

# Supporting Information

Gilbert et al. 10.1073/pnas.0802315105

## SI Text: Amino Acid Differences Between the M1 and M25 mtDNA Genomes

Sequence data from 13 mammoth mitochondrial proteins showed between 1 and 5 sequence differences between individuals M1 and M25 (one case, *ND4*, showed a C-terminal insertion of five residues). In all cases, the substitutions were conservative. That is, all involve substitutions within the set of hydrophobic residues, extended to include Thr and Ser, except for one: an Asn–Asp change in *ND4*. It was possible to predict structures for four of the proteins by homology modeling techniques [using SWISS-MODEL (1, 2)]. Table S3 shows the details of these calculations. The sequence similarities between parent and target molecules were all at least 78% residue identity, suggesting that the quality of the models should be quite high. (All four parent structures were bovine homologues.)

Pictures of the modeled structures show that the sites of mutation are on the surface. Fig. S2 shows the sites of mutation in the structures for *Cytb* and *COX1*.

**Comparison with Elephant Sequences.** Alignment of *COX1* and *Cytb* sequences between the two mammoth individuals and two of the extant elephant species showed alignments with no insertions or deletions. The sets of sequences are very similar, with almost all

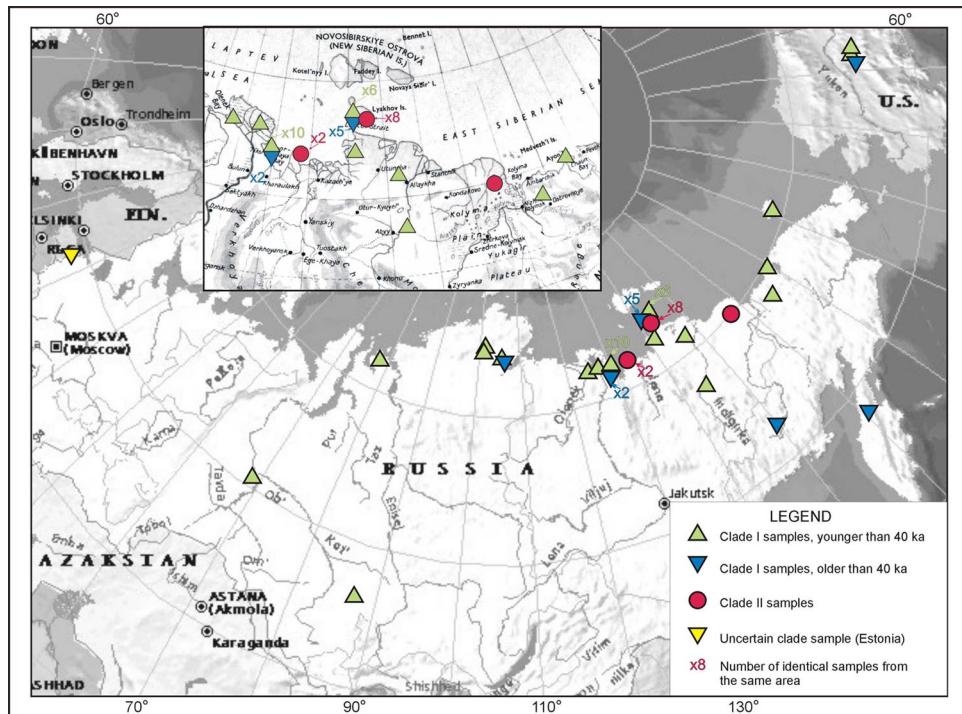
positions constant. When judging only from the *COX1* and *Cytb* protein sequences, the data suggest that mammoths are more closely related to African rather than Asian elephants. In this regard, we find 7 out of 16 sequence differences in *Cytb* and 3 out of 7 in *COX1*, at which the two mammoths and *L. africana* have a common residue but *E. maximus* has a different one. However, at position 371 of the *Cytb* alignment, both mammoths and *E. maximus* have Met, and *L. africana* has Val.

The alignments are shown in Fig. S3. At position 14 of the *CYTB* alignment, M25 has the same residue as the extant elephants, different from M1. At positions 159 and 238 of the *CYTB* alignment, both mammoths have the same residue (159 Asp, 238 Phe), and both extant elephants have the same residue (159 Asn, 238 Leu); a similar distribution appears at residues 329, 330, and 334. At position 303 of the *Cytb* alignment, M25 and M1 have different residues (Ile and Met, respectively), whereas the two extant elephants have Leu. At position 345, M1 and *L. africana* have Tyr but M25 and *E. maximus* have His.

The four *COX1* sequences are very similar until near the C terminus, where the extant elephants have a C-terminal extension.

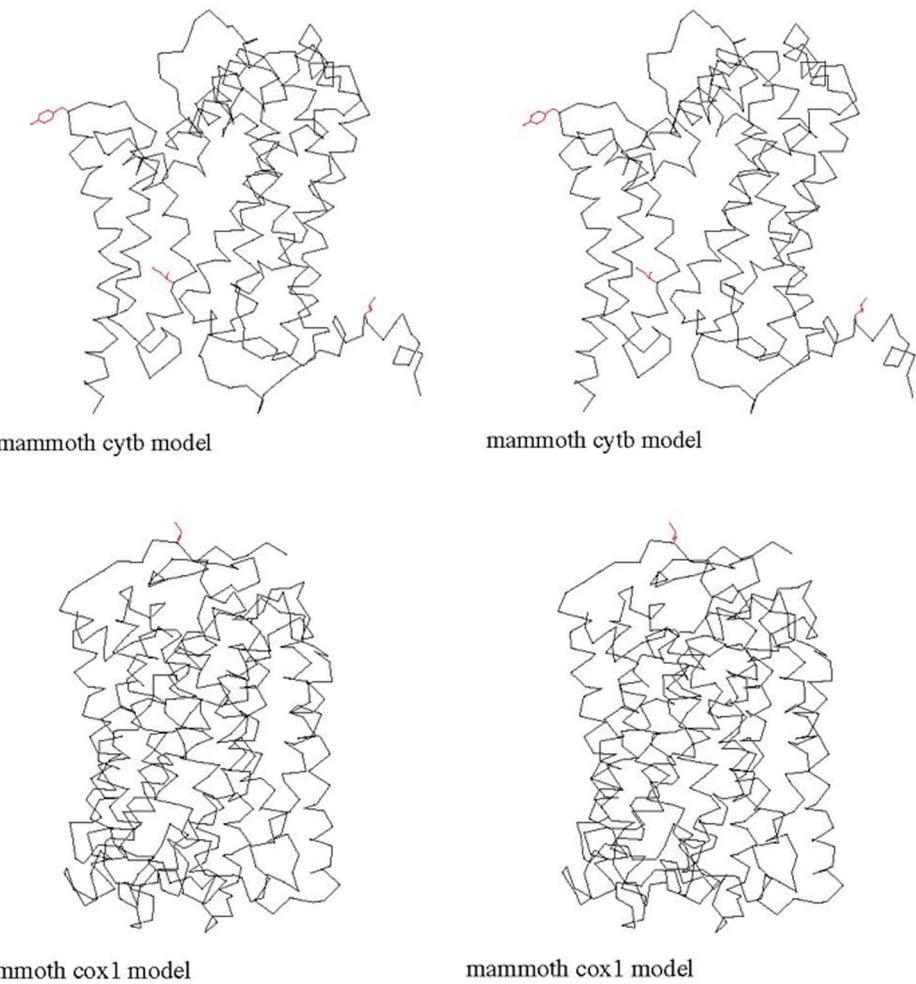
**Conclusion.** There is no reason to suggest that any of the sequence changes is non-neutral.

1. Guex N, Peitsch M.C (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis* 18:2714–2723.
2. Schwede T, Kopp J, Guex N, Peitsch PE (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 31:3381–3385.



**Fig. S1.** Spatial and temporal distribution of clade I and clade II mammoth specimens. Indicated is the geographic distribution of the combined mammoth clade I and clade II specimen dataset, as reported both here and previously by Barnes *et al.* (1). Also indicated is the Estonian mammoth from ref. 1, which is basal to the mammoth phylogeny and may or may not represent a potential unconfirmed third mammoth clade (1) (not discussed here). Because of the high density of samples recovered from the Lena–Kolyma region, this region is expanded in *Inset*. The clade I mammoths are indicated as either younger or older than 40,000  $^{14}\text{C}$  years, to enable comparison of their geographic distribution with the clade II samples, most, if not all, which also date to  $>40,000$   $^{14}\text{C}$  years. The geographical limitation of clade II mammoths in comparison with contemporary clade I mammoths is clearly visible.

1. Barnes I, *et al.* (2007) Genetic structure and extinction of the woolly mammoth, *Mammuthus primigenius*. *Curr Biol* 17:1–4.



**Fig. S2.** Sites of mutation in the structures for Cytb and Cox1.

# Cytochrome b

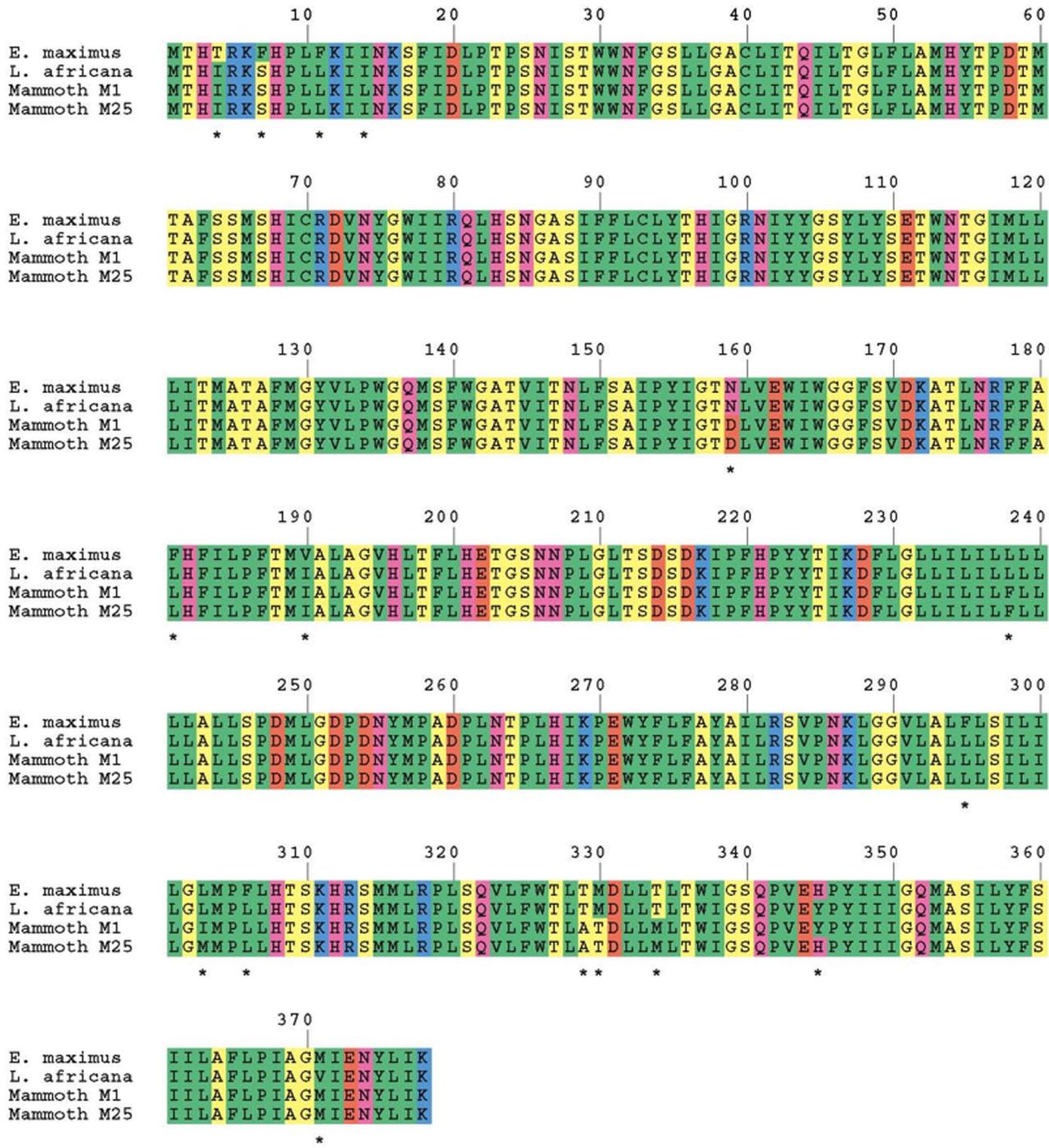
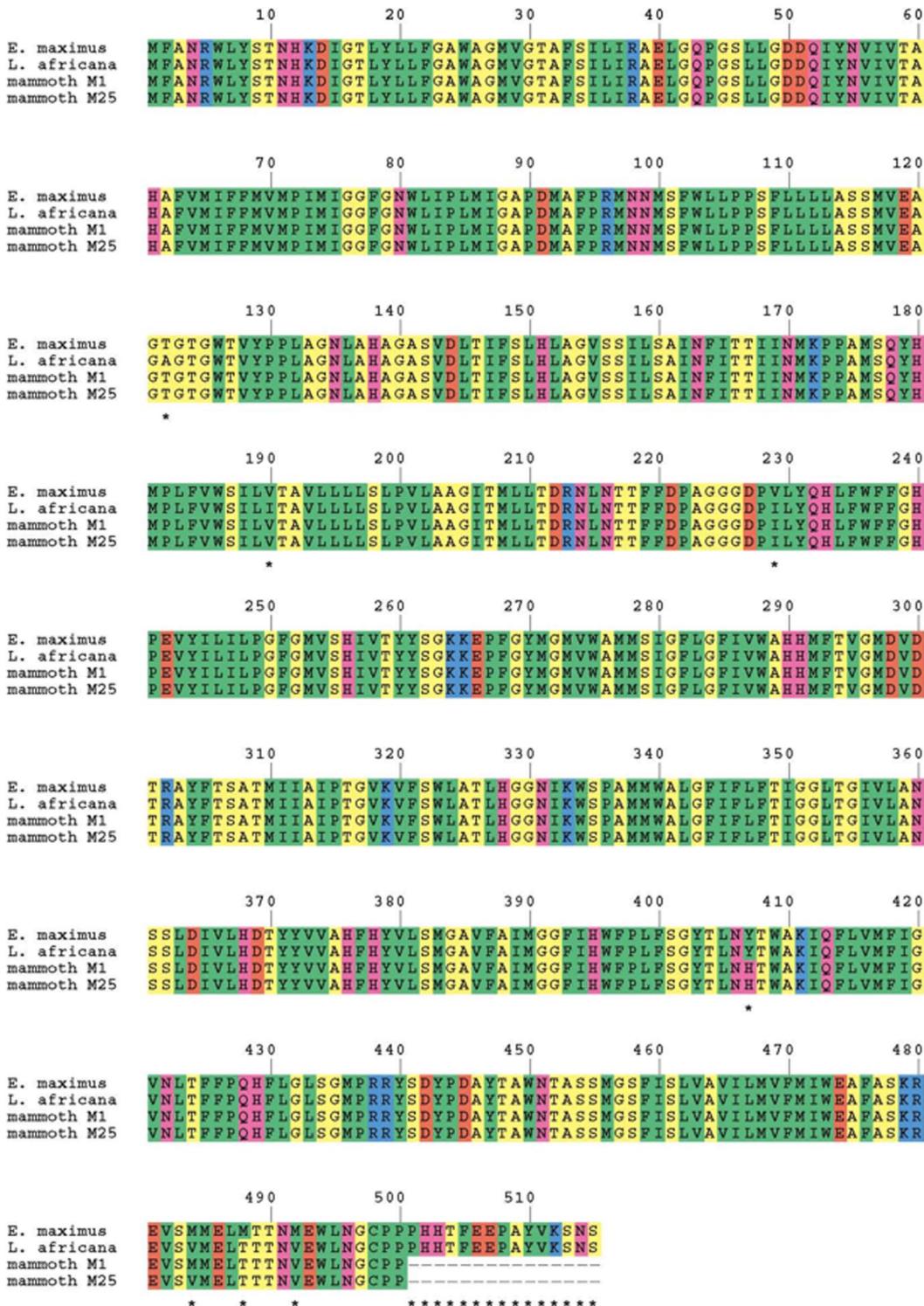


Fig. S3. (Figure continues on the next page.)

COX1



**Fig. S3.** Alignments of elephant and mammoth *Cytb* and *COX1* sequences.

**Table S1. Sample details**

Sample	Museum/ other ID	Geographic origin	Sample source	Amount of material used, g	Size of digestion, ml	GPS east	GPS north	<sup>14</sup> C accession no./source	<sup>14</sup> C age	Collected by	Collection year	Sampled by
M15	Site 5 (ref. 1)	Ayon Island, Russia	Universita Degli Studi Di Parma, Italy	0.43	15	169	69.8	OXA-19605	13,995 ± 55	No information	P. Iacumin available	
M19	Yukagir Mammoth	Yukagir Village, Yakutsk, Russia	Zoological Institute, Russian Academy of Sciences	0.6	15	140.34.873	71.52.988	GrN-28258, GrN-28259, GrN-24288 (ref. 2)	18,560 ± 50	No information	2003 available	A. Tikhonov
M20	LDR-P72	Bolshoy Lyakhovsky Island, Russia	Lena Delta Reserve	0.35	15	142.89	73.64	OXA-19608	>63,500	F. Selyakhov	2000	T. Kuznetsova
M21	LDR-P73	Bolshoy Lyakhovsky Island, Russia	Lena Delta Reserve	0.2	15	143.6	73.21	OXA-19609	>58,000	I. Mikolauskas	2001	T. Kuznetsova
M25	F-0299	Bolshaya Chukochya River, Russia	Ice Age Museum	5.2	30	157.7	69.79	OXA-19610	59,300 ± 2,700	F. Shidlovskiy	2003	A. Sher

1. Iacumin P, Davanzo S, Nikolaev V (2006) Spatial and temporal variations in the <sup>13</sup>C/<sup>12</sup>C and <sup>15</sup>N/<sup>14</sup>N ratios of mammoth hairs: Palaeodiet and palaeoclimatic implications. *Chem Geol* 231:16–28.
2. Aptroot A, van Geel B (2006) Fungi of the colon of the Yukagir Mammoth and from stratigraphically related permafrost samples. *Rev Palaeobot Palyno/* 141:225–230.

**Table S2. Primers and regions PCR amplified and sequenced for specimen M25 to confirm SNPs that differ from the Krause reference mammoth**

Primer set	Forward primer 5' → 3'	First base of F primer*	Reverse primer 5' → 3'	Amplicon size	Annealing temperature
1	TGTAGCTTAAACAAAAGCAAGG	6	GGCGAGACGTATGAGCTAC	215	60
3	GACGGTCAAAGATGGGCTAC	775	AATCTGTAATATGATTGTTGATGTGG	191	60
4	AGTACCGCAAGGGAATGATG	1212	GGCTCGATTGGTATTTCACC	225	60
6	GATCGGATGCCCACTGATAG	1782	GGCCGTTAACACTTGGTCA	221	60
7	TAACAATTGGTGGGGTGA	2235	TCGAGGTGTAACACCTGCT	190	60
8	AGCCGCTATTAAGGGTCTG	2452	CTGTTCTGGATTGGGTGGT	210	60
9	GCAGTAGCCTCCTCACCTT	2786	TCAGACTTGATATTGCGAGAATAAA	289	60
10	ACCATATCCCCTAATTAACTTAAATCT	3013	GGTTAGTGTAAAGGACCCATTATT	217	60
13	AGACAACCAATCCACGATCC	4042	TGGTGCTAATTTGTCAGGTT	278	60
14	GGAATGATCTTAAACCTGACAAAAA	4284	TGAAGTTGTGAGGGTATTGCT	300	60
15	TCCCTAACTCTAATATGAAACAAAACA	4590	AGATAGGGTGATTAGAGTCGGTAA	315	60
17	TCTTGGAGACGACCAAATCT	5471	TGCCTGTCAGCTTCTACT	230	60
18	CACCTTGAGGAGTATCTCTATT	5781	CCTGCAGGGTCAAAGAAAGT	221	60
19	CTACCCCTCATGGCGGTATAA	6304	GAATTGGCTCATGTATGGTTT	264	60
20	TTTATCGGTGTTAATTGACATTCTT	6582	CGTTTGTGTTGTGAGTTCTATCA	223	60
21	CAAGCCAACCTATAACCTTATGC	6911	TGGACGTTTATTATATTGTGTTGGA	286	60
22	TCCACACAAATATAAAACGTCCA	7172	AGCTCCAGAACATTTGGTGT	171	60
23	CCTACAGATCTCCAGTTCG	7453	TGCTGACCGAGCTTCAAAGTA	240	60
24	TCTCCTTAATGAATGGAACGAA	7762	TCGAAGAAGGGCGGATAGTT	202	60
25	TCCAACCTCAAATGCCTAA	8015	TTTATATCGGAAGGCCAGGA	268	60
26	TCACAAAACACATGCCTATCA	8608	AAAAGGCTCAGAACACCTG	299	60
27	CAGGTTCTCTGAGCCTTTA	8886	GCCATAGATACCGTCGGAGA	299	60
28	ACCTGAGCTCATCACAGCTTAAT	9035	ACGTCTACGAAATGTCAATATCAA	308	60
29	CATTACTGATGAGGATCTTATTCTT	9367	GCGATTCTAGGTCAAATAGGAGA	299	60
30	TTTACCCCTGAGCTATCCAAGC	9678	ATTGAGGGATAATAGTCATTAG	346	60
31	TGTTTACCTTAAATGCACTATTATCC	9992	GGCTAGTTAGACTAATGCGAAC	287	60
32	CCTCCGGATCAACTTTCTAG	10417	AATGGATTGGCTAGGTATGG	305	60
33	GGATCTCTCCCCCTACTAGTAACC	10629	AAGGCTATGTGGCTACTGAA	422	60
34	CCATGGCCTAACTCCTCAA	11117	TGTTCTCGGGTAAATGAAGG	350	60
35	TTTACCCGAGAGAACACCTTG	11451	GAAAGCCACATTGTTAGATGAGG	216	60
36	TCAACCCAATCTAACCTTA	11741	CTGTTCGCCGTGCCATC	486	60
37	ACGGACGAACAGACGCTAAC	12215	TGATGGAAGTCACGGATGAA	230	60
38	ACCTCATCCGCTTTATCC	12514	GCTGTTGAAGTAATGGGTATTGTT	341	60
39	GGTTTACTAAAAGACCCAATTAT	12902	CGGAATAGTATTGAGATGAAGA	260	60
40	CGCATACAAACACCCAAATA	13160	GGCTTAGGTAGGGCAGGATT	204	60
41	CCAAAACCTCCCTCACCAAG	13675	GGGGGATCTTCTGGGTCT	260	60
42	CCACTATCATCCACCGAAA	13877	TGGTAGGAGTCCATGTAGAGGT	220	60
43	GACTAATGATCTGAAAACCATCG	14097	TGCTCCGTTGAGTGTAGTTG	315	60
44	CCTACCTATACTCGGAAACCTGAA	14467	AGCCTGTTCTGTGAAGGAAG	297	60
45	ACCTTCCCTCACGAAACAGG	14742	ATCGTAGGATGGCTAAGCA	256	60
46	CCTAGCCCTACTCCTATCAATCC	15023	GGGTTACTTAATGAGGTAGTTCG	268	60
47	GGTATTCAGGGAAGAGGTCCA	15349	CACGAATATGACTGACACATTGA	255	60
49	CCATCTTCGTCGCCCTTTC	15784	ACCAAATGCATGACACCCACA	200	60

\*Numbering with reference to the revised Krause sequence [Krause J, et al. (2006) Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* 439:724–727], GenBank entry DQ188829.2.

**Table S3. Full sequence statistics**

Sample	Tissue	<sup>14</sup> C date	Year collected*	Sequencing technology	% mitochondrial <sup>†</sup>	Contigs <sup>‡</sup>	Fold coverage	Average untrimmed read length <sup>§</sup>	% C → T damage	% Trimmed read identity <sup>  </sup>	% diff. vs. M25 <sup>  </sup>	% diff. vs. 591 bp M1**	% diff. vs. 591 bp M25**	GenBank no.
M1	Hair	n.d.	n.d.	454	1.99	1	48.0	119.0	0.243	99.86	0.00	1.26	0.00	2.54
M2 "Jarkov"	Hair	20,380 ± 140	1997	454	0.76	1	13.2	99.5	0.427	99.79	0.09	1.23	0.51	2.03
M3 "Fishhook"	Hair	20,620 ± 70	1990	454	1.21	1	20.3	128.1	0.347	99.82	0.03	1.24	0.51	2.03
M4	Hair	18,545 ± 70.	n.d.	454	0.65	1	7.8	75.9	0.314	99.85	0.16	1.23	0.68	1.86
M5	Hair	n.d.	n.d.	454	1.30	1	19.8	112.6	0.537	99.76	0.15	1.21	0.68	1.86
M8 "Dima"	Hair	46,900 ± 700	1977	454	1.43	1	27.7	71.1	0.899	99.60	0.14	1.22	0.68	1.86
M13 "Adams"	Hair	35,800 ± 1,200	1799	454	0.76	1	19.1	60.5	0.713	99.73	0.09	1.26	0.34	2.20
M15	Hair	13,995 ± 55	n.d.	454/PCR	0.30	2	4.8	86.5	0.576	99.82	0.07	1.24	0.68	2.54
M18	Hair	17,125 ± 70	n.d.	454	1.50	1	24.6	129.8	0.388	99.83	0.05	1.27	0.34	2.20
M19 "Yukagir"	Hair	18,560 ± 50	2003	454	1.86	1	72.7	199.8	0.202	99.88	0.15	1.22	0.68	1.86
M20 <sup>†</sup>	Hair	>63,500	n.d.	454/PCR	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.14	0.24	2.88	0.34
M21	Hair	>58,000	2001	454	0.43	1	12.7	73.7	0.677	99.65	1.30	0.12	2.71	0.51
M22	Hair	50,200 ± 900	2000	454	2.09	1	17.0	96.8	0.556	99.72	0.17	1.23	0.85	2.03
M25	Hair	59,300 ± 2,700	2003	454	1.43	1	25.4	113.4	0.712	99.64	1.26	0.00	2.54	0.00
M26	Hair	24,740 ± 110	2003	454	0.46	1	7.3	91.2	0.253	99.84	0.22	1.28	1.02	1.52
"Poinar"	Bone	27,740 ± 220	2005	454	0.08	2	7.7	101.1 <sup>‡‡</sup>	1.699	99.51	0.09	1.31	0.34	2.20
"Rogaev"	Muscle	33,750–31,950	Sanger	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.21	1.24	0.85
"Krause"	Bone	12,170 ± 50	Sanger	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.13	1.25	1.02	1.86

Where not newly reported (mammoths M15, M19, M20, M21, and M25), the sample details are taken from refs. 1–5, as described in the main text.

\*Year sample was discovered (where known).

†Number of contigs assembled out of mitochondrial sequences.

‡Average read length before trimming based on Krause (1) sequence.

§Average percent identity with respect to assembly after automatic computational quality processing (i.e., final read used in alignment).

||Percent difference from M1 (clade I representative) and M25 (clade II representative) sequences, respectively.

\*\*Percent difference over the 591-bp intervals used for mammoth–elephant comparison from M1 and M25, respectively.

††Summary statistics could not be calculated due to the low sequencing-by-synthesis coverage of sample M20.

‡‡The sequencing technology (Roche GS 20) used for the generation of the Poinar (3, 4) sequence precluded obtaining longer average read length.

1. Krause J, et al. (2006) Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* 439:724–727.2. Rogaev EI, et al. (2006) Complete mitochondrial genome and phylogeny of Pleistocene mammoth *Mammuthus primigenius*. *PLoS Biol* 4:e73.3. Poinar HN, et al. (2006) Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA. *Science* 311:392–394.4. Gilbert MTP, et al. (2007) Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Res* 35:1–10.5. Gilbert MTP, et al. (2007) Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* 317:1927–1930.

**Table S4. Structure prediction calculations**

Gene name	Parent structure	% sequences similarity to parent	Residues predicted
<i>COX1</i>	1occA	93	1–511
<i>COX2</i>	1v54B	84	2–226
<i>COX3</i>	1v54C	87	3–261
<i>Cytb</i>	2fyuC	78	2–378
<i>ATP8</i>	n.a.		
<i>ATP6</i>	n.a.		
<i>ND1</i>	n.a.		
<i>ND2</i>	n.a.		
<i>ND3</i>	n.a.		
<i>ND4L</i>	n.a.		
<i>ND4</i>	n.a.		
<i>ND5</i>	n.a.		
<i>ND6</i>	n.a.		