A comprehensive genomic history of extinct and living elephants

Proceedings of the National Academy of Sciences of the United States of America, 2018; 115(11):2566-2574

The author(s) retains copyright to individual PNAS articles, and the National Academy of Sciences of the United States of America (NAS) holds copyright to the collective work and retains an exclusive License to Publish these articles, except for open access articles submitted beginning September 2017. For such open access articles, NAS retains a nonexclusive License to Publish, and these articles are distributed under a CC BY-NC-ND license.

http://doi.org/10.1073/pnas.1720554115

27 November 2018
A comprehensive genomic history of extinct and living elephants


*Department of Genetics, Harvard Medical School, Boston, MA 02115; †Broad Institute of MIT and Harvard, Cambridge, MA 02142; ‡Bioinformatics Research Centre, Aarhus University, DK-8000 Aarhus, Denmark; §Unit of General Zoology, Evolutionary Adaptive Genomics, Institute of Biochemistry and Biology, Faculty of Mathematics and Life Sciences, University of Potsdam, 14467 Potsdam, Germany; ¶McMaster Ancient DNA Centre, Department of Anthropology, McMaster University, Hamilton, ON L8S 4L9, Canada; †‡Department of Biochemistry, McMaster University, Hamilton, ON L8S 4L8, Canada; ‡§The Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, ON L8S 4L8, Canada; ¶§Department of Genetics and Evolution, School of Biological Sciences, The University of Adelaide, Adelaide, 5005 SA, Australia; †¶School of Computer Science, The University of Adelaide, 5005 SA, Australia; ¶‡Center of Archaeological and Cultural Human History, Danube Private University, A-3500 Krems, Austria; †¶Department of Biomedical Engineering, University Hospital Basel, University of Basel, CH-4123 Basel, Switzerland; ‡¶Integrative Prehistory and Archaeology, University of Basel, CH-4055 Basel, Switzerland; †§Institute of Maritime History, Tall Timbers, MD 20690; ‡§Department of Bioinformatics and Genetics, Swedish Museum of Natural History, SE-10405 Stockholm, Sweden; ¶¶Division of Vertebrate Zoology/Mammalogy, American Museum of Natural History, New York, NY 10024; ¶¶State Office for Heritage Management and Archaeology, 06114 Halle (Saale), Germany; †§Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801; †¶Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801; ‡¶Institute for Conservation Research, San Diego Zoo, Escondido, CA 92027; ‡§Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, NC 27607; ¶¶European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, CB10 1SD Cambridge, United Kingdom; ¶¶Wellcome Sanger Institute, Hinxton, CB10 1SD Cambridge, United Kingdom; ‡¶Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01655; †¶Earlham Institute, NR4 7UZ Norwich, United Kingdom; ‡¶Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, 751 23 Uppsala, Sweden; and †¶¶Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115

Edited by David M. Hillis, The University of Texas at Austin, Austin, TX, and approved January 24, 2018 (received for review November 24, 2017)

Elephantids are the world’s most iconic megafaunal family, yet there is no comprehensive genomic assessment of their relationships. We report a total of 14 genomes, including 2 from the American mastodon, which is an extinct elephantid relative, and 12 living elephants comprising several species, of which the once circumpolar woolly mammoth (Mammuthus primigenius) survived in small isolated island populations well into the Holocene until ~4,000 y ago.

Significance

Elephantids were once among the most widespread megafaunal families. However, only three species of this family exist today. To reconstruct their evolutionary history, we generated 14 genomes from living and extinct elephantids and from the American mastodon. While previous studies examined only simple bifurcating relationships, we found that gene flow between elephantid species was common in the past. Straight-tusked elephants descend from a mixture of three ancestral populations related to the ancestor of African elephants, woolly mammoths, and present-day forest elephants. We detected interbreeding between North American woolly and Columbian mammoths but found no evidence of recent gene flow between forest and savanna elephants, demonstrating that both gene flow and isolation have been central in the evolution of elephantids.


Data deposition: The sequence data have been deposited in the European Nucleotide Archive (accession no. PRJEB24361). The most recent update of the savanna elephant reference genome (LoxAfr4) is available at ftp://ftp.broadinstitute.org/pub/assembly/mammals/elephantid/LoxAfr4. Previously published data that were reprocessed in this study are available at https://reirch.hms.harvard.edu/datasets.

1To whom correspondence may be addressed. Email: elle.palkopoulou@gmail.com or reich@genetics.med.harvard.edu.

2K.L.-T., M.H., H.P., and D.R. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1720554115/-/DCSupplemental.

Published online February 26, 2018.
(4, 5) while the more temperate North American Columbian mammoth (*Mammuthus columbi*) disappeared by the end of the last ice age ~11,000 y ago (6, 7). Straight-tusked elephants (genus *Palaeoloxodon*) potentially survived as late as ~50,000 to 35,000 y ago (8) and have been conventionally grouped within *Elephas* (3, 9), but recent genomic evidence from European straight-tusked elephants (*Palaeoloxodon antiquus*) over 100,000 y old showed that they were on average more closely related to forest elephants than to any other extant species and led to the suggestion that they were an ancient sister group of modern African forest elephants (10).

**Results and Discussion**

**A High-Quality Elephant Reference Genome.** This study formally reports the high-quality reference genome of the African savanna elephant, which first became available online in May 2005 (*LoxAfr1*) and has since been iteratively updated with the latest release available online in May 2014 (*LoxAfr4*). We used classic Sanger-sequencing methods to generate a de novo genome assembly from a savanna elephant at 6.8-fold coverage. Specifically, we performed paired-end Sanger sequencing using multiple insert sizes [4 kilobases (kb), 10 kb, 40 kb, and BAC clones]. We then used FISH mapping of BAC clones to place scaffolds containing 85% of the assembly onto chromosomes. The assembly has a median (N50) contig length of 69 kb and a median scaffold length of 48 megabases, with a total assembly length of 3.2 gigabases (*SI Appendix*, Table S1.1). The assembly contains 47.8% easily recognized repeat-derived sequences (28.9% long interspersed nuclear elements, 6.7% short interspersed nuclear elements, 6.7% long terminal repeats, 0.5% simple repeats, and 3.0% “other”) and 20,333 protein coding genes.

**Proboscidean Dataset and Genome-Wide Phylogeny.** In addition to the African savanna elephant reference genome, we generated genome-wide data from 14 proboscidean specimens, one of which was from the same savanna elephant individual from which the reference genome was sequenced (*SI Appendix*, Note 3). Using Illumina paired-end reads, we performed deep shotgun sequencing of the genomes of seven elephants: two forest, two savanna, and two Asian elephants ranging in coverage from 28- to 39-fold (Table 1), and an ~120,000-y-old straight-tusked elephant whose coverage we increased from the previously reported (10) 0.65-fold to ~15-fold. We also generated low- to medium-coverage genomes (0.5-fold to ~sixfold) from four woolly mammoths, one Columbian mammoth, and two American mastodons (*Mammut americanum*). The mastodon diverged from elephants ~20 to 30 Mya (11) and hence represents an appropriate outgroup for studying Elephantidae evolution. We analyzed these data together with previously published genomes from two woolly mammoths (12) and four Asian elephants (13, 14), as well as low-coverage genomic data from a second straight-tusked elephant (10).

To obtain an overview of the relationships among the genomes, we built phylogenetic trees based on different features of the data. Neighbor-joining trees using pairwise divergence per nucleotide recapitulated previously reported relationships (10, 15) (Fig. 1 and *SI Appendix*, Fig. S8.1), as did trees based on the presence or absence of interspersed repeats in either a maximum parsimony or maximum likelihood analysis, with the exception of the placement of straight-tusked elephants in the latter (*SI Appendix*, Fig. S9.8). While straight-tusked elephants were recently found to cluster within the mitochondrial diversity of forest elephants (10) (*SI Appendix*, Fig. S7.1), we show that the nuclear genomes of these taxa form separate clades in the reconstructed trees (Fig. 1). The two forest elephants in our dataset (one from the Guinean and one from the Congolian forest block, spanning the phylogeographic diversity of *L. cyclotis*) (Table 1) also comprise a lineage that is distinct from savanna elephants, confirming with complete nuclear genomes that the two African elephants should be classified as distinct taxa. However, our further analyses showed that the average trees do not capture the full complexity of the evolutionary history of elephantid species and in particular obscure major admixture events, which were central features of elephantid evolution.

**Interspecies Admixture Events.** To test for evidence of admixture, we computed *D*-statistics (16–18), which use patterns of shared derived alleles to assess genetic affinities within and between

---

**Table 1.** Proboscidean samples analyzed in this study

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Geographic origin</th>
<th>Date, y before present</th>
<th>Sequencing (source)</th>
<th>No. of mapped reads, million</th>
<th>Average coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. cyclotis</em> <em>A</em></td>
<td>Central African Republic</td>
<td>Modern</td>
<td>This study (BI)</td>
<td>906</td>
<td>27.78</td>
</tr>
<tr>
<td><em>L. africana</em> <em>B</em></td>
<td>Kenya</td>
<td>Modern</td>
<td>This study (BI)</td>
<td>1,001</td>
<td>30.44</td>
</tr>
<tr>
<td><em>L. africana</em> <em>C</em></td>
<td>South Africa</td>
<td>Modern</td>
<td>This study (BI)</td>
<td>1,114</td>
<td>33.42</td>
</tr>
<tr>
<td><em>E. maximus</em> <em>D</em></td>
<td>Myanmar</td>
<td>Modern</td>
<td>This study (BI)</td>
<td>1,283</td>
<td>38.94</td>
</tr>
<tr>
<td><em>E. maximus</em> <em>E</em></td>
<td>Malaysia (Borneo)</td>
<td>Modern</td>
<td>This study (BI)</td>
<td>1,107</td>
<td>32.20</td>
</tr>
<tr>
<td><em>E. maximus</em> <em>F</em></td>
<td>Sierra Leone</td>
<td>Modern</td>
<td>This study (BI)</td>
<td>1,074</td>
<td>32.06</td>
</tr>
<tr>
<td><em>M. primigenius</em> <em>G</em></td>
<td>Taimyr Peninsula, Russia</td>
<td>~31,500</td>
<td>This study (HMS)</td>
<td>55</td>
<td>0.60</td>
</tr>
<tr>
<td><em>M. primigenius</em> <em>H</em></td>
<td>Alaska, USA</td>
<td>~44,900</td>
<td>This study (HMS)</td>
<td>27</td>
<td>0.49</td>
</tr>
<tr>
<td><em>M. americanum</em> <em>I</em></td>
<td>Alaska, USA</td>
<td>&gt;50,000</td>
<td>This study (IFT, HMS)</td>
<td>399</td>
<td>3.96</td>
</tr>
<tr>
<td><em>E. maximus</em> <em>L</em></td>
<td>India*</td>
<td>Modern</td>
<td>(13)</td>
<td>889</td>
<td>27.02</td>
</tr>
<tr>
<td><em>E. maximus</em> <em>M</em></td>
<td>India*</td>
<td>Modern</td>
<td>(13)</td>
<td>1,014</td>
<td>30.27</td>
</tr>
<tr>
<td><em>P. antiquus</em> <em>N</em></td>
<td>Germany</td>
<td>~120,000</td>
<td>This study (BI, HMS)</td>
<td>1,399</td>
<td>14.64</td>
</tr>
<tr>
<td><em>P. antiquus</em> <em>O</em></td>
<td>Germany</td>
<td>~120,000</td>
<td>(10)</td>
<td>12</td>
<td>0.14</td>
</tr>
<tr>
<td><em>M. primigenius</em> <em>P</em></td>
<td>Oimyakon, Russia</td>
<td>~44,800</td>
<td>(12)</td>
<td>902</td>
<td>12.77</td>
</tr>
<tr>
<td><em>M. primigenius</em> <em>Q</em></td>
<td>Wrangel Island, Russia</td>
<td>~4,300</td>
<td>(12)</td>
<td>959</td>
<td>19.00</td>
</tr>
<tr>
<td><em>M. primigenius</em> <em>S</em></td>
<td>Yamal Peninsula, Russia</td>
<td>~45,300</td>
<td>This study (IFT, HMS)</td>
<td>132</td>
<td>0.91</td>
</tr>
<tr>
<td><em>M. columbi</em> <em>U</em></td>
<td>Wyoming, USA</td>
<td>~13,400</td>
<td>This study (IFT, HMS)</td>
<td>122</td>
<td>1.53</td>
</tr>
<tr>
<td><em>Mammuthus</em> <em>V</em></td>
<td>Wyoming, USA</td>
<td>~42,400</td>
<td>This study (IFT, HMS)</td>
<td>830</td>
<td>5.86</td>
</tr>
<tr>
<td><em>M. americanum</em> <em>X</em></td>
<td>Gulf of Maine, USA</td>
<td>~13,400</td>
<td>This study (HMS)</td>
<td>71</td>
<td>0.79</td>
</tr>
<tr>
<td><em>E. maximus</em> <em>Y</em></td>
<td>Assam, India</td>
<td>Modern</td>
<td>(13)</td>
<td>1,239</td>
<td>35.90</td>
</tr>
<tr>
<td><em>E. maximus</em> <em>Z</em></td>
<td>Karnataka, India</td>
<td>Modern</td>
<td>(14)</td>
<td>447</td>
<td>14.58</td>
</tr>
</tbody>
</table>

*BI, Broad Institute; HMS, Harvard Medical School; IFT, Illumina Fast Track Services.  
*Exact geographic origin is unknown.*
Fig. 1. Neighbor-joining tree from pairwise genetic divergence of proboscidean genome sequences. The phylogeny estimated from all substitutions is shown with results from transversions only in the dashed gray box, which differ in topology only within the woolly mammoth clade. Support values from 100 bootstrap replicates are given inside each node (values from all substitutions/transversions only). The average depth of coverage for each genome is listed inside parentheses next to the tip label. This phylogeny depicts the average relationships between elephantid species and does not fully capture their complex evolutionary history (Fig. 2A).

taxa (SI Appendix, Note 11). We integrated the observed signals of gene flow into a single historical model using qpGraph (18), which fits parameters of an admixture graph model (phylogenetic tree augmented with admixture events) by comparing empirical and predicted f-statistics (16). The admixture graph that most parsimoniously fit the data (Fig. 2A and SI Appendix, Figs. S12.2–S12.4) captured all of the patterns in the individual D-statistics and revealed a more complex history than can be captured by a simple tree-like topology (Fig. 1).

A major surprise that emerged from this analysis is the highly reticulated relationship between straight-tusked elephants and the other species. In contrast to previous work that has shown that straight-tusked elephants are on average more closely related to forest elephants than to any other species (10), we found that they do not form a simple clade with forest elephants. The fitted admixture graph revealed three major genetic components for straight-tusked elephants, the largest of which derived from a lineage that is basal to the common ancestor of forest and savanna elephants (Fig. 2A). This finding may help to reconcile the genomic data with the fossil record of elephantids in Africa because species of Palaeoloxodon predominate in the fossil record during most of the Pliocene and Pleistocene and are believed to have given rise to the Eurasian straight-tusked elephant (2, 19).

The remaining genetic contribution to straight-tusked elephants derived from two separate lineages, one related to woolly mammoths and the other related to extant forest elephants (Fig. 2A). Specifically, woolly mammoths, as well as Asian elephants, shared more derived alleles with straight-tusked elephants than expected and the signal was significantly stronger for mammoths than for Asian elephants (Z = 9.25) (Table 2). This pattern is most parsimoniously explained by 6 to 10% admixture from a population related to woolly mammoths into the straight-tusked elephant lineage (Fig. 2A), which could help to resolve an apparent discrepancy. While phylogenetic trees based on genomewide nuclear (Fig. 1) and mtDNA data (10) (SI Appendix, Fig. S7.1) place straight-tusked elephants as closest to forest elephants (due to an additional admixture event described below), morphological criteria have traditionally placed straight-tusked elephants within Elephas (3, 9). The morphological similarity to Asian elephants could be accounted for through hybridization from an ancestral population that split off from the mammoth lineage early in its history, close in time to the common ancestor of Asian elephants and mammoths. This would imply that morphological characters shared between straight-tusked and Asian elephants were present in the common ancestor of Asian elephants and mammoths, and thus became lost from the mammoth lineage. Alternatively, the morphological similarities between straight-tusked elephants and Asian elephants could also be due to homoplasies resulting from convergent evolution, for which there is considerable evidence in the elephantid fossil record (1–3).

Secondly, straight-tusked elephants shared significantly more derived alleles with one of our sequenced forest elephants (L. cyclotis_F from the Guinean forest block in West Africa) than with the other (7 ≤ |Z| ≤ 9) (Fig. 2B). The fitted admixture graph indicates that the straight-tusked elephant derives 35 to 39% of its ancestry from a lineage related to the West African forest elephant (L. cyclotis_F) (Fig. 2A). This admixture proportion explains the apparent placement of straight-tusked elephants as most closely related to forest elephants in the phylogenetic trees in Fig. 1 and ref. 10. Given the geographic separation and deep divergence between our sampled forest elephants (see below), gene flow from a derived forest elephant lineage into the straight-tusked elephant lineage is plausible and likely occurred in Africa. The intraspecies split time between the West and Central African forest elephants (L. cyclotis_A and L. cyclotis_F; 609,000 to 463,000 y ago subject to mutation rate uncertainty) (see Fig. 4A) and the approximate date of our sequenced straight-tusked elephants (~120,000 y ago) place upper and lower bounds on the date of the inferred gene flow. This interval, however, overlaps several glacial cycles. In Africa, glacial periods involved drier conditions, contraction of rainforest habitats, and expansion of grassland (20) while in interglacial periods involved the opposite. Such ecological factors may have had important consequences for the biota, including facilitating or inhibiting hybridization among related taxa. The
true evolutionary history of straight-tusked elephants could have been even more complex; the models reported here are the simplest scenarios that can explain the data.

Within the genus *Mammuthus*, we detected nuclear admixture between woolly and Columbian mammoths, confirming previous claims of interbreeding based on fossil evidence and mitochondrial DNA (7, 21). The Columbian mammoth specimen (*M. columbi, U*) is sister to all woolly mammoths in the average tree of relationships (Fig. 1). However, this specimen is not symmetrically related to Asian elephants (Fig. S11.1 and Table S11.7). Positive values indicate excess genetic affinity between L. cyclotis F and X while negative values indicate excess genetic affinity between L. cyclotis F and X. Bars correspond to one SE in either direction. The statistic highlighted in red is significant (*Z* > 3) and indicates an excess of shared derived alleles between the straight-tusked elephant and L. cyclotis F. Remaining key D-statistics supporting the admixture graph are shown in Table 2. All inferences are based on transversion polymorphisms only.

**Interspecies Demographic Inference.** We inferred effective population sizes, split times, and migration rates using three separate, complementary approaches. We converted estimates of genetic divergence to absolute time in years, assuming a point mutation rate of 0.406 × 10−9 per base per year (as calculated in *SI Appendix, Note 16*) and a generation interval of 31 y (as in ref. 4).
However, we caution that the elephantid mutation rate is highly uncertain (12) and, when more accurate estimates become available in the future, all absolute time estimates should be rescaled (but relative estimates should remain unchanged).

First, we applied approximate Bayesian computation (ABC) to fit demographic models based on a set of summary statistics consisting of the allelic states of pairs of adjacent variable sites (30) in alignments of three elephantid sequences and the mastodon, as well as estimates of pairwise divergence and D-statistics (SI Appendix, Note 16). Consistent with our pairwise sequential Markovian coalescent (PSMC) results (shown below), inferred ancestral effective population sizes (Fig. 3) were largest for the ancestors of forest, savanna, and straight-tusked elephants, followed by the ancestors of Asian elephants and woolly/Columbian mammoths, and smallest for the common ancestral population of all elephantids, although all confidence intervals (CIs) were overlapping (CI, respectively: 37,000 to 233,000; 10,000 to 130,000; and 7,000 to 78,000).

Forest and savanna elephants are inferred to have split from each other ~5 to 2 Mya, soon after their common ancestor split from the straight-tusked elephant lineage. The split between Columbian and woolly mammoths is inferred to have occurred 1.5 to 0.7 Mya, consistent with some, but not all, paleontological estimates (7, 31). Asian elephants and mammoths are estimated to have split at about the same time as the split between Loxodonta and straight-tusked elephants while the initial split within the Elephantidae is inferred to have occurred ~10 to 5 Mya, in good agreement with the divergence time of Loxodonta and Asian elephants/mammoths inferred from the fossil record (15) (9 to 4.2 Mya). All elephantids are estimated to have split from the mastodon at ~28 to 10 Mya, with the upper end of this range in line with evidence from the fossil record (19) (28 to 24 Mya).

The highest migration rate is inferred between forest and straight-tusked elephants (CI: 0.49 × 10^−6 to 1.49 × 10^−6; proportion of migrants per generation), consistent with the largest admixture proportion estimated by the admixture graph and f_2-ratio tests (Fig. 24 and SI Appendix, Table S11.8). These are followed by the migration rates between straight-tusked elephants and woolly mammoths (1.84 × 10^{-7} to 6.44 × 10^{-5}), and between straight-tusked and Asian elephants (1.32 × 10^{-5} to 5.71 × 10^{-5}), which is again in agreement with the findings from D-statistics and the admixture graph.

Second, we used a coalescent hidden Markov model (32) (CoalHMM) to infer split times and ancestral effective population

---

**Fig. 3.** A consensus demographic model for the history of elephants. Inferred parameters from three modeling approaches are shown: (i) coalescent simulations with approximate Bayesian computation (ABC), (ii) incomplete lineage sorting analysis (ILS CoalHMM), and (iii) isolation-and-migration models (IM CoalHMM). Dark red arrows indicate gene flow as inferred from the ABC analysis, with arrow thickness corresponding to the extent of gene flow. Shaded areas below the separation of species indicate a limited period of gene flow between incipient species as inferred from the IM CoalHMM analysis. Gene flow rate is shown below the shaded areas as the fraction of migrations per lineage per generation. Effective population sizes (N_e) and split times (t_e) correspond to the 95% confidence intervals obtained from the ABC analysis (green), the mean estimates obtained from the ILS CoalHMM analysis (blue), and the bootstrap intervals obtained from the IM CoalHMM analysis (black). Split times are given in million y before present, with t_e-beg referring to the initial split time and t_e-end to the end of the migration period (for the IM CoalHMM analysis). LOX refers to the common ancestor of savanna and forest elephants, L-P to the common ancestor of Loxodonta and straight-tusked elephants, MAM to the common ancestor of woolly and Columbian mammoths, E-M to the common ancestor of Asian elephants and mammoths, ELE to the common ancestor of all elephantids, and ANC to the common ancestor of elephants and the American mastodon. Branch lengths, splits, and migration rate periods are not drawn to scale.
sizes for selected trios of elephantid species based on incomplete lineage sorting (ILS) ([SI Appendix, Note 17]. ILS is reflected in regions of the genome where taxa that are not most closely related in the species tree cluster together ([15, 33, 34]). Here, we also incorporated data from chromosome X to test for evidence of sex-biased demography. These analyses support the evidence from ABC analysis that the autosomal Ne for the ancestor of forest and savanna elephants (mean: 165,000 individuals) is higher than that for the ancestor of Asian elephants and woolly mammoths (mean: 72,000) ([Fig. 3]), and for the common ancestor of all elephantids (48,000 to 53,000, range of means obtained from analyses of different elephantid trios). Forest and savanna elephants are inferred to have split at ~2 Mya, Asian elephants and woolly mammoths at 2.5 Mya, and all elephantids at 5.6 to 5 Mya ([Fig. 3]). These dates overlap with the lower end of the ranges obtained from the ABC analysis, with the younger average dates from the CoAlHMM model likely due to the absence of migration in the model (see also below).

For all analyzed species trios, the observed X-to-autosome ratio of Ne was lower than 3/4 (the baseline value for a simple demography), even though a higher ratio might be expected considering the higher variance in male reproductive success in elephants ([35, 36]). Potential factors that could explain this discrepancy include linked selection ([37] on chromosome X or male-biased gene flow ([38]).

An examination of the ILS patterns revealed that, in the forest, straight-tusked, and Asian elephant trio, a higher proportion of regions clustered together straight-tusked and Asian elephants (18.8 to 20.5%) rather than forest and Asian elephants (15.3 to 16.0%) ([SI Appendix, Figs. S17.15–S17.18]), consistent with the gene flow indicated in the best-fit admixture graph ([Fig. 2A and Table 2]). We did not observe a substantial ILS asymmetry in the trio of Asian elephants, woolly mammoths, and straight-tusked elephants ([SI Appendix, Figs. S17.13 and S17.14]), but we believe this is still compatible with the findings from the admixture graph analysis, given the proportion of woolly mammoth-related ancestry in straight-tusked elephants, and its source splitting off relatively close to the common ancestor of Asian elephants and woolly mammoths ([Fig. 2A]).

Finally, we applied CoAlHMM for pairs of elephantid species under isolation-–and-migration (IM) models, allowing for the possibility of continuing gene flow after initial population separation ([39]) ([SI Appendix, Note 18]). Our autosomal IM CoAlHMM analysis strongly supports the presence of migration after initial separation for all interspecies pairs ([Fig. 3] and [SI Appendix, Fig. S18.1]). Consistent with our other analyses, the highest gene flow rates were estimated between the forest and straight-tusked elephant lineages ([C.I.: 1.00 × 10⁻⁵ to 1.49 × 10⁻⁵]). Gene flow between the ancestors of forest and savanna elephants is inferred to have occurred from their split ~5.3 Mya ([C.I.: 5.6 to 2.6 Mya]) until 1.3 Mya ([C.I.: 3.0 to 1.2 Mya]) for pairs including L. cyclotis_A and 1.4 to 0.1 Mya for pairs including L. cyclotis_F) although the D-statistics and admixture graph analyses did not provide any evidence of recent gene flow between the two species. Overall, split times were quite similar to those estimated via ABC while estimates of ancestral Ne were mostly lower than those obtained from the ILS CoAlHMM analysis but similar (except with tighter confidence intervals) to those from ABC ([Fig. 3]).

**Within-Species Analyses: Diversity, Population Size Change, and Population Substructure.** Estimates of genetic diversity for the high-coverage genomes (n = 13) indicated, consistent with previous reports, that African forest elephants harbor the highest levels of heterozygosity (0.00085 to 0.00364) ([Fig. 4B] and sequence divergence ([SI Appendix, Table S8.1]) among extant and extinct elephants ([15, 40–42]). Mammoths, straight-tusked elephants, and Asian elephants displayed intermediate levels of heterozygosity (0.00093 to 0.00167) ([Fig. 4B]), except for E. maximus_E from Malaysian Borneo, which had extremely low heterozygosity (0.00032). Savannah elephants exhibited the lowest heterozygosity among all elephantids (0.000085 to 0.000088) ([Fig. 4B]).

To reconstruct elephantid population size changes over time, we used the PSMC ([43]) ([SI Appendix, Note 14]). The two forest elephants had similar population size histories before ~370,000 y ago but very different ones thereafter. Current effective population size (Ne) in L. cyclotis_F (from the smaller Guinean forest block in West Africa) was ~fourfold lower than in L. cyclotis_A (from the larger Congolian forest block in Central Africa) ([Fig. 4C]), in line with the ~21% lower heterozygosity in the former. The two savanna elephants had lower Ne relative to forest elephants for hundreds of thousands of years ([Fig. 4D]), potentially reflecting ecological competition from the African elephant Palaeoloxodon recki (including Palaeoloxodon iolensis) that dominated the African savannas until the Late Pleistocene ([2, 19]), or the high levels of male–male competition documented in this species.

Early in its history (>1 Mya), the straight-tusked elephant had a population size trajectory similar to that of forest and savanna elephants ([Fig. 4C]), including a period of population expansion ~2 Mya followed by decline. This observation may be explained by evidence that these species share deep ancestry ([Fig. 2A]). Asian elephants are inferred to have gone through a phase of population growth, succeeded by decline ~120,000 y ago, resulting in a current Ne estimated to be about half that of savanna elephants ([Fig. 4E]). The population sizes of the two woolly mammoths are inferred to have been similar before their split, but, subsequently, the ancestors of the Wrangel Island mammoth experienced a severe bottleneck ([Fig. 4F]), which led to an ~20% drop in heterozygosity, as shown earlier in the study that reported the Wrangel and mainland Siberian mammoth genomes ([12]).

We estimated split times of elephantids within species using the F(4fB) statistic ([17]), which measures the fraction of heterozygous positions discovered in one individual that are derived in a randomly sampled chromosome from an individual of a second population of the same species ([SI Appendix, Note 15]). This fraction is expected to decrease as a function of population separation time (reflecting the fact that, for an older split, a greater proportion of discovered mutations will have occurred after population divergence), with the exact form of the decay depending on the demographic history of the first individual, which we can infer using PSMC. The oldest intraspecific split within elephantid taxa was estimated between the two forest elephants (L. cyclotis_A and L. cyclotis_F; 609,000 to 463,000 y ago) ([Fig. 4A]). This is consistent with a hypothesis of deep population structure with limited gene flow, as well as with the high ancestral Ne among forest elephants ([15]). By contrast, the two savanna elephants were estimated to have split from each other only 38,000 to 30,000 y ago, in line with their nearly identical Ne curves ([Fig. 4D]), as well as with a previous hypothesis for a relatively recent founder event ([40, 41]), and with high levels of male dispersal documented in this species ([44]). Among Asian elephants, split times were oldest between the Bornean E. maximus_C and other individuals (190,000 to 103,000 y ago) ([Fig. 4D]), consistent with the uniqueness of the mitochondrial DNA haplogroup of elephants in Malaysian Borneo ([45]). The Asian elephant from Myanmar (E. maximus_D) exhibited higher heterozygosity than other Asian elephants and intermediate split times with elephants from India (43,000 to 24,000 y ago), compatible with a hypothesized secondary admixture of diverged populations that may have occurred in this part of Southeast Asia, as suggested by mitochondrial DNA ([46]). Within Mammuthus, the inferred interspecific split between Columbian mammoths and Eurasian woolly mammoths 712,000 to 423,000 y ago, was overlapping but mostly lower than that obtained from the ABC analysis described above (1.5 to 0.7 Mya), but still far older than that between the two Eurasian...
woolly mammoths (*M. primigenius_P* and *M. primigenius_Q*; 225,000 to 112,000 y ago) (Fig. 4).

**Conclusion**

Our genomic analyses of present-day and extinct elephantids revealed a history of multiple major interspecies admixture events. Evidence for gene flow among closely related mammalian species is not unprecedented. Examples include cases of unidirectional gene flow [e.g., from polar bears into brown bears (47), similar to the Columbian mammoth gene flow into woolly mammoths observed in our study]; emergence of admixed species [e.g., North American wolves with ancestry from coyotes and gray wolves (48), similar to the straight-tusked elephants in our study]; different extents of gene flow [e.g., between gray wolves and Eurasian/African golden jackals (49), and between bonobos and central/eastern chimpanzees (50)], as in the case of straight-tusked elephants and west African forest elephants/woolly mammoths in our study; extended periods of gene flow during the initial diversification of species [e.g., between eastern and western gorillas (39), Sumatran and Bornean orangutans (39), and the ancestors of humans and chimpanzees (39, 51), like those inferred from most pairwise species comparisons in our study]; different extents of gene flow [e.g., between gray wolves and Eurasian/African golden jackals (49), and between bonobos and central/eastern chimpanzees (50)], as in the case of straight-tusked elephants and west African forest elephants/woolly mammoths in our study; extended periods of gene flow during the initial diversification of species [e.g., between eastern and western gorillas (39), Sumatran and Bornean orangutans (39), and the ancestors of humans and chimpanzees (39, 51), like those inferred from most pairwise species comparisons in our study]; and adaptive introgression [e.g., in the great cats of the genus *Panthera* (52)], which could have played an important role in the
evolution of elephants as well. Our results in elephants thus add to the growing weight of evidence in favor of the view that capacity for hybridization is the norm rather than the exception in many mammalian species over a time scale of millions of years. Three different outcomes followed interspecies hybridization among elephants: emergence of a species with three ancestral genetic components (straight-tusked elephants); the continued isolation of species and lack of genome-wide introgression even after recurrent hybridization (forest and savanna elephants); or a modest degree of introgression (Columbian and North American woolly mammoths). An important priority for future work should be to explore whether admixture was not only an important phenomenon in the demographic history of the elephants, but also played a biologically important role in their evolution, facilitating adaptation after migration into new habitats, or in the face of fluctuating climatic conditions and resulting ecological shifts (53).

Materials and Methods
Detailed information on the samples and methods is provided in SI Appendix, including de novo genome assembly, mitochondrial phylogeny, and analysis of repetitive elements.

Genome Sequencing. Illumina libraries were prepared from genomic DNA of six modern elephants and sequenced at the Broad Institute. Illumina genomic libraries were also prepared for seven ancient proboscideans, following established methods (54, 55), and were sequenced together with previously generated libraries (10) at the Broad Institute, Harvard Medical School, and Illumina Fast Track Services.

Data Processing. Paired-end reads were trimmed and merged (ancient data) or trimmed only (modern data) with SeqPrep v.1.1 (https://github.com/jstjohn/SeqPrep), aligned against the African savanna elephant reference genome (LoxAfr4) with Burrows-Wheeler Aligner (BWA) (56), using parameters optimized for ancient DNA or default parameters, and converted to bam format with SAMtools (57) v.0.1.19. Duplicate reads were discarded using a custom python script or the SAMtools "rmdup" command. Previously published genomes for two woolly mammoths (12), two straight-tusked elephants (10), and four Asian elephants (13, 14) were also reprocessed and included in the dataset. Applied filters included base quality threshold of 30, mapping quality of ≥30 or 37, and mappability filters as described in SI Appendix, Note 6.

Sequence Divergence. Pseudohaploid sequences of chromosomes 1 to 27 were generated for each elephant with single randomly sampled alleles per site to eliminate reference alignment biases (as explained in detail in SI Appendix, Note 6). Pairwise sequence divergence was estimated from alignments ranging in size from 45 Mbp to 1,609 Mbp, based on all substitutions or only transversion SNPs only, and the reconstructed PSMC to infer the decay of this statistic as a function of population split time. Time estimates were rescaled assuming the mutation rate and generation time described above. Within-species population split times were estimated using the F(4) statistic (17) as implemented in the software POPSTATS, using transversion SNPs only, and the reconstructed PSMC to infer the decay of this statistic as a function of population split time. Time estimates were rescaled assuming the mutation rate and generation time described above.

ACKNOWLEDGMENTS. We thank Bruce Upchurch and Teri Hermann (Woodland Park Zoological Gardens), Melissa Dickson (Dickerson Park Zoo), David Shepherson (Oregon Zoo), and Nicholas J. Georgiadis (University of Washington) for providing elephant samples; Bernard Buigues (International Mammoth Committee), Pamela Groves (Institute of Arctic Biology), Daniel Fisher (University of Michigan), Mark Clementz (University of Wyoming), and Paul Matheus and Dale Guthrie (University of Alaska, Fairbanks) for providing ancient proboscidean samples; Jacob Enk (McMaster University and University of Victoria) for assisting in laboratory analyses; Pontus Skoglund (Harvard Medical School) for providing sequencing for population genetic analyses; Yasuko Ishida (University of Illinois) for quantifying and sending elephant samples to the Broad Institute; and Karol Schauer for the proboscidean drawings in Figs. 1 and 3. Deep sequencing of the straight-tusked elephant sample (P. antiquus, X) was funded by National Human Genome Research Institute Grant US4 HG003067-08. A.L.R. was supported by the US Fish and Wildlife Service African Elephant Conservation Fund. B.L.A. and M.R. were funded by Wellcome Trust Grants WT098051 and WT108749/Z/15/Z and by the European Molecular Biology Laboratory. M.H. was supported by European Research Council Consolidator Grant 310763 GeneFlow. H.P. was funded through a Natural Sciences and Engineering Research Council of Canada Discovery Grant R4105-10539150 and the Canada Research Chairs program. D.R. was funded by NSF (HOMINID) Grant BCS-1032255 and NIH (National Institute of General Medical Sciences) Grant GM100233 and is an Investigator of the Howard Hughes Medical Institute.


