Complete mitochondrial genome and phylogeny of Pleistocene mammoth Mammuthus primigenius

Evgeny I. Rogaev
University of Massachusetts Medical School

Let us know how access to this document benefits you.
Follow this and additional works at: https://escholarship.umassmed.edu/oapubs

Part of the Life Sciences Commons, and the Medicine and Health Sciences Commons

Repository Citation

This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in Open Access Publications by UMass Chan Authors by an authorized administrator of eScholarship@UMassChan. For more information, please contact Lisa.Palmer@umassmed.edu.
Introduction

*Mammuthus, Elephas,* and *Loxodonta* (family Elephantidae, subfamily Elephantinae) are closely related genera that evolved in the African Pliocene, possibly from the genus *Primelephas* [1–4]. The woolly mammoth *Mammuthus primigenius* became extinct across most of its former range along with other late Pleistocene megafauna, although small, isolated mammoth populations survived into the mid-Holocene [5]. The phylogeny of Elephantinae has not been resolved. Morphological analyses have yielded conflicting phylogenies for *M. primigenius, E. maximus,* and *L. africana* [1–4]. Dental characters suggest a closer relationship between *M. primigenius* and *E. maximus,* trunk tip morphology supports a grouping of *M. primigenius* and *L. africana,* and immunological and hair structure characters could not confidently resolve the phylogeny of these three taxa [1–4]. Molecular analyses have also generated conflicting conclusions [1–4]. Data on short mitochondrial DNA (mtDNA) sequences have variously supported a monophyletic clade of the extant elephant species, have grouped *M. primigenius* with either *E. maximus* [6] or *L. africana* [7–9], or have been inconclusive [10,11]. DNA template modifications caused by oxidation or hydrolysis are a potential source of artificial mutations and may partly explain the high polymorphism initially observed for some short mammoth DNA sequences, although pseudogenes of mitochondrial origin located in the nuclear genome, which are common in elephants, may also be a potential factor [10]. A meta-analysis taking into account potential sequencing and other errors favored the *M. primigenius–L. africana* clade [9]. Clearly, a comprehensive analysis of longer genomic sequences is necessary to resolve this phylogeny, but so far only short DNA fragments have been retrieved from *M. primigenius."

**Results/Discussion**

The woolly mammoth leg with intact muscle and skin tissue was found in the Enmynveem River valley (Chukotka) in north-eastern Siberia in 1986 (Figures 1 and S1). Since then, the collected muscle specimens of the mammoth (hereafter called Enmyn) were kept frozen. Radiocarbon method dated the specimens to ~33,750–31,950 y BP (before present) [12]. An initial examination suggested that the soft tissue was remarkably well preserved, and no signs of tissue decay were noticed when the specimen was excavated from the permafrost. Because no tissue damage by insects or other animals was observed, the remains presumably were quickly buried and have never been defrosted. The cells with nuclei were observed in epithelia and muscle tissue. The treatment with DAPI efficiently stained the nuclei, indicating the existence of

well-preserved genomic DNA (Figure 2A). To our knowledge, this is the first documented cytogenetically detectable nucleus with genomic DNA in such an ancient tissue (>30,000 y BP).

In accordance with these primary observations, a substantial amount of genomic DNA was extracted and detected on an electrophoresis gel. Although the extracted DNA was apparently degraded, it is yet of remarkable quality and quantity. For example, the major fraction of DNA fragments from one extract (shown in Figure 2B) ranged from 100 base pairs (bp) to 600 bp, with diminishing amounts at higher molecular weights. Microbial DNA may potentially contribute to the high-molecular-weight DNA fraction. Nevertheless, the PCR analysis described below (Figure 2C and 2D) supports the assertion that this mammoth specimen contains preserved DNA. Independent replication in PCR analysis is absolutely imperative for the study of ancient DNA. Therefore, several DNA extracts from the tissue specimens were obtained for the mitochondrial genome analysis.

Importantly, the replication in this study was not limited to selected mitochondrial genome regions, and the entire mitochondrial genome was sequenced in duplicate. Different DNA extracts from mammoth muscle tissue were used to generate PCR products and reconstruct the complete mitochondrial genome in Moscow (“MOS contig”, completed in 2000) and, independently, at the University of Massachusetts Medical School (UMASS MS laboratories (“UM contig”, completed in 2005) (see Materials and Methods). No contamination by extraneous DNA was found in multiple PCR experiments. Apart from a variable number of tandem repeats (VNTR) in the control region characterized by high somatic hypervariability, the mtDNA sequences obtained from different extracts and different laboratories matched exactly. This accuracy was achieved with the following protocol. (I) Redundant oligonucleotide primers were designed and tested to yield a sufficient amount of PCR products (Figures S2–S4 and Table S1). Preliminary tests with these primers were conducted to demonstrate the applicability of DNA extracts for PCR. We consistently found efficient recovery of at least ~500- to 700-bp PCR fragments from mammoth DNA extracts. Moreover, PCR of complete gene sequences of ~1,200–1,700 bp was also efficient, but the mtDNA sequence larger than 3,000 bp was not amplifiable (Figure 2C and 2D and Methods and Materials). The data are consistent with excellent preservation coupled with some degradation due to the ancient origin of the DNA. (II) The large amount of DNA enabled us to obtain PCR products sufficient for sequencing after the first round of PCR reaction and minimized the risk that sequence errors arose due to template switching during PCR, because of the low number of original DNA templates [13,14]. (III) Direct sequencing of PCR products gave clean chromatogram reads (Figure S5). In all cases, the replication of sequencing in each nucleotide position was performed from different PCR products using the protocols described in Materials and Methods. In addition, the PCR products were cloned and the independent clones were also sequenced. As expected, the cloned mtDNA fragments occasionally contained random mutations with an average rate of ~6/1,000 bp. No cloned sequences with identical sets of mutations were identified. These sequence modifications were mostly transitions corresponding to type I (A → G/T → C) and type II (C → T/G → A) mutations, which have previously been detected in ancient DNA (Figure S6) [13–15].
As described previously for ancient DNA, type II mutations were predominant in the cloned mammoth mtDNA sequences (>70%). The sequence modifications were not detectable in direct PCR product sequences of the same region (Figure S5). The data indicate that the mutations in cloned DNA were random and rare in each site of individual molecules. No contamination with nuclear mitochondrial DNA (numt) was found for any PCR product of mammoth DNA (Materials and Methods). Direct sequencing of multiple PCR products and independent clones provided evidence of accurate reconstruction of authentic mitochondrial genome sequence of *M. primigenius*.

The sequence of the small region (VNTR) in the control region (D-loop) could not be determined by direct sequencing of PCR products. Thus, the analysis of cloned PCR fragments was undertaken. The sequence analysis demonstrated somatic heteroplasmy and high molecular heterogeneity in this region. The heterogeneity resulted from hypervariability in a number of short hexa-nucleotide tandem repeats (CGCATA)_n resembling VNTR (Figure S4). Similar VNTR was found in the control mtDNA region in *Loxodonta* and *Elephas* also.

To determine the evolutionary history of the woolly mammoth, we sequenced complete mitochondrial genomes of *L. africana* and *E. maximus*. To exclude the possibility of DNA contamination, the elephant specimens were obtained only after the primary mitochondrial genome sequence of *M. primigenius* was completed (Materials and Methods). Long-range PCR was initially used for the amplification of the overlapping ~3–5 kb mtDNA fragments of African and Asian elephants. The sequence analysis was also conducted for cloned PCR fragments and short PCR fragments as described in mammoth mtDNA analysis. The *E. maximus* and *L. africana* mitochondrial genomes sequenced in this study were highly similar but not identical to those submitted to GenBank previously.

Similar to other placental mammals, the mammoth mitochondrial genome contains 13 protein coding genes, 22 tRNA and two rRNA genes, and the D-loop control region (Figure 3). The length of the genome varied due to somatic variability of tandem DNA repeats (CGCATA)_n in the control region. The length of the mammoth mitochondrial genome was 16,842 bp, if the longest VNTR tract was included (Figure S4). Start and stop codons were identical in *M. primigenius*, *L. africana*, and *E. maximus*; however, we found a substitution (G→A) in the ND4 gene leading to a premature stop codon in the C-terminus of the protein. Thus, the ND4 protein is four amino acids shorter in *M. primigenius* than in elephants. A multiple alignment of this gene showed that the C-terminal end of the ND4 protein is variable in sequence and in length in other animal species. Substitutions in the mammoth lineage were predominantly transitions (19/339 transversion/transition ratio in mammoth versus 13/351 in *E. maximus* when polarized with *L. africana*). The total number of nucleotide and amino acid differences between the mammoth and *E. maximus* genomes was 722 and 100, and between the mammoth and *L. africana* genomes 780 and 107, respectively (Table S2).

The phylogeny of Elephantinae was inferred using the newly obtained *M. primigenius*, *E. maximus*, and *L. africana* sequences, and other complete mitochondrial genome sequences of *E. maximus* and *L. africana*, while the mitochondrial genomes of the closest extant taxa [16], the dugong (*Dugong dugon*) and the hyrax (*Procavia capensis*), were used as outgroups. The same topology was recovered by a variety of tree reconstruction methods (Figure 4). As expected, individuals from the same species, *E. maximus* and *L. africana*, clustered together on the tree. *M. primigenius* was determined to be a sister species to *E. maximus*, i.e., the woolly mammoth shared a common ancestor with the Asian elephant more recently than with the African elephant. A maximum likelihood (ML) ratio test comparing all three possible topologies of the Elephantinae species corroborates this conclusion (p < 0.01; see Materials and Methods). We also reconstructed the phylogeny of these species by using only individual protein and tRNA genes (tRNA genes are too short and contain too few substitutions). The majority, but not all (Table S3), of trees reconstructed with the sequence of individual genes supported the topology recovered using the complete genome. Although the topology of this tree appears robust, the unavailability of an outgroup more closely related than the dugong and hyrax calls for the sequencing of nuclear *M. primigenius* genes to further test the results reported here.

The resolution of the Elephantinae phylogeny makes it possible to estimate the time of divergence of *M. primigenius* from its sister species *E. maximus*. Such an estimate usually relies on the existence of a molecular clock, i.e., a uniform rate of evolution of all compared clades, and on having calibration points from the fossil record. However, previous studies suggested that the evolution of the Elephantinae mitochondrial genomes may be inconsistent with the molec-
The analysis of complete mtDNA sequences places *M. primigenius* with *E. maximus* on the tree. The Sirenia (*D. dugon*) and Hyaenodonta (*P. capensis*), most closely related species among extant taxa to Elephantinae, were taken as outgroups. Bootstrap values and posterior probabilities were calculated using a Bayesian approach [29,31] assuming a gamma distribution of the rates of evolution across sites with a General Time Reversible model (normal font), HKY model (bold), with a parsimony approach (italic), and by neighbor joining (italic and bold) [28]. The scale is 0.1 substitutions per site. The mitochondrial genomes of *M. primigenius*, *E. maximus* A, *E. maximus* B, *L. africana* A, *L. africana* B, *D. dugon*, and *P. capensis* were used in the analysis.

**Figure 4. Paenungulata Tree and Phylogenetic Relationship of the Woolly Mammoth**

The analysis of complete mtDNA sequences places *M. primigenius* with *E. maximus* on the tree. The Sirenia (*D. dugon*) and Hyaenodonta (*P. capensis*), most closely related species among extant taxa to Elephantinae, were taken as outgroups. Bootstrap values and posterior probabilities were calculated using a Bayesian approach [29,31] assuming a gamma distribution of the rates of evolution across sites with a General Time Reversible model (normal font), HKY model (bold), with a parsimony approach (italic), and by neighbor joining (italic and bold) [28]. The scale is 0.1 substitutions per site. The mitochondrial genomes of *M. primigenius*, *E. maximus* A, *E. maximus* B, *L. africana* A, *L. africana* B, *D. dugon*, and *P. capensis* were used in the analysis.

DOI: 10.1371/journal.pbio.0040073.g004

ular clock [3,8], and so we tested this possibility using complete genomes. Tajima’s relative rate test showed no difference in the rate of evolution of different Elephantinae lineages (unpublished data). Similarly, a likelihood ratio test did not reject the assumption of equal rate of evolution of the *M. primigenius* versus the *E. maximus* lineage using the *L. africana* genome as an outgroup (Materials and Methods). In contrast, the assumption of a molecular clock was rejected when the likelihood ratio test was performed on simulations involving a comparison of the *M. primigenius–E. maximus* clade with the *L. africana* lineage when the dugong and hyrax species were used as outgroups (without molecular clock Lnl.1 = −56661.45; with molecular clock Lnl.2 = −56877.63; \( p < 0.01 \)).

Molecular clock analyses are highly sensitive to the molecular distance of the outgroups, which can affect the correct root placement. Clearly, the dugong and hyrax are not ideal outgroups for these analyses, as their approximate nonsynonymous nucleotide divergence (dn) with the Elephantinae species is ~0.15 while synonymous nucleotide divergence is at saturation (ds ~ 1.2). Thus, the change in the rate of evolution implied by the ML analysis must be treated with caution. However, a simple parsimony assay corroborated the possibility that the mitochondrial genomes of different Elephantinae species evolve at different rates (Table S4). A ML estimate of the number of substitutions per lineage yielded quantitatively similar results (unpublished data).

These observations, coupled with previous reports [3,8], imply that a molecular clock assumption may be inappropriate when estimating the time of divergence of Elephantinae species. Thus, we used a heuristic rate smoothing procedure for ML estimates [17], which takes into account different rates of evolution among different branches of the tree. Using the calibration points of 6 million y for the *E. maximus–L. africana* split and 65 million y for the Elephasidae–Sirenia split (see [8] and references within), the time of divergence of *M. primigenius* and *E. maximus* was estimated as ~4 million y ago (±0.01 s.e.). This estimate must be treated with caution because it relies on the correct paleontological dating of the *E. maximus–L. africana* divergence. Nevertheless, these data clearly indicate that the divergence of the woolly mammoth from the Asian elephant occurred soon after the divergence of their ancestor lineage from African elephants.

To further verify the accuracy of our sequence and to investigate the level of nucleotide diversity in the woolly mammoth population, we compared our sequence to the longest available sequences of the mammoth mitochondrial genome from other individuals. Two CytB gene sequences from mammoth individuals found in north-eastern Siberia (Magadan region, Kirkilyakh Creek near Kolima River) [18] and north-central Siberia (Pyasina River valley in Taimyr Peninsula) [19] show very high similarity, while the 12S RNA gene [19] was identical to the corresponding genes in our mitochondrial genome sequence from Chukotka’s mammoth (Table S5). All these sequences were recovered from muscle specimens enriched in mtDNA and, thus, very likely had no numt contamination. The Chukotka Ermynveem River valley is ~1,000 km from Kirkilyakh Creek and ~2,900 km from the Pyasina River valley. The data suggest a relatively low genetic diversity of mammoth maternal lineages in mammoth population spanning vast territory in Northern Siberia. Two recent studies have also demonstrated the isolation of multiple fragments of mammoth DNA [20,21]. In one report, a novel approach was applied for direct cloning and shotgun sequencing of random DNA fragments. The set of multiple fragments corresponding to partial mitochondrial genome sequence contained sequences with ancient DNA artifact mutations and potential numts and cannot yet be reliably used for comparative analysis of complete mitochondrial genomes [21]. Another group reported the complete mitochondrial genome sequence [20] of a ~12,170-y-old woolly mammoth that was found in the Berelyekh Yakutia region (70°35’ N, 145°00’ E), which is ~900 km from the Ermynveem River valley (68°10’ N, 165°56’ E). A comparison of the two complete genomes revealed a pattern of high similarity with the nucleotide diversity of ~0.3% (Table S5). This diversity may be even lower since sequence errors may be present in the genome sequences recovered from ancient specimens. When the differences between the two mammoth mitochondrial genomes were polarized with *E. maximus* sequence, the number of sequence-specific polymorphisms for the mitochondrial genome sequence determined here was slightly lower (by ~25%) than in the mitochondrial genome sequence described in the other
study [20]. The sequence [20] also showed a higher ratio of nonsynonymous to synonymous substitutions in protein coding genes and a higher number of nucleotide substitutions in the entire genome (C→T/G→A transitions) that are often associated with ancient DNA modifications (Tables S6 and S7). Although these differences were not statistically significant, a large number of nonsynonymous substitutions was observed in the ND2 gene in sequence [20], but not in our sequence (Table S6). The possibility that these changes appeared due to positive selection in one of the mammoth lineages seems highly unlikely, the more reasonable explanation being that the genome sequence from [20] may harbor some post-mortem ancient DNA mutations (most ND2 nonsynonymous mutations were C→T/G→A). However, these artifacts appear to be rare. In sum, the independently determined mitochondrial genome sequences from ~33,000-y-old (present study) and ~12,000-y-old [20] animals were highly similar, and, owing to a few potential errors, the true level of mitochondrial nucleotide variability may be slightly lower than 0.3%.

The low nucleotide diversity of mammoth mitochondrial sequence (π ~ 0.003; Table S5) is an order of magnitude lower than that reported for the overall populations of L. africana (π ~ 0.02) [22] and E. maximus (π ~ 0.017) [23,24], but similar to the values reported for select populations of L. africana (π ~ 0.00084–0.027) [25]. However, these artifacts appear to be rare. In sum, the independently determined mitochondrial genome sequences from ~33,000-y-old (present study) and ~12,000-y-old [20] animals were highly similar, and, owing to a few potential errors, the true level of mitochondrial nucleotide variability may be slightly lower than 0.3%.

The low nucleotide diversity of mammoth mitochondrial sequence (π ~ 0.003; Table S5) is an order of magnitude lower than that reported for the overall populations of L. africana (π ~ 0.02) [22] and E. maximus (π ~ 0.017) [23,24], but similar to the values reported for select populations of L. africana (π ~ 0.00084–0.027) [25]. However, these artifacts appear to be rare. In sum, the independently determined mitochondrial genome sequences from ~33,000-y-old (present study) and ~12,000-y-old [20] animals were highly similar, and, owing to a few potential errors, the true level of mitochondrial nucleotide variability may be slightly lower than 0.3%.

The low nucleotide diversity of mammoth mitochondrial sequence (π ~ 0.003; Table S5) is an order of magnitude lower than that reported for the overall populations of L. africana (π ~ 0.02) [22] and E. maximus (π ~ 0.017) [23,24], but similar to the values reported for select populations of L. africana (π ~ 0.00084–0.027) [25]. However, these artifacts appear to be rare. In sum, the independently determined mitochondrial genome sequences from ~33,000-y-old (present study) and ~12,000-y-old [20] animals were highly similar, and, owing to a few potential errors, the true level of mitochondrial nucleotide variability may be slightly lower than 0.3%.

The low nucleotide diversity of mammoth mitochondrial sequence (π ~ 0.003; Table S5) is an order of magnitude lower than that reported for the overall populations of L. africana (π ~ 0.02) [22] and E. maximus (π ~ 0.017) [23,24], but similar to the values reported for select populations of L. africana (π ~ 0.00084–0.027) [25]. However, these artifacts appear to be rare. In sum, the independently determined mitochondrial genome sequences from ~33,000-y-old (present study) and ~12,000-y-old [20] animals were highly similar, and, owing to a few potential errors, the true level of mitochondrial nucleotide variability may be slightly lower than 0.3%.

The low nucleotide diversity of mammoth mitochondrial sequence (π ~ 0.003; Table S5) is an order of magnitude lower than that reported for the overall populations of L. africana (π ~ 0.02) [22] and E. maximus (π ~ 0.017) [23,24], but similar to the values reported for select populations of L. africana (π ~ 0.00084–0.027) [25]. However, these artifacts appear to be rare. In sum, the independently determined mitochondrial genome sequences from ~33,000-y-old (present study) and ~12,000-y-old [20] animals were highly similar, and, owing to a few potential errors, the true level of mitochondrial nucleotide variability may be slightly lower than 0.3%.

The low nucleotide diversity of mammoth mitochondrial sequence (π ~ 0.003; Table S5) is an order of magnitude lower than that reported for the overall populations of L. africana (π ~ 0.02) [22] and E. maximus (π ~ 0.017) [23,24], but similar to the values reported for select populations of L. africana (π ~ 0.00084–0.027) [25]. However, these artifacts appear to be rare. In sum, the independently determined mitochondrial genome sequences from ~33,000-y-old (present study) and ~12,000-y-old [20] animals were highly similar, and, owing to a few potential errors, the true level of mitochondrial nucleotide variability may be slightly lower than 0.3%.
ochondrial sequences were found. These diverged numts were found when some pairs of primers were used for long-range PCR using elephant DNA isolated by Qiagen kits. These pairs of primers have been used for mammoth mtDNA PCR analysis. The mammoth mtDNA was also isolated from muscle tissue that has a relatively high proportion of mtDNA to nuclear DNA, The elephant numt sequences were determined using a second alignment of all protein-coding genes was used, and then they were sequenced using other sets of primers and PCR products. Finally the complete mitochondrial genome sequences were determined from multiple independent PCR products. Comparison of the sequences determined in this study (animal A) with the corresponding L. africana, B and E. maximus B sequences from the GenBank revealed their high similarity. The divergence of the sequences of these animals is likely due to polymorphisms. The potential sequence errors in the GenBank E. maximus mtDNA and L. africana mtDNA sequences might be of some concern. However, the results of phylogenetic reconstructions were qualitatively unchanged. The same results were obtained using two different priors: the General Time Reversible model (Table S2) and the HKY model (Table S2), which estimates fewer rate parameters. Both models were run assuming a gamma-distribution of substitution rates across sites for 1 million iterations (mcmc ngen = 1000000 in MrBayes). Phylogenetic trees using individual gene sequences were tested with Bayesian inference set to the HKY model (Table S3). An ML ratio test was performed for ML values obtained for the possible topologies of the three Elephantidae species. The topology with M. primigenius and E. maximus as sister species had a log ML of −4240.629, compared with −4245.549 and −4254.866 for the topology linking E. maximus with L. africana and M. primigenius with L. africana, respectively (p < 0.01; ML ratio test). The log ML scores reported here are averages of log ML scores obtained for 10,000,000 iterations of the BEAST program and ran under the HKY model with set tree topologies.

The molecular clock was tested with the likelihood ratio test, using the log likelihood scores obtained with the baseml program in PAML when the dataset was analyzed with the assumption of no molecular clock (clock = 0 in baseml) and a local clock that tested the difference between the species in question (clock = 2 in baseml). To estimate the branch length for synonymous and nonsynonymous substitutions, the branches of interest were assumed to have different nni/M ratios (model = 2 in codeml) and an absence of a molecular clock (clock = 0 in codeml). To estimate the divergence time of M. primigenius, a heuristic rate smoothing procedure for ML estimates [16] was used as implemented in PAML. The number of synonymous and nonsynonymous substitutions (Table S4) was obtained by parsimony using the African elephant as the outgroup for the mammoth–Asian elephant comparison and dugong as the outgroup for the mammoth–African elephant and Asian–African elephant comparisons without correcting for multiple substitutions and with the assumption that the outgroup accurately represents the ancestral state. These results yielded were quantitatively very similar to those obtained with a ML approach that corrected for multiple substitutions (unpublished). Control files for all stand-alone programs ran here and other methodological materials are available on request.

Supporting Information
Figure S1. Mammoth Specimen Found in Siberia
(A) Restored exhibit of mammoth Enmyn leg found in the Emnynveen River valley, Chukotka, northeastern Siberia. (B) The geographic location of the find is designated by the red dot on the map. Found at DOI: 10.1371/journal.pbio.0040073.sg001 (2.5 MB PDF).

Figure S2. Electrophoresis Analysis of the PCR Products Used for Sequencing of M. primigenius Mitochondrial Genome (“MOS Mammoth mtDNA”)
A and B agarose gel panels. The negative controls are shown in (A). * designates the area of oligonucleotide primer dimers formed in PCR. Found at DOI: 10.1371/journal.pbio.0040073.sg002 (631 KB PDF).

Figure S3. Electrophoresis Analysis of the PCR Products Used for Sequencing of M. primigenius Mitochondrial Genome (“UM Mammoth mtDNA”)
A–D agarose gel panels. * designates lines with low amount of PCR products. Found at DOI: 10.1371/journal.pbio.0040073.sg003 (522 KB PDF).
Figure S4. Extreme Heteroplasmy in Hypervariable Region HVR (VNTR Region) in the Control Region of Mitochondrial Genome of *M. primigenius*  
(A) Electrophoresis of PCR products from two DNA extracts (m1 and m22) and (B) corresponding PCR clones with VNTR region mtDNA insertions showed variability in their length. (C) Nucleotide sequences of cloned PCR products with VNTR. The sequence for the longest VNTR was incorporated in a reported complete sequence of *M. primigenius*.

Found at DOI: 10.1371/journal.pbio.0040073.sg004 (455 KB PDF).

Figure S5. Sequencing Electrophoresis Data for the Mitochondrial Genome of *M. primigenius* with High-Quality Chromatograms of Direct Sequencing of PCR Products, in Comparison with the Cloned PCR Fragments

Three mtDNA regions are shown: (A) L11-H11, (B) L17-H17, and (C) L22-H22. Lines 1 and 2 (A) and line 1 (B and C) from the bottom are the PCR products. Other lines are cloned sequences determined from opposite strands. The rare mutations are observed in cloned sequences but not in direct PCR product sequences.

Found at DOI: 10.1371/journal.pbio.0040073.sg005 (503 KB PDF).

Figure S6. Comparison of *M. primigenius* mtDNA Sequences in Individual Clones and Direct PCR Products

The rare, random “ancient DNA” mutations were identified in cloned sequences. No sequences with identical series of mutations were found in individual clones, providing evidence that the PCR products were obtained from the significant amount of original DNA templates. Mutations were confirmed by sequencing of both DNA strands. The G→A and C→T mutations (type II mutations) were prevalent, as is expected for “ancient DNA” modifications.

Found at DOI: 10.1371/journal.pbio.0040073.sg006 (22 KB PDF).

Table S1. Primer Oligonucleotides and Predicted PCR Fragments Used for Sequencing of Complete Mitochondrial Genome of *M. primigenius*, *E. maximus*, and *L. africana*

Redundant primer oligonucleotides were used to amplify mammoth mtDNA. The sequences of mitochondrial genomes of *E. maximus* and *L. africana* were determined using both long-range PCR fragments and short PCR products.

Found at DOI: 10.1371/journal.pbio.0040073.st001 (30 KB DOC).

Table S2. Different Types of Substitutions Revealed by Pairwise Comparisons of Elephantine Genomes

Number of synonymous and nonsynonymous substitutions calculated with the modified Nei-Gojobory method as implemented in MEGA [28]; the results were independent of the transition/transversion ratio.

Found at DOI: 10.1371/journal.pbio.0040073.st002 (25 KB DOC).

Table S3. Bayesian Phylogeny Inference [29] of Elephantine Using Single Mitochondrial Protein Coding and rRNA Genes

Posterior probabilities are shown in parentheses.

Found at DOI: 10.1371/journal.pbio.0040073.st003 (26 KB DOC).

Table S4. Lineage-Specific Substitutions Inferred by Pairwise Comparisons of Elephantine Genomes and Polarized with One Outgroup

We used the sequences of *E. maximus* and *L. africana* determined in this study for these comparisons. Pairwise comparisons of Elephantine species using a single outgroup to polarize substitutions suggested that the *M. primigenius*—*E. maximus* clade evolves faster than the *L. africana* lineage in both synonymous and nonsynonymous sites.

Found at DOI: 10.1371/journal.pbio.0040073.st004 (26 KB DOC).

Table S5. Nucleotide Diversity Inferred by Pairwise Comparisons of Different *M. primigenius* mtDNA Sequences

Sequences are reported in the following studies: DQ316067 (present study), NC_007596 [20], D83047 [18], and D50841–D50842 [19].

Found at DOI: 10.1371/journal.pbio.0040073.st005 (31 KB DOC).

Table S6. Genome-Specific Nonsynonymous and Synonymous Substitutions Inferred by Pairwise Comparisons of Protein-Coding Genes from Two *M. primigenius* Genomes (DQ316067 versus NC_007596) and Polarized with *E. maximus* (DQ316068) Sequence

A large number of nonsynonymous substitutions was observed in the ND5 gene (seventeen nonsynonymous versus two synonymous mutations in NC_007596 accumulated in a short ~360-bp region; whereas the DQ316067 sequence determined in this study had zero nonsynonymous and one synonymous substitution in the same gene).

Found at DOI: 10.1371/journal.pbio.0040073.st006 (38 KB DOC).

Table S7. Genome-Specific Substitutions Inferred by Pairwise Comparison of Two *M. primigenius* Genomes (DQ316067 versus NC_007596) and Polarized with *E. maximus* (DQ316068) Sequence

While the C→T and G→A substitutions may be a sign of ancient DNA modification (most frequent type II mutations), they represent also real polymorphisms. The NC_007596 mammmoth genome shows more of these substitutions, suggesting that a few of these substitutions may have occurred after death of the sampled mammoth; however, this difference is not statistically significant (Fisher’s exact test p > 0.05).

Found at DOI: 10.1371/journal.pbio.0040073.st007 (25 KB DOC).

Accession numbers

The GenBank (http://www.ncbi.nlm.nih.gov) accession numbers for the complete mitochondrial genome sequences determined in this paper are *E. maximus* (DQ316068), *L. africana* (DQ316069), and *M. primigenius* (DQ316067). The GenBank accession numbers for other *M. primigenius* mtDNA sequences are D83047 [18], D50841–D50842 [19], and NC_007596 [20]. The GenBank accession numbers for mitochondrial genomes of *E. maximus* and *L. africana* [33] are NC_005129 and NC_000934, respectively, and for *D. dugon* and *P. capensis* are NC_003514 and NC_004919, respectively.

Acknowledgments

We thank the Natural History Museum, North-Eastern Research Center, Far Eastern Branch of the Russian Academy of Sciences for photographic material of *M. primigenius* leg, V. A. Nikishina for artwork and technical support, Y.B. Yurov, G. Dvoryanchikov, N. Riazanskaya and T. Kolesnikova for technical support, K. Mehren and C. Gray for elephant specimens, and V. V. Solovyev for help with artwork of animal images.

Author contributions.

EIR conceived and designed the experiments, EIR, YKM, and APG performed the experiments, EIR, YKM, and FAK collected and analyzed the data. BAM, MVD, and IC contributed reagents/materials/analysis tools and performed some experiments. EIR, FAK, APG and YKM wrote the paper.

Funding. FAK is supported by the NSF Graduate Research Fellowship.

Competing interests. The authors have declared that no competing interests exist.


