LETTERS

Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae

Johannes Krause[1], Paul H. Dear[2], Joshua L. Pollack[3], Montgomery Slatkin[3], Helen Spriggs[2], Ian Barnes[4], Adrian M. Lister[4], Ingo Ebersberger[5], Svante Pääbo[1] & Michael Hofreiter[1]

In studying the genomes of extinct species, two principal limitations are typically the small quantities of endogenous ancient DNA and its degraded condition[1], even though products of up to 1,600 base pairs (bp) have been amplified in rare cases[2]. Using small overlapping polymerase chain reaction products, longer stretches of sequences or even whole mitochondrial genomes[3,4] can be reconstructed, but this approach is limited by the number of amplifications that can be performed from rare samples. Thus, even from well-studied Pleistocene species such as mammoths, ground sloths and cave bears, no DNA sequences of more than about 1,000 bp have been reconstructed[5–7]. Here we report the complete mitochondrial genome sequence of the Pleistocene woolly mammoth *Mammuthus primigenius*. We used about 200 mg of bone and a new approach that allows the simultaneous retrieval of multiple sequences from small amounts of degraded DNA. Our phylogenetic analyses show that the mammoth was more closely related to the Asian than to the African elephant. However, the divergence of mammoth, African and Asian elephants occurred over a short time, corresponding to only about 7% of the total length of the phylogenetic tree for the three evolutionary lineages.

We have developed a multiplex polymerase chain reaction (PCR) approach that in principle allows an entire mitochondrial genome to be amplified from ancient DNA using just two initial amplifications. This is accomplished by using primer pairs that define overlapping DNA sequence fragments representing the complete mitochondrial genome. These primer pairs are combined into two sets, each containing every second primer pair. Each of these two sets is used in a multiplex PCR amplification that requires the same amount of ancient template DNA as would be used for amplifying a single target sequence. Subsequently, the two primary amplifications are diluted and used as templates in secondary PCR reactions, in which each product is amplified individually.

To test whether this approach works for Pleistocene DNA, we designed 46 primer pairs, which together we expected to amplify the entire mitochondrial (mt) genome of the woolly mammoth (Fig. 1; see Supplementary Information). Forty-one of the secondary amplifications yielded products of the expected size (Fig. 2 and Supplementary Information). Of the five PCR amplifications that failed, four were successful in later attempts, suggesting that the initial failure was due to the absence of even a single template molecule in the PCR amplification. The fifth PCR failed repeatedly, and inspection of the sequences of the overlapping amplimers revealed differences at the priming sites between the mammoth and elephant sequences that can account for the failure. New primers were designed for this amplimer on the basis of the flanking mammoth sequence. The sequences of all 46 amplimers were found to be similar to (but distinct from) corresponding sequences of extant elephants.

To ensure the authenticity of the ancient DNA sequences[1], we repeated the primary and secondary amplifications and the sequencing of the amplimers (see Methods). In 17 of the amplimers, we found differences at 1–5 nucleotide positions between the two independent experiments, presumably due to cytosine deamination of the template[8] (Supplementary Information). In these cases, a third round of primary PCR amplifications was performed, and the consensus of the three sets of sequences was inferred to represent the correct sequence. In total, fourteen primary PCR amplifications, corresponding to ~200 mg of mammoth bone, were required for determination and confirmation of the entire mtDNA sequence.

To further confirm the reproducibility of the results, samples of the bone were extracted, amplified and sequenced in laboratories in Cambridge and London, UK, that did not have access to the initial results obtained in Leipzig. In total, 5,024 bp of the mitochondrial sequence were independently reproduced. The reproduced sequences

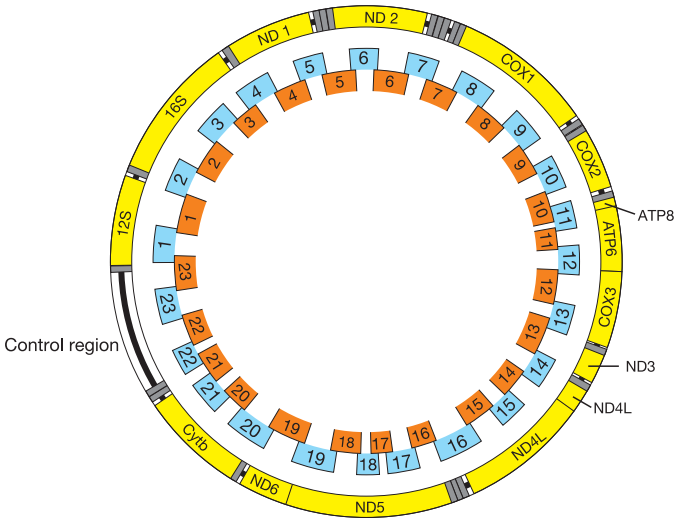

Figure 1 | Map of the mammoth mitochondrial genome. Circular genome (yellow), showing the positions of the control region and the genes encoding 22 transfer RNAs (grey boxes), 2 ribosomal RNAs and 13 proteins. The positions and relative lengths of the 46 amplification products used are depicted in blue (first set) and red (second set).

[1]Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany. [2]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK. [3]Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA. [4]Department of Biology, University College London, London WC1E 6BT, UK. [5]WE Informatik, Heinrich Heine Universität, Universitätsstr. 1, D-40225 Düsseldorf, Germany.

were identical to those initially determined, except for one position determined from a single PCR reaction in London that shows a thymine (T), when the sequence determined from several PCR amplifications in Leipzig shows a cytosine (C). This discrepancy is most likely due to cytosine deamination of the template DNA[8].

One concern when determining mtDNA sequences is that nuclear insertions of mtDNA fragments might be mistaken for the organelle copies[9]. This problem can be particularly severe in some species, including elephants[10]. However, the sequence we have determined is circular and therefore cannot represent a nuclear insertion. Although it is conceivable that some of the amplimers could have arisen from nuclear insertions, no mismatches were observed in a total of 2,030 overlapping base pairs between adjacent amplimers. We thus conclude that the results obtained indeed represent the complete mtDNA sequence of the mammoth from Berelekh.

The phylogenetic relationship of the mammoth to its closest living relatives, the African and Asian elephants, has until now been unresolved. Numerous studies using small mitochondrial and nuclear DNA[11] sequences have variously suggested that mammoths were the sister group of African[12–14] or of Asian elephants[7,11,15]. It has been argued that the failure to resolve this phylogenetic relationship is due to the small amount of DNA sequence information available[13], because long mtDNA sequences are often necessary to obtain a robust phylogenetic tree[16]. To address this question, we aligned the entire 16,770-bp mammoth mtDNA with mtDNA from the two elephant species and two potential outgroup species, dugong and hyrax. Phylogenetic trees identified either the Asian or African elephant as the sister taxon of the mammoth, depending on the outgroup and tree reconstruction method used (Supplementary Table S2).

Two lines of evidence indicate that this lack of resolution results from a too-distant relationship of elephants and mammoths to the two outgroup species, both of which diverged from the proboscidean lineage at least 65 million years ago[17]. First, about half of the third-codon positions in the protein-coding sequences differ when the mammoth or either elephant species is compared to dugong or hyrax. Second, the ratio of transitions to transversions among elephants and mammoth is 20:1, but is 2:1 between these species and dugong or hyrax. Thus, the phylogenetic signal is blurred by multiple substitutions when dugong or hyrax is included in the analysis[18].

As no closer outgroup species exists, we restricted our analysis to the mammoth and the two elephant species using three different methods that assume a molecular clock (that is, a uniform rate of nucleotide substitutions in all lineages). We first tested whether the mtDNAs of the three species evolved at equal rates using a clock test that uses only three taxa and treats transitions and transversions separately[19], as well as by a likelihood ratio test. The assumption of a molecular clock among the mammoth and two elephant mtDNAs could not be rejected ($P = 0.61$). We therefore estimated a maximum-likelihood tree under the assumption of a molecular clock, applying midpoint rooting. This analysis supports a sister-group relationship between mammoth and Asian elephant with a 97% bootstrap value. We then calculated the likelihood of the data assuming a star-like tree topology. A likelihood ratio test reveals that the more complex model, resulting in the resolved sequence tree, explains the data significantly better, and thus rejects the simpler model of a star-like tree ($P < 0.01$). Notably, the tree that places the Asian elephant and mammoth as the closest relatives to each other has a posterior probability of 99.8%. This indicates that the two alternative scenarios, each having posterior probabilities of ~0.1%, can be neglected. Finally, we used Felsenstein's[20] method for testing whether the most parsimonious tree is significantly different from a tree with a 'trifurcation' (Fig. 3a). Again, we rejected the null hypothesis of a trifurcation ($P = 0.032$ and $P = 0.035$ for the C and S statistics, respectively; see Methods). Thus, the monophyly of the mammoth and the Asian elephant, to the exclusion of the African elephant, is significantly supported.

In this context, two points should be noted. First, the length of the internal branch leading to the mammoth and Asian elephant mtDNA is only 7.3% of that leading to the African elephant (Fig. 3b). Paleontological data[21] suggest a divergence of Asian and African elephants and mammoths about six million years ago in Africa. This

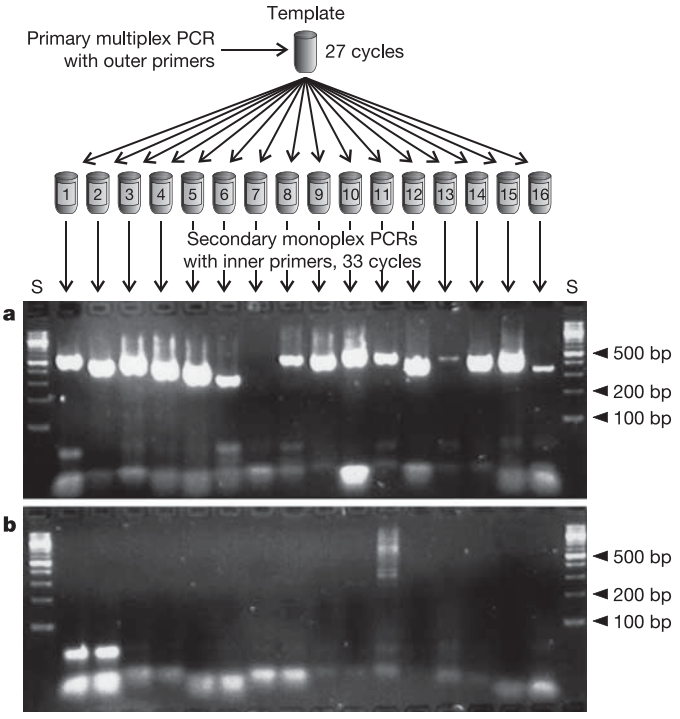

Figure 2 | Principle of the multiplex approach and typical results. Amplification results for the mammoth ancient DNA extract, with 15 of 16 amplifications yielding positive results (**a**). Negative controls for the same 16 primer pairs are shown in **b**. Gel lanes marked with S contain molecular size markers. The size of the amplification products ranges from 306 to 554 bp.

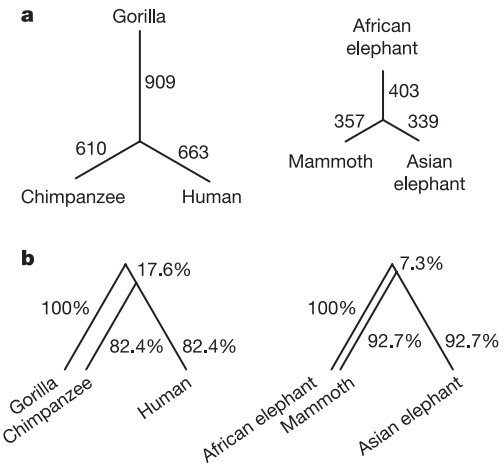

Figure 3 | Comparison of three-species trees. Comparisons based on the complete mtDNA sequences for human, chimpanzee, gorilla, mammoth and African and Asian elephants. **a**, Absolute number of phylogenetically informative positions on which the three-species test is based. **b**, Relative length of the internal branch in the two groups of species for clock-like phylogenetic trees using the substitution data from **a** and rooted by midpoint rooting. Note the differences in the absolute numbers of substitutions and relative branch length of the internal branch between the two species groups (see also Supplementary Information).

date implies that the divergence between the mammoth and Asian elephant took place only 440,000 years after the divergence of the African elephant. Second, this time is short enough that polymorphisms in the ancestral species may have persisted between the two speciation events[22], as has been observed for humans, chimpanzees and gorillas[23]. As the probability of such events increases with increasing effective population size[24], mtDNA is more likely to reflect species phylogeny than autosomal genes that have an approximately four times larger effective population size. The sister-group relationship of the mtDNAs of the mammoth and the Asian elephant therefore suggests that these two species shared a common ancestor after their separation from the African elephant (Supplementary Information). Thus, any morphological similarities between the African elephant and mammoth[13] are likely to be either ancestral or convergent. However, owing to the short time between the two divergence events, this is also likely to be true for many morphological similarities between the Asian elephant and mammoth.

In summary, the multiplex approach described here allows the retrieval of complete mtDNA sequences from Late Pleistocene fossils, enabling robust phylogenetic inferences to be made. The small amounts of sample material necessary allow complete mtDNA sequences to be determined even from very valuable specimens. This approach makes it possible to sample the mitochondrial genomes of Pleistocene animals as widely as those of extant species, and provides the opportunity to answer detailed questions about the structure and history of extinct populations.

METHODS

Detailed descriptions of the methods used to date the mammoth bone and extract, amplify and sequence its DNA, together with the phylogenetic methods used, are provided in the Supplementary Information.

DNA extraction and amplification. For DNA extraction we used 747 mg bone powder from a mammoth bone found in Berelekh, Yakutia (Supplementary Information), yielding 70 μl of extract. We used 1.5 μl of the DNA extract for each multiplex PCR reaction, in a total volume of 20 μl. After 27 cycles of PCR, the appropriate multiplex product was diluted and 0.625% of this material was used for each of 46 individual amplifications. Amplification products of the correct size were cloned using the TOPO TA cloning kit (Invitrogen), and a minimum of three clones were sequenced on an ABI3730 capillary sequencer (Applied Biosystems).

We designed new primers for those fragments that gave only weak (or no) amplification products in the first attempts and for which the sequences of adjacent fragments showed differences from the elephant sequence in the primer sites. These new primers, together with the previously successful primers, were used to amplify the remaining segments of the mammoth mtDNA and to replicate all positions at least twice from independent primary amplifications (see also Supplementary Information). Extraction protocols and PCR conditions for the laboratories in Cambridge and London are available in the Supplementary Information.

Phylogenetic analyses. We initially aligned the mtDNA sequence of the mammoth to the corresponding sequences from Asian (*Elephas maximus*) and African (*Loxodonta africana*) elephant and dugong (*Dugong dugon*), excluding the control region in all analyses. Applying Modeltest[25] on this alignment, we obtained a general time-reversible substitution model with gamma-distributed substitution rates across sites. We estimated neighbour-joining and maximum-likelihood trees correcting for multiple substitutions using this substitution model and maximum parsimony trees using the program package PAUP*[26], and bayesian trees using MrBayes[27]. Depending on the tree-building method and outgroup, we recovered both the Asian and African elephant as sister taxa of the mammoth. We therefore added the mtDNA sequence from hyrax (*Procavia capensis*) as additional outgroup in an attempt to gain resolution. Again, we could not resolve the phylogeny (Supplementary Table S2).

Our analyses indicate that both dugong and hyrax are too distantly related to the mammoth and African and Asian elephants to resolve the phylogeny. However, by assuming a molecular clock, rooted phylogenies can be obtained without outgroups. We first tested the clock assumption by applying a simple test that uses only three taxa[19]. We also performed a clock test based on a likelihood ratio test with the program TREE-PUZZLE[28], using the Hasegawa–Kishino–Yano (HKY) model of DNA sequence evolution[29] and a gamma-distribution using eight gamma rate categories to model substitution rate heterogeneity among sites. Neither of the two tests could reject the clock assumption. We subsequently used a parsimony method developed to estimate the phylogenetic relationship of three sequences that evolve under a molecular clock[20] and a likelihood ratio test (Supplementary Information) in order to resolve the relationship between mammoth, African and Asian elephant.

The parsimony method uses two statistics, C and S in Felsenstein's notation[20]. Both are based on the number of positions at which one of the three sequences differs from the other two (which are identical to each other). C is the highest number of such positions obtained from one of the three possible comparisons n_1, n_2 and n_3 (n_1, species A differs from species B and C; n_2, B differs from A and C; n_3, C differs from A and B). S is described by $n_1 - n_2$, given that n_1 and n_2 are the largest and second largest numbers obtained for n, respectively. Excluding the control region, 403 (n_1) positions support a sister-group relationship between the mammoth and Asian elephant, 357 (n_2) positions support a sister-group relationship between the mammoth and the African elephant, and 339 (n_3) positions support a sister-group relationship between the two elephant species. This yields a C statistic of 403 ($C = n_1$) and an S statistic of 46 ($S = n_1 - n_2$)[20]. On the basis of a Monte Carlo randomization test (10^6 simulations), the monophyly of the mammoth and the Asian elephant, to the exclusion of the African elephant, is significantly supported by both the C and the S statistic ($P = 0.032$ and $P = 0.035$, respectively).

To test whether the phylogenetic signals in the data are strong enough to warrant resolution of the sequence tree under a likelihood framework, we proceeded as follows. The likelihood of the data was assessed under two alternative models: (1) a simple model with only a single free parameter, corresponding to a star-like tree topology, and (2) a more complex model with two free parameters, resembling the resolved tree topology. A subsequent likelihood ratio test revealed that the more complex model, yielding the resolved sequence tree, explains the data significantly better, and thus rejects the simpler model of a star-like tree ($P < 0.01$).

Received 1 July; accepted 14 November 2005.
Published online 18 December 2005.

1. Pääbo, S. *et al.* Genetic analyses from ancient DNA. *Annu. Rev. Genet.* **38,** 645–679 (2004).
2. Lambert, D. M. *et al.* Rates of evolution in ancient DNA from Adelie penguins. *Science* **295,** 2270–2273 (2002).
3. Cooper, A. *et al.* Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409,** 704–707 (2001).
4. Haddrath, O. & Baker, A. J. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proc. R. Soc. Lond. B* **268,** 939–945 (2001).
5. Höss, M., Dilling, A., Currant, A. & Pääbo, S. Molecular phylogeny of the extinct ground sloth *Mylodon darwinii*. *Proc. Natl Acad. Sci. USA* **93,** 181–185 (1996).
6. Loreille, O. *et al.* Ancient DNA analysis reveals divergence of the cave bear, *Ursus spelaeus*, and brown bear, *Ursus arctos*, lineages. *Curr. Biol.* **11,** 200–203 (2001).
7. Yang, H., Golenberg, E. M. & Shoshani, J. Phylogenetic resolution within the Elephantidae using fossil DNA sequences from the American mastodon (*Mammut americanum*) as an outgroup. *Proc. Natl Acad. Sci. USA* **93,** 1190–1194 (1996).
8. Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, A. & Pääbo, S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **29,** 4793–4799 (2001).
9. Bensasson, D., Zhang, D. X., Hartl, D. L. & Hewitt, G. M. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol.* **16,** 314–321 (2001).
10. Greenwood, A. D. & Pääbo, S. Nuclear insertion sequences of mitochondrial DNA predominate in hair but not in blood of elephants. *Mol. Ecol.* **8,** 133–137 (1999).
11. Greenwood, A., Capelli, C., Possnert, G. & Pääbo, S. Nuclear DNA sequences from late Pleistocene megafauna. *Mol. Biol. Evol.* **16,** 1466–1473 (1999).
12. Noro, M., Masuda, R., Dubrovo, I. A., Yoshida, M. C. & Kato, M. Molecular phylogenetic inference of the woolly mammoth *Mammuthus primigenius*, based on complete sequences of mitochondrial cytochrome *b* and 12S ribosomal RNA genes. *J. Mol. Evol.* **46,** 314–326 (1998).
13. Thomas, M. G., Hagelberg, E., Jone, H. B., Yang, Z. & Lister, A. M. Molecular and morphological evidence on the phylogeny of the Elephantidae. *Proc. R. Soc. Lond. B* **267,** 2493–2500 (2000).
14. Debruyne, R., Barriel, V. & Tassy, P. Mitochondrial cytochrome *b* of the Lyakhov mammoth (Proboscidea, Mammalia): new data and phylogenetic analyses of Elephantidae. *Mol. Phylogenet. Evol.* **26,** 421–434 (2003).
15. Ozawa, T., Hayashi, S. & Mikhelson, V. M. Phylogenetic position of mammoth and Steller's sea cow within Tethytheria demonstrated by mitochondrial DNA sequences. *J. Mol. Evol.* **44,** 406–413 (1997).
16. Cummings, M. P., Otto, S. P. & Wakeley, J. Sampling properties of DNA sequence data in phylogenetic analysis. *Mol. Biol. Evol.* **12,** 814–822 (1995).

17. Shoshani, J. Understanding proboscidean evolution: a formidable task. *Trends. Ecol. Evol.* **13,** 480–487 (1998).
18. Felsenstein, J. *Inferring Phylogenies* 196–221 (Sinauer Associates, Sunderland, 2004).
19. Tajima, F. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135,** 599–607 (1993).
20. Felsenstein, J. Confidence-limits on phylogenies with a molecular clock. *Syst. Zool.* **34,** 152–161 (1985).
21. Tassy, P. in *Lothagam: The Dawn of Humanity in Eastern Africa* (eds Leakey, M. G. & Harris, J. M.) 331–358 (Columbia Univ. Press, New York, 2003).
22. Nei, M. in *Evolutionary Perspectives and the New Genetics* (eds Gershowitz, H., Rucknagel, D. L. & Tashian, R. E.) 133–147 (Alan R. Liss, New York, 1986).
23. Chen, F. C. & Li, W. H. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* **68,** 444–456 (2001).
24. Tajima, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105,** 437–460 (1983).
25. Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14,** 817–818 (1998).
26. Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer Associates, Sunderland, 2003).
27. Ronquist, F. & Huelsenbeck, J. P. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572–1574 (2003).
28. Schmidt, H. A., Strimmer, K., Vingron, M. & Haeseler, A. TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18,** 502–504 (2002).
29. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22,** 160–174 (1985).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the members of our laboratories for discussions and support, G. Khlopatchev for providing the mammoth bone and K. Finstermeier for help with figure design. This work was supported by the Max Planck Society. J.L.P. and M.S. were supported by an NIH grant (to M.S.).

Author Information The complete mammoth mitochondrial DNA sequence has been deposited in GenBank under accession number DQ188829. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.H. (hofreiter@eva.mpg.de).