaris-Sørensen K, Greenwood AD, Kahlke R-D, Kosintsev P: **Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (Ovibos moschatus) population dynamics**. *Proceedings of the National academy of Sciences* 2010, **107**(12):5675-5680.
28. Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MT, Barnes I, Binladen J *et al*: **Rise and fall of the Beringian steppe bison**. *Science* 2004, **306**(5701):1561-1565.
29. Campos PF, Kristensen T, Orlando L, Sher A, Kholodova MV, Götherström A, Hofreiter M, Drucker DG, Kosintsev P, Tikhonov A: **Ancient DNA sequences point to a large loss of mitochondrial genetic diversity in the saiga antelope (Saiga tatarica) since the Pleistocene**. *Mol Ecol* 2010, **19**(22):4863-4875.
30. Prost S, Smirnov N, Fedorov VB, Sommer RS, Stiller M, Nagel D, Knapp M, Hofreiter M: **Influence of climate warming on arctic mammals? New insights from ancient DNA studies of the collared lemming Dicrostonyx torquatus**. *PloS one* 2010, **5**(5):e10447.
31. Barnett R, Shapiro B, Barnes I, Ho SY, Burger J, Yamaguchi N, Higham TF, WHEELER H, Rosendahl W, Sher AV: **Phylogeography of lions (Panthera leo ssp.) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity**. *Mol Ecol* 2009, **18**(8):1668-1677.
32. Maglio VJ: **Origin and evolution of the Elephantidae**. *Transactions of the American Philosophical Society* 1973:1-149.
33. Osborn HF: **Proboscidea. A Monograph of the Discovery, Evolution, Migration and Extinction of the Mastodonts and Elephants of the World: Vol. I-[II]**: American Museum of Natural History; 1942.
34. Agenbroad LD: **North American Proboscideans: Mammoths: The state of Knowledge, 2003**. *Quatern Int* 2005, **126**:73-92.
35. Kurtén Br, Anderson E: **Pleistocene mammals of North America**. New York: Columbia University Press; 1980.
36. Archie EA, Maldonado JE, HOLLISTER-SMITH JA, Poole JH, Moss CJ, Fleischer RC, Alberts SC: **Fine-scale population genetic structure in a fission–fusion society**. *Mol Ecol* 2008, **17**(11):2666-2679.
37. Roca AL, Georgiadis N, O'Brien SJ: **Cytonuclear genomic dissociation in African elephant species**. *Nature genetics* 2005, **37**(1):96-100.
38. Vidya TNC, Sukumar R, Melnick DJ: **Range-wide mtDNA phylogeography yields insights into the origins of Asian elephants**. *P R Soc B* 2009, **276**(1658):893-902.
39. Fisher DC: **Season of death, growth rates, and life history of the North American mammoths**. In: *Mammoth Site Studies Proceedings of the First International Conference on Mammoth Site Studies, Lawrence, KS, March 11-13, 1998: 2001*. 121-135.
40. Lister AM, Sher AV, van Essen H, Wei GB: **The pattern and process of mammoth evolution in Eurasia**. *Quatern Int* 2005, **126**:49-64.
41. Reich D, Patterson N, Kircher M, Delfin F, Nandineni MR, Pugach I, Ko AM-S, Ko Y-C, Jinam TA, Phipps ME: **Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania**. *The American Journal of Human Genetics* 2011, **89**(4):516-528.
42. Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D: **The genomic landscape of Neanderthal ancestry in present-day humans**. *Nature* 2014.
43. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai WW, Fritz MHY *et al*: **A Draft Sequence of the Neandertal Genome**. *Science* 2010, **328**(5979):710-722.

44. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prufer K, de Filippo C *et al*: **A high-coverage genome sequence from an archaic Denisovan individual**. *Science* 2012, **338**(6104):222-226.
45. Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, Garcia N, Paabo S, Arsuaga JL *et al*: **Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments**. *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**(39):15758-15763.
46. Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid KA, Willerslev E, Krogh A, Orlando L: **Improving ancient DNA read mapping against modern reference genomes**. *BMC genomics* 2012, **13**:178.
47. Orlando L, Ginolhac A, Zhang GJ, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I *et al*: **Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse**. *Nature* 2013, **499**(7456):74-+.
48. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ: **Direct selection of human genomic loci by microarray hybridization**. *Nature methods* 2007, **4**(11):903-905.
49. Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, McPhee JB, DeWitte SN, Meyer M, Schmedes S *et al*: **A draft genome of Yersinia pestis from victims of the Black Death**. *Nature* 2011, **478**(7370):506-510.
50. Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajkovic D, Kucan Z *et al*: **Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes**. *Science* 2009, **325**(5938):318-321.
51. Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, Meyer M, Good JM, Maricic T, Johnson PL, Xuan Z *et al*: **Targeted investigation of the Neandertal genome by array-based sequence capture**. *Science* 2010, **328**(5979):723-725.
52. Devault AM, Golding GB, Waglechner N, Enk JM, Kuch M, Tien JH, Shi M, Fisman DN, Dhody AN, Forrest S: **Second-Pandemic Strain of Vibrio cholerae from the Philadelphia Cholera Outbreak of 1849**. *New England Journal of Medicine* 2014.
53. Horn S, Prost S, Stiller M, Makowiecki D, Kuznetsova T, Benecke N, Pucher E, Hufthammer AK, Schouwenburg C, Shapiro B: **Ancient mitochondrial DNA and the genetic history of Eurasian beaver (Castor fiber) in Europe**. *Mol Ecol* 2014.
54. Sawyer S, Krause J, Guschanski K, Savolainen V, Paabo S: **Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA**. *PloS one* 2012, **7**(3):e34131.
55. Lister AM, Sher AV: **The origin and evolution of the woolly mammoth**. *Science* 2001, **294**(5544):1094-1097.
56. Enk JM, Yesner DR, Crossen KJ, Veltre DW, O'Rourke DH: **Phylogeographic analysis of the mid-Holocene Mammoth from Qagnax Cave, St. Paul Island, Alaska**. *Palaeogeogr Palaeocl* 2009, **273**(1-2):184-190.