

Supplemental Data

Out of America: Ancient DNA Evidence

for a New World Origin

of Late Quaternary Woolly Mammoths

Regis Debruyne, Genevieve Chu, Christine E. King, Kirsti Bos, Melanie Kuch, Carsten Schwarz, Paul Szpak, Darren R. Gröcke, Paul Matheus, Grant Zazula, Dale Guthrie, Duane Froese, Bernard Buigues, Christian de Marliave, Clare Flemming, Debi Poinar, Daniel Fisher, John Southon, Alexei N. Tikhonov, Ross D. E. MacPhee, and Hendrik N. Poinar

Supplemental Experimental Procedures

DNA extractions

A total of 135 mammoth samples were collected from bone (nB = 84) and molar tooth lamellae or roots (nT = 51), and prepared using either a hammer and chisel or a dremel tool. Depending upon original conditions of storage, samples were kept in dark conditions in a -20°C freezer or at room temperature (see Table S1).

DNA was extracted in a dedicated clean room separated from PCR, post-PCR and sequencing facilities

(<http://socserv.mcmaster.ca/adna/labs.htm>). Samples (100 mg per sample \pm 20%) were processed in batches of 12 (including one extraction blank containing no sample) starting with demineralization in 1.5 mL of EDTA (0.5 M, pH = 8.0) rotating overnight at room temperature in the dark. After collecting the solid remains by

centrifugation and removal of the supernatant, the decalcified bone material was digested at 50–55°C rotating overnight, again in the dark, in 1.5 mL of bone lysis buffer as previously described [1]. The digested pellet was extracted with 500 µL of phenol/chloroform/isoamyl alcohol (25:24:1) and 500 µL of chloroform before concentration to a final volume of 100 µL in 0.1X TE (pH 8.0) by ultrafiltration using a Microcon YM-30 column (Millipore, USA).

Hierarchical quantitative screening method

The yield of endogenous DNA and the level of PCR inhibition of the extracts (used straight and diluted 1:10) was assessed through an 84 bp elephantid-specific qPCR assay targeting a fragment of the mitochondrial *cytochrome b* gene [1]. No contamination was observed in any of the extraction blanks. Based on absolute quantitation (as measured by the Ct value relative to a standard curve of known-concentration), extracts were assigned to four categories: type 1, > 10k copies (per reaction using 5 µL of template); type 2, > 1k to ≤ 10k; type 3, > 100 to ≤ 1k; and type 4, < 100. Of the 26 type 4 samples, only 12 (less than 9% of the total number of samples) yielded no amplifiable mammoth DNA in the quantitative PCR assay. Type 4 extracts were not processed further; extracts in other categories were subjected to PCRs of different

amplification lengths based upon their quality in order to complete the full length of the selected 743 bp fragment of the mitochondrial genome (see below).

Specific amplifications, cloning and sequencing

The mitochondrial fragment used for phylogenetic analysis spans the 3' end of the *cytochrome b* gene, *threonine* and *proline* tRNAs, and HVR1 of the control region. It was selected based on its high variation level as documented in phylogeographic surveys of both Asian [2] and African elephants [3] with which it largely overlaps. It also shows 705 bp in common with the fragment used by Barnes et al. [4] in the first large-scale study of the woolly mammoth diversity.

Primers were designed using PRIMER-XPRESS software (ver1.5; Applied Biosystems, Foster City, California) using the first two mitochondrial sequences published for the woolly mammoth [5, 6] and optimized for annealing temperature to amplify specific elephantid sequences. Ten primers (numbered 1–10, odd numbers assigned to forward direction, even numbers to reverse direction; Table S2) were used in different combinations to generate either a single large PCR product (primers 1 and 10; 788 bp, primers included) targeted for type 1 extracts, medium (334–448 bp) or small (190–237 bp)

overlapping fragments targeted respectively for type 2 and type 3 extracts (Table S3). However, deviations from this amplification model did occur. All PCRs attempted are documented in Table S1.

PCR reactions were carried out in 20 μ L reaction volume with 1X PCR-buffer, 2.5 mM MgCl₂, 250 μ M of each dNTP, 300 nM of each primer, 1 unit of taq polymerase (AmpliTaq Gold; Applied Biosystems, Foster City, California), and 5 μ L of template. Cycling conditions were as follows: initial denaturation at 95°C for 5', 45 cycles of denaturation 30'' at 95°C, annealing 30'' at 50–60°C (Table S3), extension at 72°C for 40–60'' (Table S3) and a final extension at 72°C for 7'. Products of amplification were visualized under UV light on 2% agarose gels stained with ethidium bromide. When applicable, positive reactions showing dimer formations and/or secondary products were excised directly from the gel with glass pipets. After gel dissolution in 0.1X TE for 10' at 50°C, those templates were reamplified with the same primer pair.

To ensure accuracy, each position in a given final consensus sequence was covered by at least two independent amplifications (of the same or different fragment). Each PCR product was cloned using the TOPO-TA cloning kit with TOP-10 chemically competent cells in a scaled-down reaction volume to a quarter of the manufacturer's recommendations (Invitrogen, Canada). Insert-carrying clones were

identified via blue/white selection and amplified via colony PCR using the M13 F/R primer pair (with annealing temperature set at 57°C). Colony PCR products were purified using either AcroPrep 96 Filter Plates (Pall, USA) or Multiscreen PCR μ 96 plates (Millipore, USA). Except where colony PCR failed to provide enough positive clones, a minimum of two positive clones per independent PCR were sequenced using the M13F primer on an ABI 3130, as per manufacturer's suggestions for DNA concentration but 1 in 7 μ L reactions using 0.3 μ L of BigDye ver1.1 (Applied Biosystems, Foster City, California).

Sequences were aligned using the clustal algorithm (with default options) as implemented in BIOEDIT (ver5.07; [5]) with no editing other than removal of the primer sequences. In order to construct a consensus sequence for each specimen, we applied a nested consensus approach (Figure S1): all the clones derived from a single PCR were used to design a step 1 consensus; step 1 consensus sequences were then amalgamated to produce the step 2 consensus sequence. This approach enabled detection of polymerase errors occurring in the early stages of amplification as well as consistent DNA damage among clones from PCRs resulting from amplification of very low-template molecules. Whenever step 2 consensus sequences displayed undetermined positions, we

attempted additional PCR amplifications until all uncertainties were accounted for.

Sample 2005/915 (Cerpolex, table S1) was used by Poinar et al.[1] in their metagenomic study on the GS-20 sequencer (Roche Applied Science, USA). In this study, we retrieved the same hypervariable region sequence as Poinar et al. in the original assembly of the 7X coverage of the mitochondrial genome, using only 2 clones, each from 2 independent amplifications of the 788 bp amplification. Fourteen additional clones of the first amplification were sequenced and confirmed our consensus, hence validating our approach, far more cost-effective than using a metagenomic approach for our purpose. Specimen AM8744 (AMNH, Table S1) had been directly sequenced previously [4]. Although our consensus approach retrieved the original sequence, the limitations of direct sequencing versus cloning will be more fully explored elsewhere (Debruyne and Poinar, submitted). One sequence derived from Barnes et al. (GENBANK#AM746179) had to be removed from the alignment because it contained two ambiguities at variable positions that prevented its assignment to any haplotype and we think could be the result of damage to the original template.

Collagen extraction and ^{14}C dating

Of the samples used in this study, 61 originated from specimens that had been previously dated by AMS ^{14}C dating methods; of these, 42 were successfully sequenced. In order to augment our database, 44 samples for which DNA sequences were successfully generated were dated at the KCCAMS [Keck Carbon Cycle Accelerator Mass Spectrometry] Laboratory (Table S1), of which five represented updates of previously dated specimens (Table S4). Preparation of bone/tooth collagen was performed according to the KCCAMS protocol for chemical pretreatment of bone (version May 2007, available online at <http://www.ess.uci.edu/ams/>). Radiocarbon ages have been corrected for sample preparation blanks equivalent to 45–50 ka, as determined using aliquots of ultrafiltered collagen from a whalebone older than 70 ka.

When combined with previously published ^{14}C -dated sequences (Table S5), our dataset of 131 dates (105 of which are finite), constitutes the second largest assembly of dated ancient DNA sequences so far [8]. All those ^{14}C dates were calibrated in calendar years using the Calpal-Hulu calibration curve implemented in CALPAL^{online} (<http://www.calpal-online.de/>; [6]).

Of the 44 newly dated mammoth samples, 5 had previously been dated but had yielded very wide confidence intervals [7] or displayed limited between-date consistency [8]. Table S4 summarizes these results and shows the improvement in the dating of those specimens. The ultrafiltration step, used in our sample preparation protocol for AMS dating, filters out degraded peptides while leaving only high molecular weight collagen fragments. This step greatly improves consistency in the quality of the “collagen” (i.e., various bone proteins) recovered and therewith the precision of age estimates.

Sequence Alignments

Three different NEXUS matrices were used for the phylogenetic and networking analyses. The first matrix contained the most complete dataset, consisting of 705 bp for 166 terminals including 3 African elephants representing both forest and savannah haplogroups ([3]; GENBANK accession numbers AF132528, AY359278 and DQ316069); 3 Asian elephants representing the two main clades ([2]; accession numbers AF135520, AF135521 and NC005129); and 160 individual woolly mammoths, 108 from this study (Table S1) and 52 retrieved from GENBANK (Table S5). This matrix was used to determine the fundamental number of haplotypes and their distribution using SPLITSTREE

(ver4.8; [9]) and TCS (ver1.21; [10]). Before proceeding with further phylogenetic analysis, we analyzed the theoretical completeness of haplotype sampling (See Supplementary Discussion and Figure S2).

After trimming the original matrix so that it consisted only of individual haplotypes (i.e. 80 plus 5 outgroup sequences), that dataset was subjected to parsimony and bayesian analyses using MRBAYES (ver3.1; [11]) to reconstruct the reticulated tree.

A third dataset was constituted by pruning the original matrix to the actual number of 'dated' specimens (138 mammoths + 6 modern sequences). Of the mammoths, 131 had individual ^{14}C dates and 7 others stem from the dated Berelekh archeological site in Eastern Siberia (Table S1). This site has been documented to be a strictly post-LGM site with most evidence [12] dating it between 12 and 14 ky BP (uncalibrated ^{14}C dates). We dated three specimens from that site and our dates span $12125 \pm 30 - 12350 \pm 35$ ^{14}C ky BP, thus calibrated between 14 and 15 ka. We arbitrarily applied the mean of that time span (i.e. 14.5 ka) as an individual date for the 7 other specimens from the Berelekh site and used them together with the ^{14}C dated specimens in BEAST (ver1.4.6; [13]) for inferences on the timing of split events and demographic history. Infinite dates were used with the minimum age selected as a proxy for the actual age of the samples (see below).

Reconstruction of the Rooted Reticulated Tree (RRT)

MRBAYES was used to infer the bayesian tree describing best our dataset using an HKY+I+G model, as selected by MODELTEST (ver3.7; [14]) and the both hLRT and AIC criteria. The Bayesian analysis was run using default options except for parameters stipulated as follows: nst = 2, rates = invgamma, ngen = 10,000,000, and with a burnin period of 1,000,000 and sampling every 1,000 generations for both tree and model parameters. Values for models generated by distribution of posteriors in bayesian analysis converged evenly (potential scale reduction factor = 1.00 in all cases) and were as follows: $\kappa = 84.88$ (95% CI [Credibility Interval] = 41.90–166.38), $\text{freq}(A) = 0.32$ (95% CI = 0.29–0.36), $\text{freq}(C) = 0.24$ (95% CI = 0.21–0.27), $\text{freq}(G) = 0.13$ (95% CI = 0.10–0.15), $\text{freq}(T) = 0.31$ (95% CI = 0.28–0.34), $\text{Pinvar} = 0.69$ (95% CI = 0.64–0.73), $\alpha = 0.18$ (95% CI = 0.17–0.19).

The full topology of the tree derived from the bayesian analysis is shown in Figure S3 in which all compatible nodes have been retained with their individual posterior probabilities. The five haplogroups presented in Figure 2 of the main text are all supported by a posterior probability ≥ 0.75 and are highlighted with the same color scheme (see

Supplementary Discussion).

Phylogenetic analyses performed at the intraspecific level are prone to topology errors using outgroup rooting, due either to the presence of polymorphisms shared with the outgroup (the ingroup thus being non-monophyletic), or conversely to extreme outgroup divergence [15, 16]. However, running the bayesian analysis with or without modern elephant sequences did not affect the topology or support of any of the nodes discussed in the text; the only changes were at nodes with low support (below 0.50) within each main haplogroup (data not shown).

For the interpretation of within-species variation, "network" methods are thought to provide a more realistic representation of among-group relationships than the bifurcating trees of conventional phylogenetic analyses [16]. However, Cassens et al. [15] have noted inherent biases in the available algorithms for so-called "network" reconstructions of intraspecific phylogenies (e.g., minimum-spanning network approach implemented in ARLEQUIN and statistical parsimony implemented in TCS), especially in regard to internal haplotypes that have not been sampled. Cassens et al. [17] have thus produced a new method (i.e. UMP for Union of Maximum Parsimony trees) which, in their view, is less biased relying on conventional parsimony reconstruction to derive a consensus graph that summarizes all combinable parsimonious

trees into a single, potentially reticulated tree (here called RRT).

We applied the UMP method to our matrix of 80 haplotypes, rooted with the six modern elephants in our dataset. Parsimony analysis (step 1 of the UMP) was performed with TNT (ver1.1, downloadable at <http://www.zmuc.dk/public/phylogeny/TNT/>) using the New Tech algorithms [18] with default options implemented, except that we added 10,000 random sequences with both sectorial search (SECT) and tree-branch-reconnection (TBR) rearrangements implemented in Driven Search until the same stable consensus was reached twice. Performing two independent runs of the analysis yielded the same 284 parsimonious trees of 234 steps (Consistency Index = 0.51, Retention Index = 0.80), with 23 nodes in the strict consensus. All trees were saved with their branch lengths and processed with TRECOMBINE [17] to generate the best reticulated tree (step 2 of the UMP), displayed as Figure 1 of the main text.

In our specific case, the RRT obtained is similar in pattern to other network reconstructions using either statistical parsimony (TCS ver1.21, [10]) or median-joining network (NETWORK ver4.2.0.1, [19]), apart from minor changes in the assignment of tip groups within each of the five haplogroups which are not discussed here (data not shown).

Substitution rates analyses

The software BEAST [13] was used to derive the most accurate substitution rate for the present dataset. We first analyzed 15 published full mitochondrial genomes within Elephantoidea: 1 American mastodon, 2 savannah African elephants, 1 Asian elephant and 11 dated (out of 13) mammoths. The tandem repeats found in the control regions were excluded, and the rest of the genome (16469 bp total) was analyzed as a single partition under the GTR+I+G model. The effect of the fragment size on the rate calculation was addressed by comparing the posterior estimates derived from the analysis of the complete genome (referred to as "full mt") versus those of the analysis of the 705 bp fragment sequenced for the mammoth dataset. A more comprehensive analysis of the effect of fragment size is performed elsewhere (Debruyne et al. in prep.).

Two different calibration approaches were used and contrasted. Full bayesian analyses (see Supplemental Discussion) were performed using only tip calibrations of the samples (with an age of 90 ky BP selected for the mastodon specimen based on fossil evidence, [20]), whereas another set of constrained analyses combined the same tip calibrations together with a paleontological calibration of the node separating the lineage of the American mastodon from the Elephantidae estimated at

24–28 Ma [21]. Those Analyses were run according to a strict clock model using 10,000,000 generations. Sampling for all parameters occurred every 1,000 generations with a burnin of 1,000,000 and TRACER (ver1.4; <http://beast.bio.ed.ac.uk/Tracer>) was used to check for evidence of convergence and adequate sampling (all parameters ESS > 100). In order to minimize the number of nuisance parameters when determining the evolutionary rate, analyses were conducted using a constant-size demographic model.

Dating divergence events in the RRT

The splitting events within the mammoth dataset were dated using the 138 dated mammoths and 6 modern elephant sequences of 705 bp. Based on the observations above, we performed a BEAST analysis using both tip calibrations of the individual specimens and two node calibrations derived from the analysis of the full mt calibrated with the mastodon/Elephantidae split (Supplementary Discussion). Those two calibrations are based on undisputed nodes (with posterior probabilities $pp = 1.0$) in both the full mt and the 705 bp analyses and describe the separation of *Elephas* and *Loxodonta* set as a prior following a normal distribution with mean = 7.699 Ma (and standard deviation of 500 ka, so that 95% CI = 6.674–8.766), and the separation of the eleven

mammoths sequences used in both analyses from the elephants (mean of normal distribution = 308.3 ka, with 95% CI = 238.5–379.5).

A comparison of both strict and relaxed clock model performances for the analysis of the 705 bp revealed that the relaxed clock framework was more effective: the comparisons of Bayes factor between the two clock environments showed that the relaxed clock was significantly better with $\log_{10} \text{BF} > 5$. Hence we choose the relaxed clock model for all subsequent analyses, which causes the relatively wide confidence intervals obtained (see Fig. 3 in the main text).

To increase both reliability and accuracy of our dates, the final dating analysis was based on the combination (using TRACER) of three runs of 10,000,000 generations each (with sampling every 1,000 generations and burnin period of 1,000,000) under the HKY+I+G model selected by MODELTEST. In order to minimize the number of nuisance parameters when determining the evolutionary rate, that analysis was performed under a constant-size demographic model.

Bayesian Skyline analysis

Those dates were confirmed by an independent Bayesian skyline analysis using only the 138 dated mammoths without outgroup consideration. In order to prevent from sampling artefactual accelerated

rates, the mean rate of that analysis was enforced using a normal prior based on the full Bayesian analysis of the full mt genome: mean rate of 9.1×10^{-9} with a standard deviation of 5.4×10^{-9} (see Supplemental Discussion). The coalescent model used in that analysis (linear Bayesian skyline) was used to derive the demography of the mammoth populations. Again, three independent analysis of 10,000,000 generations were pooled to generate the skyline.

A separate analysis using the constant population demographic prior was generated in order to perform Bayes factor comparisons. This comparison shows that, despite apparent variations in the effective population size derived from the Bayesian skyline analysis, this model does not perform better than the actual hypothesis of constant population to explain the data ($\log_{10} \text{BF cte/skyline} < 1$).

SUPPLEMENTAL DISCUSSION

Regarding the completeness of haplotype sampling

Prior to any analysis of the genetic diversity of a species at the population level it is worthwhile to test whether the dataset is in fact drawn from a truly representative sample of the populations that comprise the species in question. This issue is especially complicated for species such as the woolly mammoth that have no living members.

Dixon's [22] web-based tool

(<http://www.botanik.univie.ac.at/plantchorology/haplo.htm>) performs a bayesian analysis to measure the completeness of a given haplotype sampling effort by producing an estimate of what the Stirling probability of haplotype richness should be within a 95% confidence interval, provided assumptions of (i) panmixia and (ii) even geographic sampling are met. When applied to a recent phylogeographic analysis of late Quaternary *Mammuthus* [4] based on the recovery of 30 haplotypes from 44 different mammoth specimens, the Dixon statistic yields a mean total diversity of 52 haplotypes, but with a very wide confidence interval (95% CI = 40–93; Figure S2). Given that the mean estimate of diversity is accurate within these limits, the Barnes et al. study should have sampled 58% of the total diversity. Our recovery of 80 haplotypes indicates that the mean estimate substantially underestimates the actual

number of haplotypes present in mammoth populations during the last part of the Quaternary, was mainly due to a limited sample ($n=3$) of American mammoths. When applied to the present dataset, the Dixon statistic estimates an expected mean diversity of as much as 100 haplotypes (95% CI = 90–116; Figure S2) with much narrow confidence intervals. If correct, this estimate would imply that we have sampled between 70% and 89% of the theoretical expected haplotype richness of Late Pleistocene mammoths.

For several reasons our results need to be treated with caution. First, the mammoths sampled stem from very different geological ages and therefore can hardly be treated as a single panmictic unit. However, generally, failure to meet the assumption of panmixia should result in an overestimate of diversity under a bayesian model, as the total number of haplotypes in our dataset is probably higher than would be expected in a single contemporaneous population. Additionally, the fact that the main haplogroups in the study display a fairly different geographic distribution (see Figure 3 in main text) defies the assumption of panmixia in the model. Again, a deviation from panmixia to more structured populations would produce an overestimate of the total diversity, even if those populations are evenly sampled, due to isolation-by-distance [22]. Finally, our study has its own lack of representation, in

this case of southern North American and western European mammoths, besides a single sequence from Estonia retrieved by Barnes et al. (Table S5). This sequence haplotype supports the retention, until radiocarbon times, of mammoth populations containing possible *M. trogontherii* (or primitive *M. primigenius*) diversity consistent with the paleontological hypothesis of a late replacement of those forms by the *primigenius* form in Europe. Thus, one might expect a much greater diversity of haplogroup B (and potentially A) than presently seen due to our limited sampling of that range. Hence, this third limitation, contrarily to the two former ones would lead to an underestimate of the total diversity in mammoths due to incomplete geographic sampling.

Haplotype grouping and network/tree description

Not all the five haplogroups considered in our study are actual clades, a situation not surprising considering that we are working at the intraspecific level. They are thus better regarded as Evolutionarily Significant Units (ESUs). In figure S3 it can be seen that haplogroups C and D are paraphyletic, embedded within monophyletic supergroups (C+D+E and D+E, respectively). Although there is strong support for the five main haplogroups, within-ESUs nodal support is otherwise weak, due to overarching similarity among grouped haplotypes (often

differing by only 1 position) and retention of ancestral haplotypes (e.g., C1, D1 or E1) within each haplogroup.

Interestingly, the monophyly of the haplogroup C (only documented in North-American populations) was never sampled in the posteriors (hence its posterior probability of 0.00), so that the most likely biogeographical explanation of the pattern of distribution of the haplotypes within the C+D+E clade is that its origin be found in North America. One could argue that the absence of C haplotypes in Eurasia is due to either (i) a complete loss of that haplogroup in the course of evolution of the Eurasian populations due to a selective advantage of the other haplogroups (namely D+E) or (ii) an insufficient sampling in the present dataset. Both those hypotheses are highly unlikely in the light of the distribution of all Eurasian haplotypes: (i) the local extinction of the C haplogroup in Eurasia earlier than any of the samples assayed conflicts with the reciprocal retention of the haplogroups A and B in Eurasia until radiocarbon times together with the potentially selectively advantaged D+E populations, and the observation that even low frequency haplogroups (namely B which has a frequency not higher than 1% in the Eurasian sequences) were sampled in our dataset.

On a side note, although the phylogenetic pattern we retain is more inclusive, one might want to compare our clustering system to the

one put forward by Barnes et al. [4]. Thus our D+E clade is synonym to their clade 1, whereas our clade A is a putative synonym to clade 2.

Selection of the relevant evolutionary rate/calibration method using BEAST

In our analyses we have adopted a very conservative approach to derive a consistent substitution rate for our mammoth dataset, to address the legitimacy of the rate acceleration hypothesis, recently proposed by Ho et al.[23], but independently questioned by Emerson[24] and Bandelt[25]. Regarding the selection of priors, Ho et al.[26] have advocated for a full bayesian approach, where all parameter analyses are optimized during the analysis of sequences calibrated by their associated ^{14}C ages only (referred to as tip calibrations). Emerson[24] and Bandelt[25] on the other hand have insisted on greater control over the priors and have emphasized the importance of background knowledge associated with the dataset being analyzed, such as paleontological calibration dates (i.e. node calibrations). We implemented both approaches in BEAST and analyzed the effect of the fragment size on the rate calculation.

Figure S5 summarizes the results of the bayesian analyses of the substitution rate for either 705 bp or the complete mitochondrial

genome. In the full bayesian framework, it depicts a strong decrease in the posterior mean of the substitution rate when the fragment length increases, associated with an extreme narrowing of the 95% highest posterior density (HPD). The mean rate estimate for the complete genome is strikingly close to the former phylogenetic rate published by Rohland et al. ($4.2e-9$ substitution / site / year; [20]), whereas the rate based on the 705 bp fragment yields an average rate ~ 15 times higher ($6.7e-8$).

No obvious biological mechanism can be advocated for explaining that result: only random sets of different fragment lengths from the same data matrix are analyzed and should *a priori* yield comparable posterior mean rate estimates. This is not the case as the primary effect of using longer sequences on the estimate of the substitution rate is a decrease of the posterior mean and a narrowing of the 95% HPD by limiting the posterior sampling of elevated rates compared with the analysis of short sequences. It is thus apparent that the estimate of the mean posterior substitution rate is affected by the amount of information contained within the data matrix. However, when using the posterior rate to derive the ages of internal nodes, the mean age of the split between the mastodons and the elephants would vary from 2.4 Ma (95% HPD: 0.1–6.2 Ma) using 705 bp, to 28.3 Ma (95% HPD: 3.2–57.3

Ma) for the entire mt sequence. While the estimate produced with the full mt suffers a wide confidence interval, its posterior mean appears close to the paleontological divergence estimate between the elephant and the mastodon: between 24 and 28 Ma [21], and supports the full bayesian approach when sequences of sufficient length are studied. Conversely we show here that, when only short sequences are available for large phylogeographic studies, typical of aDNA [8, 30], a full bayesian approach is likely to yield erroneous posterior mean estimates and their associated confidence intervals.

Limited phylogenetic content of short sequences appears to relax the constraint over the substitution rates which can vary so greatly that their mean estimate becomes irrelevant for use and leads to reproducible variation trends of the rates over the entire fragment analyzed. Whether this computational limit is of general significance concerning the time-dependency hypothesis proposed by Ho et al.[23] on ancient DNA datasets is explored elsewhere (Debruyne and Poinar submitted). However we clearly show here that a full bayesian analysis of the mammoth dataset (i.e. tip calibration only) would lead to erroneous posterior estimates of the substitution rate and divergence dates when the phylogenetic content of the dataset explored is limited to 705 bp.

The effect of node calibration on the substitution rate was addressed by considering a single node calibration based on a strongly documented paleontological record that sets the divergence of the lineages of the Elephantidae vs. *Mammut americanum* between 24 and 28 Ma. When that node calibration was added to the priors of the same set of analyses of the different sizes (as a uniform prior on the root height), the posterior estimates of the rates were strikingly different: for both the fragment sizes addressed, a narrow and almost constant 95% HPD that encompass the 4.2×10^{-9} substitution/site/year estimate previously recovered [20] was obtained; very little variation was observed for the mean rate estimate: from 3.0×10^{-9} up to 3.8×10^{-9} (Fig. S5).

This effect is also retrieved for the divergence estimates of the uncalibrated nodes and enables accurate dating of those splitting events. Based on the full mt analysis, the differentiation of Elephantinae (namely here the node that separates *Loxodonta* from the *Elephas+Mammuthus* lineages, returns a posterior estimate of 7.7 Ma (95% HPD: 6.7–8.8 Ma) which accords well with the paleontological evidence: the differentiation of Elephantidae (Elephantinae+*Stegotetrabelodon*) is expected to have taken place not much before the earliest *Stegotetrabelodon* acknowledged, ca 7.5–9

Ma[27], and cannot be younger than the earliest evidence of *Loxodonta* and *Elephas* fossils (5.4–7.3 Ma and 5.2–6.7 Ma respectively[28]).

The obvious benefit (in normalization of the rate and of its HPD over the fragment length considered) of the node calibration + tip calibration over the tip calibration alone in the present study using 705 bp only for our full mammoth dataset (138 ¹⁴C dated specimens) led us to use both the separation of *Elephas* and *Loxodonta* (normal prior with mean at 7.699 Ma, 95% CI = 6.674–8.766 Ma) and the divergence of the 11 complete mitochondrial sequences (normal prior with mean at 308.3 ka, 95% CI = 238.5–379.5 ka) as nodal priors for the BEAST analysis (Supplemental Experimental procedures). The results of that analysis were used to derive the ages of nodes based on their median and 95% HPD (Fig. 3 in the main text).

SUPPLEMENTAL REFERENCES

1. Poinar, H.N., Schwarz, C., Qi, J., Shapiro, B., Macphee, R.D., Buigues, B., Tikhonov, A., Huson, D.H., Tomsho, L.P., Auch, A., Rampp, M., Miller, W., and Schuster, S.C. (2006). Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* 311, 392-394.
2. Fleischer, R.C., Perry, E.A., Muralidharan, K., Stevens, E.E., and Wemmer, C.M. (2001). Phylogeography of the asian elephant (*Elephas maximus*) based on mitochondrial DNA. *Evolution Int J Org Evolution* 55, 1882-1892.
3. Debruyne, R. (2005). A case study of apparent controversy between molecular phylogenies: the interrelationships of African elephants. *Cladistics* 21, 31-50.
4. Barnes, I., Shapiro, B., Lister, A., Kuznetsova, T., Sher, A., Guthrie, D., and Thomas, M.G. (2007). Genetic structure and extinction of the woolly mammoth, *Mammuthus primigenius*. *Curr Biol* 17, 1072-1075.
5. Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41, 95-98.
6. Weninger, B., and Jöris, O. (in press). Towards an Absolute Chronology at the Middle to Upper Palaeolithic Transition in Western Eurasia: A New GreenlandHulu Time-scale Based on U/Th Ages. *Journal of Human Evolution*.
7. Guthrie, R.D. (2004). Radiocarbon evidence of mid-Holocene mammoths stranded on an Alaskan Bering Sea island. *Nature* 429, 746-748.
8. Mol, D., TIKHONOV, A., VAN DER PLICHT, J., KAHLKE, R.-D., DEBRUYNE, R., VAN GEEL, B., VAN REENEN, G., PALS, J.P., DE MARLIAVE, C., and REUMER, J.W.F. (2005). Results of the CERPOLEX/Mammuthus expeditions on the Taimyr Peninsula, Arctic Siberia, Russian Federation. *Quaternary International* 142-43, 186-202.
9. Huson, D.H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23, 254-267.
10. Clement, M., Posada, D., and Crandall, K.A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9, 1657-1660.

11. Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17, 754-755.
12. Vereshchagin, N.K. (1977). The Berelekh mammoth "cemetery". In *Mammoth fauna of the Russian plain and Eastern Siberia*, Y.I. Starobogatov, ed. (Leningrad, USSR: Zoological Institute Press), pp. 5-50.
13. Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7, 214.
14. Posada, D., and Crandall, K.A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817-818.
15. Cassens, I., Van Waerebeek, K., Best, P.B., Crespo, E.A., Reyes, J., and Milinkovitch, M.C. (2003). The phylogeography of dusky dolphins (*Lagenorhynchus obscurus*): a critical examination of network methods and rooting procedures. *Molecular Ecology* 12, 1781-1792.
16. Posada, D., and Crandall, K.A. (2001). Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution* 16, 38-45.
17. Cassens, I., Mardulyn, P., and Milinkovitch, M.C. (2005). Evaluating Intraspecific "Network" Construction Methods Using Simulated Sequence Data: Do Existing Algorithms Outperform the Global Maximum Parsimony Approach? *Systematic Biology* 54, 363-372.
18. Goloboff, P.A. (1999). Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics* 15, 415-428.
19. Bandelt, H.-J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16, 37-48.
20. Rohland, N., Malaspinas, A.S., Pollack, J.L., Slatkin, M., Matheus, P., and Hofreiter, M. (2007). Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biol* 5, e207.
21. Shoshani, J., Walter, R.C., Abraha, M., Berhe, S., Tassy, P., Sanders, W.J., Marchant, G.H., Libsekal, Y., Ghirmai, T., and Zinner, D. (2006). A proboscidean from the late Oligocene of Eritrea, a "missing link" between early Elephantiformes and Elephantimorpha, and biogeographic implications. *Proc Natl Acad Sci U S A* 103, 17296-17301.
22. Dixon, C.J. (2006). A means of estimating the completeness of haplotype sampling using the Stirling probability distribution. *Molecular Ecology Notes* 6, 650-652.

23. Ho, S.Y., Phillips, M.J., Cooper, A., and Drummond, A.J. (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol* 22, 1561-1568.
24. Emerson, B.C. (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.
25. Bandelt, H.J. (2007). Clock debate: when times are a-changin': Time dependency of molecular rate estimates: tempest in a teacup. *Heredity*.
26. Ho, S.Y., Shapiro, B., Phillips, M.J., Cooper, A., and Drummond, A.J. (2007). Evidence for time dependency of molecular rate estimates. *Syst Biol* 56, 515-522.
27. Tassy, P. (2003). Elephantoidea from Lothagam. In *Lothagam: The Dawn of Humanity in Eastern Africa*, M.G. Leakey and J.M. Harris, eds. (New York: Columbia University Press), pp. 331-358.
28. Shoshani, J., and Tassy, P. (2005). Advances in proboscidean taxonomy & classification, anatomy & physiology, and ecology & behavior. *Quaternary International* 126-128, 5-20.
29. Macphee, R.D.E., Tikhonov, A., Mol, D., de Marliave, C., van der Plicht, H., Greenwood, A.D., Flemming, C., and Agenbroad, L. (2002). Radiocarbon chronologies and extinction dynamics of the Late Quaternary Mammalian Megafauna of the Taimyr Peninsula, Russian Federation. *Journal of Archaeological Science* 29, 1017-1042.

SUPPLEMENTAL FIGURES

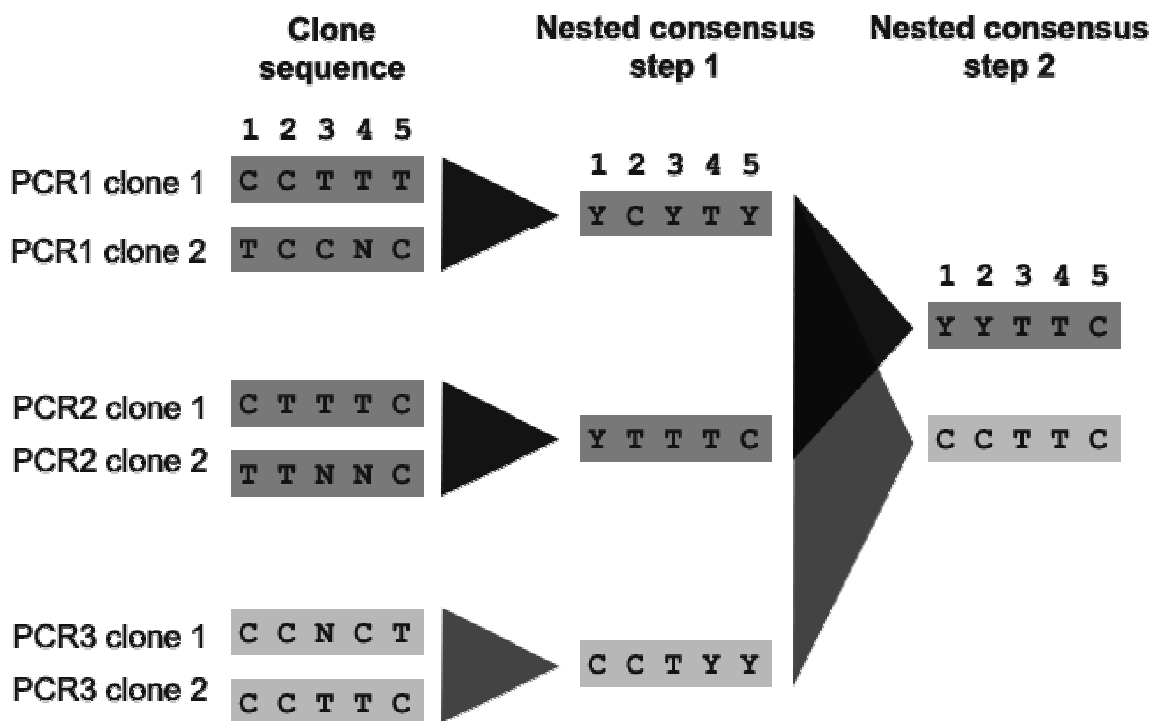


Figure S1: Nested consensus approach: when the first two PCRs (1 and 2) were insufficient to obtain a fully resolved step 2 consensus, additional PCRs (here PCR 3) were added until all ambiguous positions had been resolved.

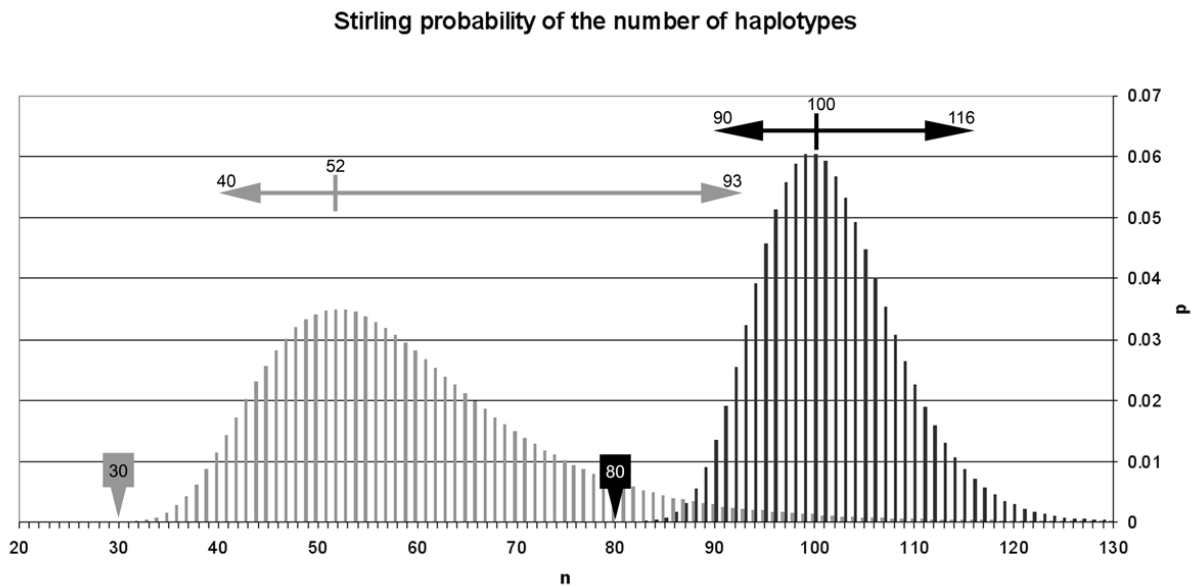


Figure S2: Stirling probability estimates (Y axis) of haplotype richness (X axis of the graphs), with means and 95% confidence intervals (horizontal bars): in gray, Barnes et al's dataset; in black, this paper's dataset.

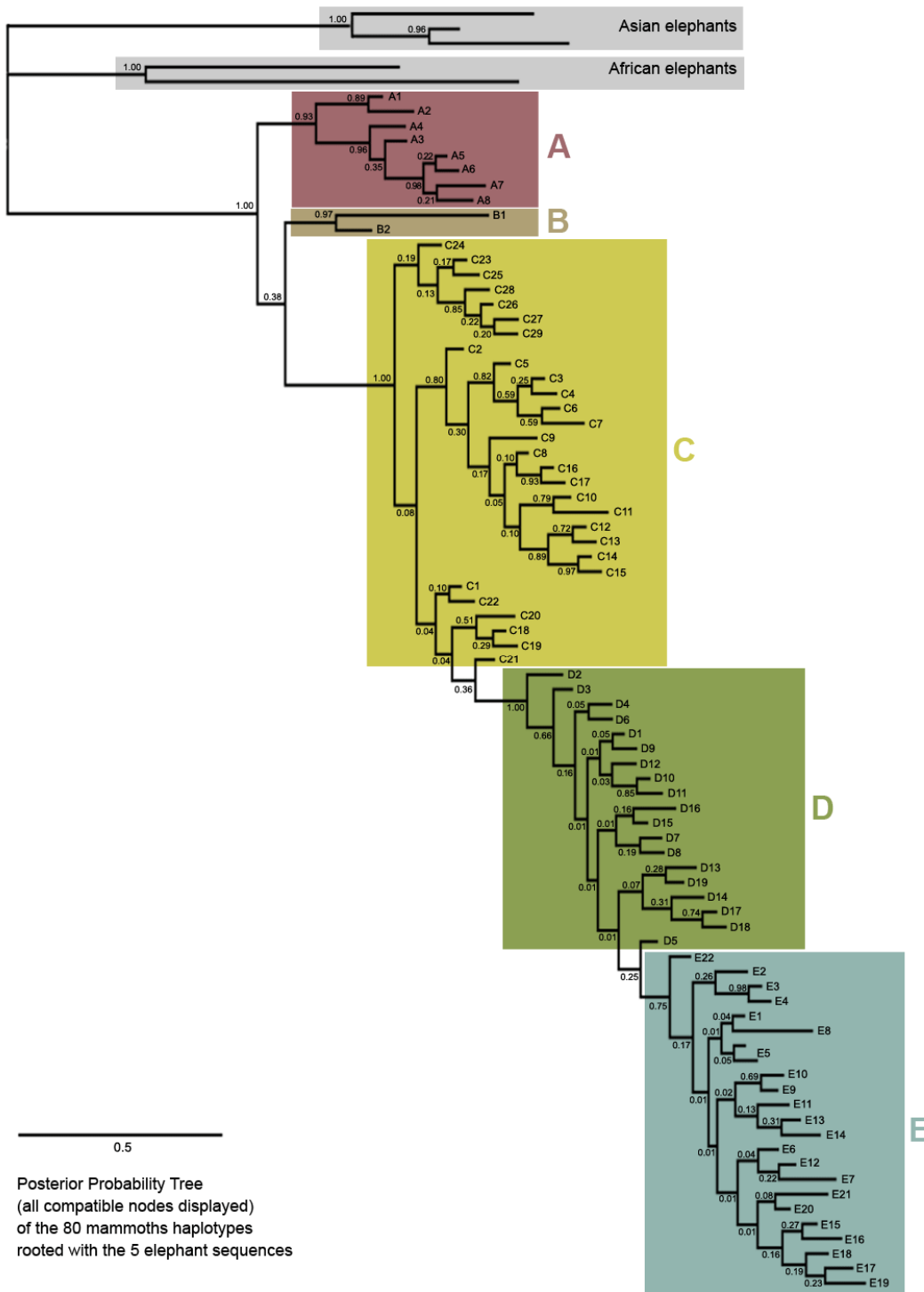


Figure S3: MrBayes phylogram of 80 mammoth haplotypes (posterior probabilities indicated at nodes).

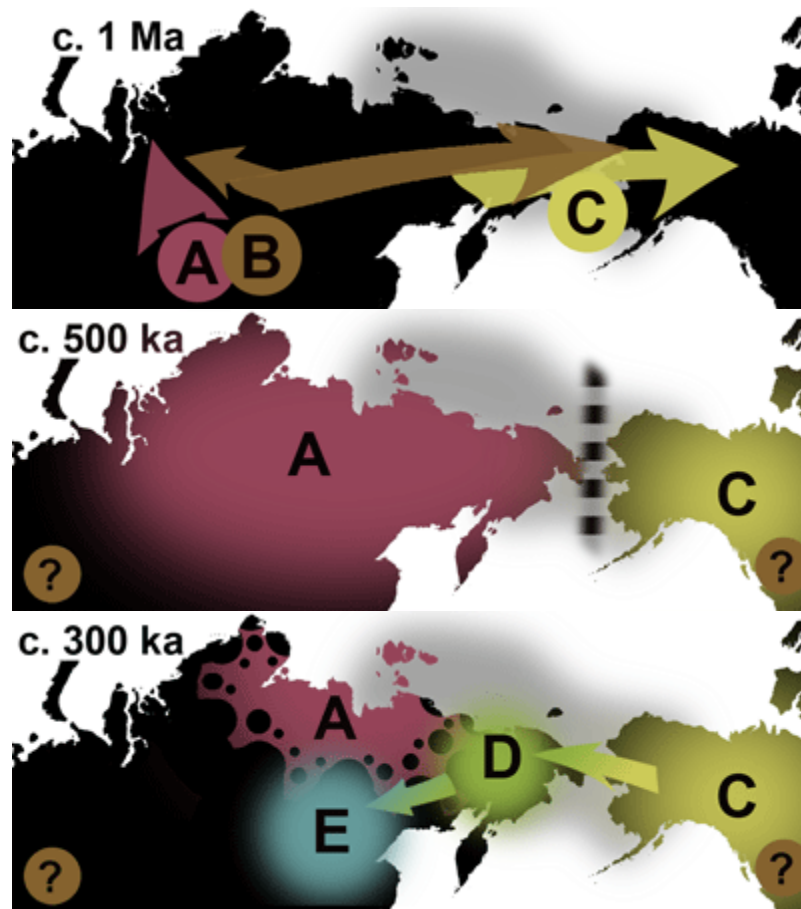


Figure S4: The putative origin, divergence and migration of *Mammuthus* populations during Early/Middle Pleistocene, based upon the dates of divergence estimated from our data (Fig. 4 and main text).

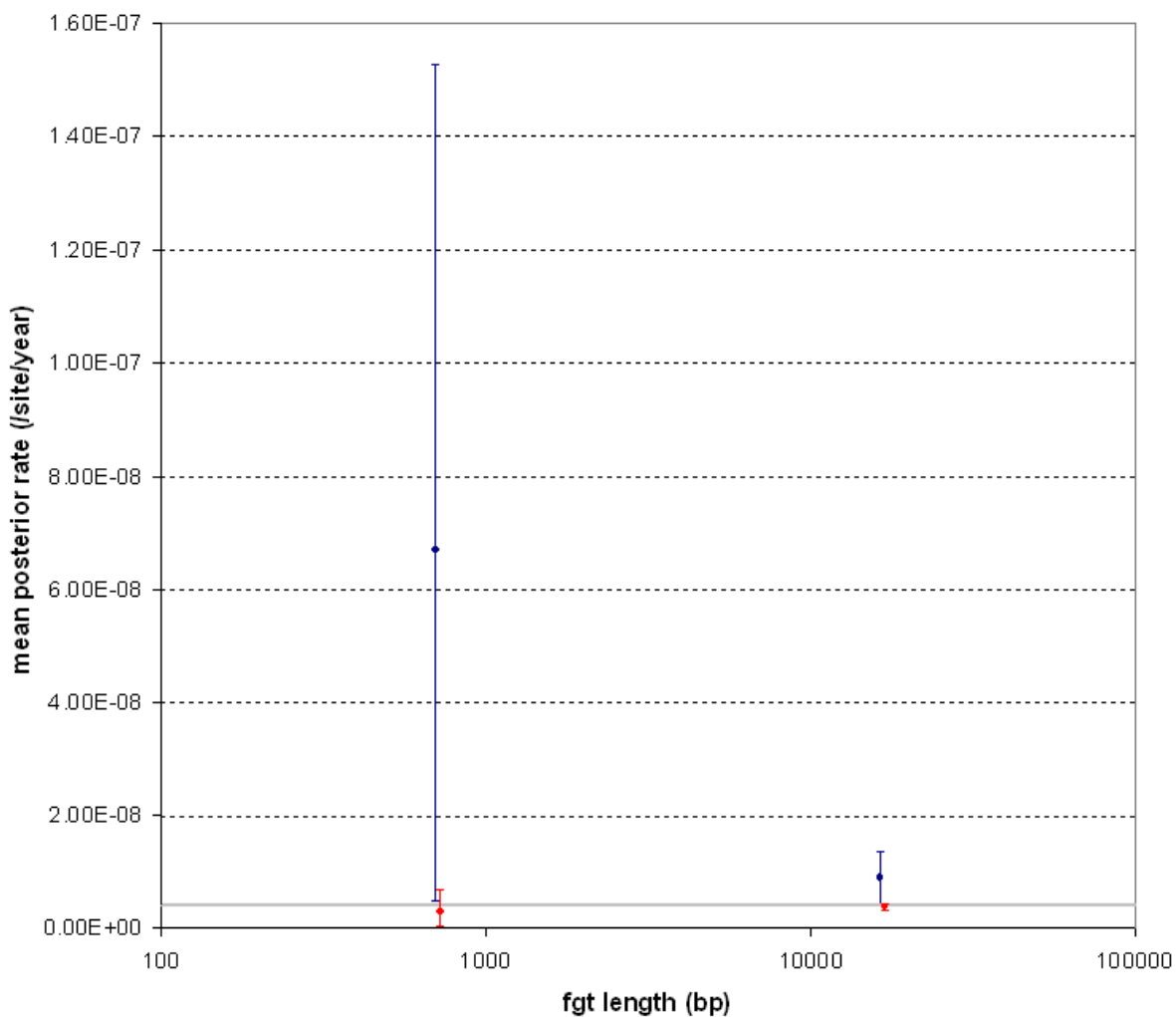


Figure S5: Analysis of posterior substitution rates (mean and 95% HPD) according to the fragment length analyzed (705 bp or complete genome) in a full bayesian framework (in blue), or using an extra node calibration (in red). The grey line is set at the constant formerly published phylogenetic rate for the full mt genome of Elephantoidea: 4.2×10^{-9} substitutions/site/year [20].

SUPPLEMENTAL TABLES

Table S1: Sample statistics. Samples in grey did not provide the full 743 bp sequence or were not attempted in PCR at all due to low copy numbers in *cytochrome b* quantitation. Column headings-- *Lab ID*: extraction number (two IDs listed when two different extractions were performed); *Collection*: institutional acronyms as follows: Zoological Institute of Saint-Petersburg (*ZIN*), Cerporex/Mammuthus collections, Khatanga (*Cerporex*), Matheus personal collection (*MPC*), National Museum Of Natural History, USA (*UNM*), United States Bureau of Land Management (*USBLM*), American Museum of Natural History (*AMNH*), National Museum of Canada (*NMC*), and Yukon Paleontological Collections (*YPC*); *Coll #*: accession/catalog number; *B/T*: sample source (bone/tooth); *Temp*: current preservation conditions of sample; *Taxon, Area, Locality, latitude, longitude*: as given by collector/institution; *14C date*: in radiocarbon years before present; *Sigma*: \pm standard error; *Calendar age, Calendar sigma*: same values calibrated in calendar years (see supplementary text); *Ref.*: source of date information (1, MacPhee et al.[29]; 2, Poinar et al.[1]; 3, Matheus (unpublished material); 4, Guthrie [7]; and N, this paper); *Dating Lab ID*: institutional dating reference with acronyms as following: University of California (*UCIAMS*), Arizona Accelerator (*AA*), Beta Analytic (*Beta*),

Radiocarbon Laboratory of Gröningen (*GrA*), Lawrence Livermore National Laboratory's Center for Accelerator Mass Spectrometry (*CAMS*); *qPCR raw*, *qPCR 1:10 dilution*: yield per reaction for undiluted extract or diluted 1:10 in 0.1X TE; *Fragments sequenced*: number of fragments sequenced to complete the 743 bp sequence following the primer number in Table S2, with positive (1 or more) clones sequenced per amplification separated by a '+' (negative PCRs shown as (0)); *Haplotype*: sample haplotype assignment (see Figure 1).

Current Biology, Volume 18

90	UNM	8572	B	room	<i>M. primigenius</i>	Alaska	South East AK	58° 30'N 135° 00'W	18090	250	21778	442	4	AA26004	243	105	-	-	-	-	2+2	2+2	2+2	2+2	3+2	E14	
89	USBLM	50AEM47	B	room	<i>M. primigenius</i>	Alaska	Bristol Bay	57° 30'N 159° 00'W	28440	850	32891	765	4	AA25997	0	0	-	-	-	-	-	-	-	-	-	nd	
59	AMNH	AM1001A	B	room	<i>M. primigenius</i>	Alaska	MudCr. Seward	58° 30'N 159° 30'W	42562	1795	46421	1966	4	AA14901	4	0	-	-	-	-	-	-	-	-	-	nd	
65	UNM	8846	B	room	<i>M. primigenius</i>	Alaska	Norton Sound	63° 26'N 163° 30'W	37100	2500	41119	2413	4	AA26009	0	0	-	-	-	-	-	-	-	-	-	nd	
51	NMC	6746	T	room	<i>M. primigenius</i>	Alaska	Tanana	55° 18'N 151° 54'W	23150	460	27745	662	4	AA17574	7903	1835	-	2+2	2+2	-	-	2	-	-	-	C4	
A48/76	AMNH	AM1131	T	room	<i>M. primigenius</i>	Alaska	Koyuk	67°N 156°W	30485	437	34763	449	4	AA14907	159	31	-	-	-	-	2+2	2+2	2+2	2+2	2+2	C14	
53	AMNH	AM1187	T	room	<i>M. primigenius</i>	Alaska	Inglutalik Cr.	65°N 165°W	>41081	na	>44641	na	4	AA14365	428	91	-	-	-	2+2	2+2	-	2	2+2+2	2+2	C20	
88	AMNH	AM1188	T	room	<i>M. primigenius</i>	Alaska	Inglutalik Cr.	65°N 165°W	37359	2532	41339	2422	4	AA14362	76	10	-	-	-	-	2	0	2	0	0	nd	
A36/74	AMNH	AM1189	T	room	<i>M. primigenius</i>	Alaska	Inglutalik Cr.	65°N 165°W	31360	1251	36090	1520	4	AA14363	225	26	-	-	-	-	2+2	2+2	2+2	2+2+1	3+2+1	C18	
A37/73	AMNH	AM1193	T	room	<i>M. primigenius</i>	Alaska	Inglutalik Cr.	65°N 165°W	16319	292	19566	467	4	AA14364	115	12	-	-	-	-	3+2	2+2	2+2	3+2+1	2+2+2	C6	
A41/37	AMNH	AM1208	T	room	<i>M. primigenius</i>	Alaska	Sullivan Creek	65° 10'N 151°W	12677	142	15047	367	4	AA14888	937	136	-	-	-	-	2+2	2+2	2+2	2+2	2+2	C7	
A45	AMNH	AM2378	T	room	<i>M. primigenius</i>	Alaska	Cripple Creek	64° 60'N 148°W	>49894	na	>54251	na	4	AA14905	0	0	-	-	-	-	-	-	-	-	-	nd	
85	AMNH	AM2442	T	room	<i>M. primigenius</i>	Alaska	Cripple Creek	64° 60'N 148°W	>41094	na	>44655	na	4	AA14858	0	0	-	-	-	-	-	-	-	-	-	nd	
A40/75	AMNH	AM2446	T	room	<i>M. primigenius</i>	Alaska	Cripple Creek	64° 60'N 148°W	26022	640	30889	585	4	AA14855	140	16	-	-	-	-	2+2	2+2	2+2	2+2	2+2	C2	
63	AMNH	AM2527	T	room	<i>M. primigenius</i>	Alaska	Cripple Creek	64° 60'N 148°W	>41086	na	>44646	na	4	AA14857	0	0	-	-	-	-	-	-	-	-	-	nd	
70	NMC	11349	T	room	<i>M. primigenius</i>	Alaska	Cripple Hill	64° 60'N 148°W	37800	2700	41672	2525	4	AA17545	6001	931	-	0	3+2+2	2+2	2+2	-	-	-	-	C22	
A47	AMNH	AM1104	T	room	<i>M. primigenius</i>	Alaska	Clary Creek	65° 10'N 147° 30'W	42764	1737	46578	1947	4	AA14906	163000	20100	-	-	2+2+2	2+2	-	-	2	2	2	C28	
A44	AMNH	AM502	T	room	<i>M. primigenius</i>	Alaska	Clary Creek	65° 10'N 147° 30'W	40415	1341	44181	1145	4	AA14911	4	0	-	-	-	-	-	-	-	-	-	nd	
87	AMNH	AM504	T	room	<i>M. primigenius</i>	Alaska	Clary Creek	65° 10'N 147° 30'W	14093	163	17354	272	4	AA14892	5	0	-	-	-	-	-	-	-	-	-	nd	
A43	AMNH	AM523	T	room	<i>M. primigenius</i>	Alaska	Clary Creek	65° 10'N 147° 30'W	43239	1878	47030	2219	4	AA14904	972	136	-	0+0	0+0	-	2+2	2+2	2+2	2+2+2	2+2+1	C28	
62	AMNH	AM8052	T	room	<i>M. primigenius</i>	Alaska	Clary Creek	65° 10'N 147° 30'W	18379	124	20219	328	4	AA14935	8576	1278	-	3+2	2+2+1	-	-	-	-	-	-	C1	
A46	AMNH	AM529	T	room	<i>M. primigenius</i>	Alaska	Goldstream	64° 50'N 148° 00'W	43373	1875	47136	2222	4	AA14902	0	0	-	-	-	-	-	-	-	-	-	nd	
A42	AMNH	AM6677	T	room	<i>M. primigenius</i>	Alaska	Goldstream	64° 50'N 148° 00'W	12576	147	114902	365	4	AA14880	6	0	-	-	-	-	-	-	-	-	-	nd	
64	AMNH	AM1114	T	room	<i>M. primigenius</i>	Alaska	Esther Creek	64° 50'N 148° 00'W	46391	2699	50593	3611	4	AA14939	115	27	-	-	-	-	0+0	-	-	0+0	0+0	nd	
A39	AMNH	AM8744	T	room	<i>M. primigenius</i>	Alaska	Esther Creek	64° 50'N 148° 00'W	16789	108	20007	300	4	AA14896	224	7666	-	0	-	-	3+3	2+2	2+2	3+3+3	2+2	D19	
63	AMNH	AM140	T	room	<i>M. primigenius</i>	Alaska	Fairbanks Creek	64° 50'N 148° 00'W	39805	1225	43732	962	4	AA14927	62	31	-	-	-	-	0	0	2+2	0	0	nd	
62	AMNH	AM5025	T	room	<i>M. primigenius</i>	Alaska	Fairbanks Creek	64° 50'N 148° 00'W	39151	3232	43058	3042	4	AA14856	0	0	-	-	-	-	-	-	-	-	-	nd	
57	USBLM	AK-2036-5	T	room	<i>M. primigenius</i>	Alaska	Fairbanks area	64° 50'N 147° 40'W	39700	3400	43846	3132	4	AA22572	4	0	-	-	-	-	-	-	-	-	-	nd	
55	NMC	42135	T	room	<i>M. primigenius</i>	Alaska	Eldorado Cr.	64° 50'N 147° 40'W	30000	1000	34276	985	4	AA17538	11790	1175	0+0	2+2	2+2	-	-	-	2	-	2+2	2	C8
86	AMNH	AM5332	T	room	<i>M. primigenius</i>	Alaska	Fairbanks area	64° 50'N 147° 40'W	>41007	na	>44561	na	4	AA14884	0	0	-	-	-	-	-	-	-	-	-	nd	
A49	AMNH	AM1006	T	room	<i>M. primigenius</i>	Alaska	Dome Creek	64° 00'N 141° 10'W	14372	92	17531	257	4	AA14919	24	21	-	-	-	-	2+2	2+2	2+2	2+2	2+2	D19	
6	YPC	57.0001	T	room	<i>M. primigenius</i>	Yukon	VGFN Foot	68° 11'N 140° 32'W	>47800	na	>51652	na	N	UCIAMS41490	119	30	-	-	-	-	2+2	2+2	2+2	2+2	2+2+2	C2	
8	YPC	42.0001	T	room	<i>M. primigenius</i>	Yukon	Old Crow River	67° 47'N 139° 55'W	nd	nd	nd	nd	nd	na	0	0	-	-	-	-	-	-	-	-	-	nd	
91	YPC	181.0004	B	room	<i>M. primigenius</i>	Yukon	Bar on Old Crow River	67° 45'N 139° 55'W	nd	nd	nd	nd	nd	na	0	0	-	-	-	-	-	-	-	-	-	nd	
7	YPC	173.001	T	room	<i>M. primigenius</i>	Yukon	Chijee's Bluff	67° 29'N 139° 55'W	>45400	na	>48612	na	N	UCIAMS41492	26140	4727	-	2+0	2+2	2	2	-	-	-	-	C2	
56	NMC	47732	T	room	<i>M. primigenius</i>	Yukon	Sixty Mile	66° 00'N 141° 00'W	36800	2400	40853	2331	4	AA17524	0	0	-	-	-	-	-	-	-	-	-	nd	
9	YPC	79.0001	T	room	<i>M. primigenius</i>	Yukon	Last chance Creek	63° 59'N 139° 06'W	>49800	na	>52893	na	N	UCIAMS41491	2662	614	-	2+2+2	3+2	-	-	-	-	-	-	C1	
61/95	NMC	772	T	room	<i>M. primigenius</i>	Yukon	Last Chance	64° 01'N 139° 10'W	36800	2400	40853	2331	4	AA17544	14240	1748	-	-	-	-	2+2	2+2	2+2	2+2	2+2+1	C26	
34	YPC	5.0046	B	room	<i>Mammuthus sp.</i>	Yukon	Hunker Creek	63° 59'N 139° 02'W	22430	140	27098	492	N	UCIAMS41487	3953	559	-	2+2+2	2+2	-	-	-	-	-	-	C8	
40	YPC	5.0069	B	room	<i>M. primigenius</i>	Yukon	Hunker Creek	63° 59'N 139° 02'W	>47500	na	>51282	na	N	UCIAMS39888	280	51	-	-	-	-	2+2	2+2	2+2	2+2	2+2	C2	
A51	YPC	5.0130	B	room	<i>M. primigenius</i>	Yukon	Hunker Creek	63° 59'N 139° 02'W	32470	480	36990	872	N	UCIAMS39889	422	67	-	-	-	-	2+2	2+2	2+2	2+2	2+2	C18	
1	YPC	29.0248	B	room	<i>M. primigenius</i>	Yukon	Hunker Creek	63° 59'N 139° 02'W	43500	1900	47245	2244	N	UCIAMS41493	115	38	-	-	-	-	2+2	2+2	2+2	2+2+2	2+2	C16	
49	NMC	46308	B	room	<i>M. primigenius</i>	Yukon	Hunker Creek	63° 59'N 139° 02'W	35800	2100	39916	2168	4	AA17564	1743	280	-	0+0	0+0	2+2	2+2+2	-	-	2+2	2+2	C16	
69	NMC	9926	T	room	<i>M. primigenius</i>	Yukon	Hunker Creek	63° 59'N 139° 02'W	31300	1200	36002	1454	4	AA17555	3439	443	-	0+0	0+0	2+2	2+2	2+2	2+2	2+2+2	2+2	C18	
32	YPC	2.0007	B	room	<i>M. primigenius</i>	Yukon	Hester Creek	63° 58'N 139° 03'W	27540	270	32173	287	N	UCIAMS41488	208	33	-	-	-	-	2+2	2+2	2+2	2+2	2+2	C3	
26	YPC	130.0002	B	room	<i>Mammuthus sp.</i>	Yukon	Quartz Creek	63° 49'N 139° 02'W	36690	810	41653	540	N	UCIAMS39891	597	312	-	0	2+2	2+2	2+2	-	-	-	-	D1	
A52	YPC	52.0036	B	room	<i>M. primigenius</i>	Yukon	Indian River	63° 46'N 139° 19'W	30630	870	34944	818	N	UCIAMS39890	828	145	-	2+2	2+2	-	-	-	-	-	-	C24	
29	YPC	52.0037	B	room	<i>M. primigenius</i>	Yukon	Indian River	63° 46'N 139° 19'W	nd	nd	nd	nd	nd	na	61	15	-	-	-	-	2+0	0+0	-	-	-	nd	
35	YPC	133.0018	B	room	<i>Mammuthus sp.</i>	Yukon	Whitman Gulch	63° 43'N 138° 38'W	32140	370	36537	759	N	UCIAMS38675	1401	181	-	2+2+2	2+2	-	-	-	-	-	-	D11	
11	YPC	133.0021	B	room	<i>Mammuthus sp.</i>	Yukon	Whitman Gulch	63° 43'N 138° 38'W	34180	590	39510	1007	N	UCIAMS41489	2496	510	-	2+2	2+2	-	-	-	-	-	-	C23	
A50	YPC	137.0003	B	room	<i>M. primigenius</i>	Yukon	Whitman Gulch	63° 43'N 138° 38'W	39800	1200	43724	946	N	UCIAMS39892	48	18	-	-	-	-	2+2	2+2	2+2	2+2	2+2	D1	
A53	YPC	136.0009	B	room	<i>M. primigenius</i>	Yukon	Sulphur Creek	63° 44'N 138° 50'W	45900	2600	49985	3373	N	UCIAMS39116	808	119	-	2+2	2+2+2	-	-	-	-	-	-	C26	
28	YPC	136.0005	B	room	<i>M. primigenius</i>	Yukon	Sulphur Creek	63° 44'N 138° 50'W	nd	nd	nd	nd	nd	na	354	43	-	-	-	-	2+2	3+2	2+2	2+2	2+2	C5	
92	YPC	139.0004	B	room	<i>M. primigenius</i>	Yukon	Klondike area	63° 44'N 138° 50'W	nd	nd	nd	nd	nd	na	0	0	-	-	-	-	-	-	-	-	-	nd	
39	YPC	3.0019	B	room	<i>M. primigenius</i>	Yukon	Finning	63° 50'N 138° 15'W	44700	2200	48491	2685	N	UCIAMS39887	457	48	-	2+0+0	0+0	2	2	-	-	2+2	2+2	C25	
33	YPC	3.0025	B	room	<i>M. primigenius</i>	Yukon	Finning	63° 50'N 138° 15'W	nd	nd	nd	nd	nd	na	71	25	-	-	-	-	2+2	2+2	2+2	2+2	2+2	C18	
3	YPC	3.0133	B	room	<i>M. primigenius</i>	Yukon																					

Table S2: Primer sequences and position on the mammoth mitochondrial genome (according to published full mt genomes [5, 6]).

#	F/R	5' end position	Tm	Sequence (5'-3')
1	F	14985	56.3	GCCATCCTACGATCTGTACCA
2	R	15175	52.1	GGTTGACTGCCAATTCATG
3	F	15097	55.5	TACTTCGACCTCTTAGCCAAGT
4	R	15333	52.1	TCATTTATGGCTTACAAGACCA
5	F	15237	54.7	CTAGCTTTCCTGCCAATTGC
6	R	15432	53.3	TGATTTGTTTGCAGGGAATAGTT
7	F	15383	49.4	CAATACCCAAAACCTGAAATTCT
8	R	15619	52.9	ACATGACTTGTGAATCTACACG
9	F	15560	51.1	TAAGCAAGTACTGTTTAATCAATGT
10	R	15770	53.1	TTCTCGGAGGTAGGTAGTTAAG

Table S3: Primer pair amplification characteristics and reaction conditions. Sequence size corresponds to amplicon size excluding primers.

Forward primer #	5' position	Reverse primer #	5' position	Amplicon size (bp)	Sequence size (bp)	Annealing temp.	Extension time
1	14985	10	15770	786	743	55	60"
1	14985	6	15432	448	404	55	40"
3	15097	6	15432	334	289	56	40"
7	15383	10	15770	388	344	50	40"
1	14985	2	15175	191	151	58	30"
3	15097	4	15333	237	193	58	30"
5	15237	6	15432	196	153	55	30"
7	15383	8	15619	237	193	55	30"
9	15560	10	15770	211	164	56	30"

Table S4: Relative precision of ^{14}C age estimates for 5 previously-dated mammoth specimens after ultrafiltration of “collagen” fraction.

Sample	Original date	Former Dating IDs	Publication	New date	New Dating IDs
2002/473	55800 +4500 / -2900	GrA-21232 GrA-21293 GrN-27409 GrN-27520	Mol et al. 2005	46700 +/- 2800	UCIAMS39881 UCIAMS39893
NMC8139	>35600	AA17568	Guthrie, 2004	39500 +/- 1200	UCIAMS41485
NMC49927	>37100	AA17577	Guthrie, 2004	46600 +/- 200	UCIAMS39883
NMC49928	>37800	AA17546	Guthrie, 2004	>46200	UCIAMS39884
NMC49562	>37500	AA17583	Guthrie, 2004	>47200	UCIAMS41486

Table S5: Haplotype assignment of previously published sequences. [*]

Haplotype assignment was not possible for sample BL-0179 due to uncertainties at polymorphic sites in the published sequence. Headers: see table S1; *Publication* and *Accession #* provided for each sequence as retrieved from GENBANK.

Sample name	Publication	Accession #	Area	Locality	14C date (y BP)	std. dev. (y)	Calendar (y BP)	Cal. Dev. (y)	Haplotype
BR.10	Barnes et al. 2007	AM746158	Europe	Koosa, Estonia	40900	600	44415	850	B1
AM 4136	Barnes et al. 2007	AM746161	Alaska	Gold Hill	>41027	na	>44583	na	C11
AM 8744	Barnes et al. 2007	AM746163	Alaska	Ester Creek	16789	100	20008	297	D1
AM 493	Barnes et al. 2007	AM746162	Alaska	Ban Creek	15540	145	18806	237	C1
PIN 3751-162	Barnes et al. 2007	AM746160	Eastern Siberia	Buor-Khaya peninsula	>40800	na	>44338	na	A5
PIN 3751-161	Barnes et al. 2007	AM746159	Eastern Siberia	Buor-Khaya peninsula	>41100	na	>44661	na	A8
MKh-0447	Barnes et al. 2007	AM746164	Eastern Siberia	Bykovsky peninsula	14600	100	17905	378	E6
MKh-0621	Barnes et al. 2007	AM746165	Eastern Siberia	Bykovsky peninsula	19200	200	22987	321	E10
MKh-0422	Barnes et al. 2007	AM746166	Eastern Siberia	Bykovsky peninsula	20200	100	24146	287	E10
MKh-0428	Barnes et al. 2007	AM746167	Eastern Siberia	Bykovsky peninsula	>44000	na	>47438	na	D9
MKh-0435	Barnes et al. 2007	AM746168	Eastern Siberia	Bykovsky peninsula	13100	500	15767	869	E11
MKh-0437	Barnes et al. 2007	AM746169	Eastern Siberia	Bykovsky peninsula	>44300	na	>47676	na	E19
Mkh-0433	Barnes et al. 2007	AM746170	Eastern Siberia	Bykovsky peninsula	30300	600	34591	584	E19
MKh-0438	Barnes et al. 2007	AM746171	Eastern Siberia	Bykovsky peninsula	29400	600	33662	543	E19
MKh-0429	Barnes et al. 2007	AM746172	Eastern Siberia	Bykovsky peninsula	28900	200	33412	361	D1
MKh-0533	Barnes et al. 2007	AM746173	Eastern Siberia	Bykovsky peninsula	34000	500	39431	1051	E19
MKh-0326	Barnes et al. 2007	AM746174	Eastern Siberia	Bykovsky peninsula	33800	260	39442	1058	E20
MKh-0381	Barnes et al. 2007	AM746175	Eastern Siberia	Bykovsky peninsula	24300	200	29065	453	D1
MKh-0430	Barnes et al. 2007	AM746176	Eastern Siberia	Bykovsky peninsula	32800	800	37351	1143	E18
Nag-99-O203	Barnes et al. 2007	AM746192	Eastern Siberia	Lena Delta	30200	400	34407	309	E15
M13	Gilbert et al. 2007	EU153445	Eastern Siberia	Lena Delta	35800	1200	40398	1325	E13
nd	Rogaev et al. 2006	DQ316067	Eastern Siberia	Enmynveen	32850	900	37489	1310	D15
nd	Krause et al. 2006	NC007596	Eastern Siberia	Berelekh	12170	50	14203	228	E21
M26	Gilbert et al. 2007	EU153454	Eastern Siberia	nd	24740	110	29798	318	D13
M8	Gilbert et al. 2007	EU153458	Eastern Siberia	Magadan area	46900	700	50439	2029	D1
ILC.01(10915)	Barnes et al. 2007	AM746193	Eastern Siberia	Kamchatka	45200	800	48496	1749	D2
M22	Gilbert et al. 2007	EU153452	Eastern Siberia	main Lyakhov island	50200	900	56625	2702	D4
BL-0865	Barnes et al. 2007	AM746177	Eastern Siberia	main Lyakhov island	32500	500	37015	878	D1
BL-013	Barnes et al. 2007	AM746178	Eastern Siberia	main Lyakhov island	37800	900	42442	712	D1
BL-0179	Barnes et al. 2007	AM746179	Eastern Siberia	main Lyakhov island	>35600	na	>40682	na	na*
BL-059	Barnes et al. 2007	AM746180	Eastern Siberia	main Lyakhov island	>42400	na	>45726	na	E16
BL-0723	Barnes et al. 2007	AM746181	Eastern Siberia	main Lyakhov island	>33000	na	>37505	na	A4
BL-0585	Barnes et al. 2007	AM746182	Eastern Siberia	main Lyakhov island	40200	900	43937	862	D14
BL-0503	Barnes et al. 2007	AM746183	Eastern Siberia	main Lyakhov island	>47700	na	>51532	na	A4
BL-0208	Barnes et al. 2007	AM746184	Eastern Siberia	main Lyakhov island	25900	600	30803	594	D12
BL-0443	Barnes et al. 2007	AM746185	Eastern Siberia	main Lyakhov island	48000	2000	52258	3178	A7
BL-0240	Barnes et al. 2007	AM746186	Eastern Siberia	main Lyakhov island	>51000	na	>57769	na	A2
BL-0310	Barnes et al. 2007	AM746187	Eastern Siberia	main Lyakhov island	>44000	na	>47438	na	A5
BL-0309	Barnes et al. 2007	AM746188	Eastern Siberia	main Lyakhov island	>47000	na	>50576	na	E10
BL-0250	Barnes et al. 2007	AM746189	Eastern Siberia	main Lyakhov island	12030	60	14016	210	D1
BL-0219	Barnes et al. 2007	AM746190	Eastern Siberia	main Lyakhov island	12500	500	14960	885	E9
BL-0308	Barnes et al. 2007	AM746191	Eastern Siberia	main Lyakhov island	43600	1000	47179	1754	A5
ILC.04(11028)	Barnes et al. 2007	AM746194	Eastern Siberia	main Lyakhov island	36610	360	41668	327	E10
WS.02	Barnes et al. 2007	AM746197	Western Siberia	Lugovskoe	13455	60	16407	412	E8
WS.04	Barnes et al. 2007	AM746198	Western Siberia	Kochegur	25300	290	30194	350	E2
M18	Gilbert et al. 2007	EU153447	Central Siberia	Gydan peninsula	17125	70	20511	338	E1
ILC.06(1138a)	Barnes et al. 2007	AM746195	Taimyr peninsula	Khatanga river mouth	46600	1100	50104	2260	D1
M2	Gilbert et al. 2007	EU153449	Taimyr peninsula	nd	20380	140	24321	322	E22
M3	Gilbert et al. 2007	EU153455	Taimyr peninsula	nd	20620	70	24615	249	D5
ILC.08(10643)	Barnes et al. 2007	AM746196	Wrangel Island	nd	25890	140	30901	347	E1
M1	Gilbert et al. 2007	EU153444	northern Yakutia?	nd	nd	nd	nd	nd	E10
M4	Gilbert et al. 2007	EU153456	northern Yakutia?	nd	18545	70	22151	309	D1
M5	Gilbert et al. 2007	EU153457	northern Yakutia?	nd	nd	nd	nd	nd	D4