A comprehensive genomic history of extinct and living elephants

Eleftheria Palkopoulou[a,b,1], Mark Lipson[a], Swapan Mallick[a,b], Svend Nielsen[c], Nadin Rohland[a], Sina Baleka[d], Emil Karpinski[e,f,g,h], Atma M. Ivancevic[i], Thu-Hien To[i], R. Daniel Kortschak[i], Joy M. Raison[i], Zhipeng Qu[i], Tat-Jun Chin[j], Kurt W. Alt[k,l,m], Stefan Claesson[n], Love Dalén[o], Ross D. E. MacPhee[p], Harald Meller[q], Alfred L. Roca[r,s], Oliver A. Ryder[t], David Heiman[b], Sarah Young[b], Matthew Breen[u], Christina Williams[u], Bronwen L. Aken[v,w], Magali Ruffier[v,w], Elinor Karlsson[b,x], Jeremy Johnson[b], Federica Di Palma[y], Jessica Alfoldi[b], David L. Adelson[i], Thomas Mailund[c], Kasper Munch[c], Kerstin Lindblad-Toh[b,z,2], Michael Hofreiter[d,2], Hendrik Poinar[e,f,g,h,2], and David Reich[a,b,aa,1,2]

[a]Department of Genetics, Harvard Medical School, Boston, MA 02115; [b]Broad Institute of MIT and Harvard, Cambridge, MA 02142; [c]Bioinformatics Research Centre, Aarhus University, DK-8000 Aarhus, Denmark; [d]Unit of General Zoology–Evolutionary Adaptive Genomics, Institute of Biochemistry and Biology, Faculty of Mathematics and Life Sciences, University of Potsdam, 14476 Potsdam, Germany; [e]McMaster Ancient DNA Centre, Department of Anthropology, McMaster University, Hamilton, ON L8S 4L9, Canada; [f]Department of Biology, McMaster University, Hamilton, ON L8S 4K1, Canada; [g]Department of Biochemistry, McMaster University, Hamilton, ON L8S 4L8, Canada; [h]The Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, ON L8S 4L8, Canada; [i]Department of Genetics and Evolution, School of Biological Sciences, The University of Adelaide, Adelaide, 5005 SA, Australia; [j]School of Computer Science, The University of Adelaide, 5005 SA, Australia; [k]Center of Natural and Cultural Human History, Danube Private University, A-3500 Krems, Austria; [l]Department of Biomedical Engineering, University Hospital Basel, University of Basel, CH-4123 Basel, Switzerland; [m]Integrative Prehistory and Archaeological Science, University of Basel, CH-4055 Basel, Switzerland; [n]Institute of Maritime History, Tall Timbers, MD 20690; [o]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, SE-10405 Stockholm, Sweden; [p]Division of Vertebrate Zoology/Mammalogy, American Museum of Natural History, New York, NY 10024; [q]State Office for Heritage Management and Archaeology, 06114 Halle (Saale), Germany; [r]Department of Animal Sciences, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [s]Institute for Genomic Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [t]Institute for Conservation Research, San Diego Zoo, Escondido, CA 92027; [u]Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, NC 27607; [v]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, CB10 1SD Cambridge, United Kingdom; [w]Wellcome Sanger Institute, Hinxton, CB10 1SD Cambridge, United Kingdom; [x]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01655; [y]Earlham Institute, NR4 7UZ Norwich, United Kingdom; [z]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, 751 23 Uppsala, Sweden; and [aa]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115

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 spanning all three extant and three extinct elephantid species including an ~120,000-y-old straight-tusked elephant, a Columbian mammoth, and woolly mammoths. Earlier genetic studies modeled elephantid evolution via simple bifurcating trees, but here we show that interspecies hybridization has been a recurrent feature of elephantid evolution. We found that the genetic makeup of the straight-tusked elephant, previously placed as a sister group to African forest elephants based on lower coverage data, in fact comprises three major components. Most of the straight-tusked elephant's ancestry derives from a lineage related to the ancestor of African elephants while its remaining ancestry consists of a large contribution from a lineage related to forest elephants and another related to mammoths. Columbian and woolly mammoths also showed evidence of interbreeding, likely following a latitudinal cline across North America. While hybridization events have shaped elephantid history in profound ways, isolation also appears to have played an important role. Our data reveal nearly complete isolation between the ancestors of the African forest and savanna elephants for ~500,000 y, providing compelling justification for the conservation of forest and savanna elephants as separate species.

paleogenomics | elephantid evolution | mammoth | admixture | species divergence

Members of the family Elephantidae, known as elephantids, first appeared in Africa 5 to 10 Mya and are the only surviving family of the order Proboscidea (1, 2). Although many fossil species have been identified, high levels of within-taxon variation have complicated the delineation of species boundaries (1–3). Living elephantids include two species of the genus *Loxodonta*, the forest elephant (*Loxodonta cyclotis*) and the savanna elephant (*Loxodonta africana*), which are restricted to Africa, and one of the genus *Elephas*, which is endemic to Asia (*Elephas maximus*). Extinct mammoths (genus *Mammuthus*)

comprise 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.

(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, 8.7% 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 elephantids ~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

Sample ID	Geographic origin	Date, y before present	Sequencing (source)	No. of mapped reads, million	Average coverage
L. cyclotis_A	Central African Republic	Modern	This study (BI)	906	27.78
L. africana_B	Kenya	Modern	This study (BI)	1,001	30.44
L. africana_C	South Africa	Modern	This study (BI)	1,114	33.42
E. maximus_D	Myanmar	Modern	This study (BI)	1,283	38.94
E. maximus_E	Malaysia (Borneo)	Modern	This study (BI)	1,107	32.20
L. cyclotis_F	Sierra Leone	Modern	This study (BI)	1,074	32.06
M. primigenius_G	Taimyr Peninsula, Russia	~31,500	This study (HMS)	55	0.60
M. primigenius_H	Alaska, USA	~44,900	This study (HMS)	27	0.49
M. americanum_I	Alaska, USA	>50,000	This study (IFT, HMS)	399	3.96
E. maximus_L	India*	Modern	(13)	889	27.02
E. maximus_M	India*	Modern	(13)	1,014	30.27
P. antiquus_N	Germany	~120,000	This study (BI, HMS)	1,399	14.64
P. antiquus_O	Germany	~120,000	(10)	12	0.14
M. primigenius_P	Oimyakon, Russia	~44,800	(12)	902	12.77
M. primigenius_Q	Wrangel Island, Russia	~4,300	(12)	959	19.00
M. primigenius_S	Yamal Peninsula, Russia	~45,300	This study (IFT, HMS)	132	0.91
M. columbi_U	Wyoming, USA	~13,400	This study (IFT, HMS)	122	1.53
Mammuthus_V	Wyoming, USA	~42,400	This study (IFT, HMS)	830	5.86
M. americanum_X	Gulf of Maine, USA	~13,400	This study (HMS)	71	0.79
E. maximus_Y	Assam, India	Modern	(13)	1,239	35.90
E. maximus_Z	Karnataka, India	Modern	(14)	447	14.58

BI, Broad Institute; HMS, Harvard Medical School; IFT, Illumina Fast Track Services.
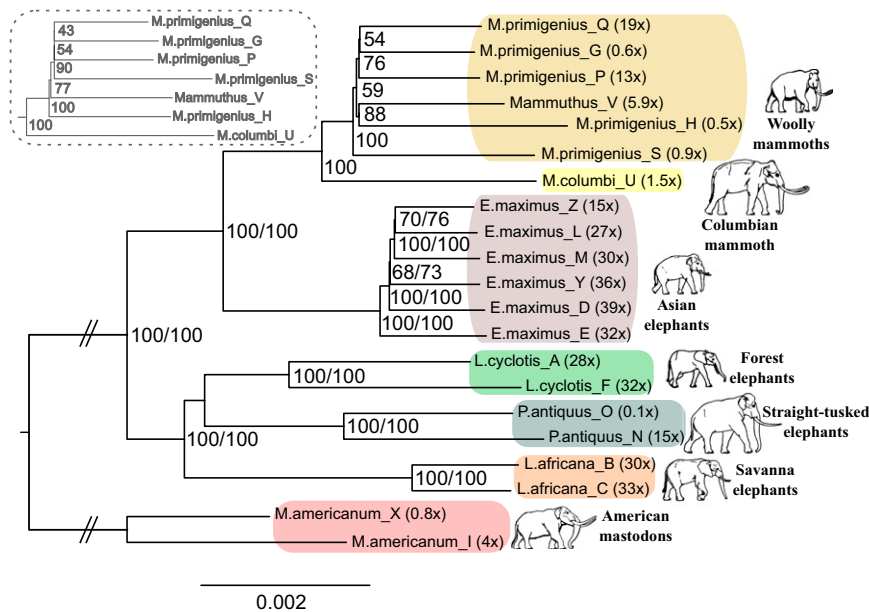*Exact geographic origin is unknown.

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. 2*A*).

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. 2*A* 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 they are 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. 2*A*). 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. 2*A*). 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. 2*A*), which could help to resolve an apparent discrepancy. While phylogenetic trees based on genome-wide 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 ($7 \leq |Z| \leq 9$) (Fig. 2*B*). 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. 2*A*). 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. 4*A*) 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 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
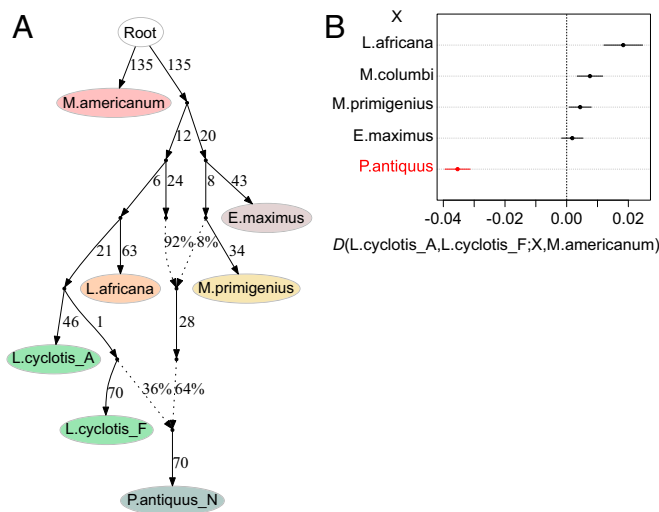
Fig. 2. Admixture graph of elephantid history and supporting *D*-statistics. (*A*) Model of the phylogenetic relationships among elephantids augmented with admixture events. Branch lengths are given in drift units × 1,000. Two admixture events are inferred in the history of the straight-tusked elephant lineage, from a population related to woolly mammoths and a population related to the West African forest elephant (*L. cyclotis_F*) while most of its ancestry derives from a lineage most closely related to the common ancestor of savanna and forest elephants. We were not able to resolve the order of the two admixture events. Inferred ancestry proportions are ~6 to 10% and 35 to 39% (with confidence intervals including uncertainty due to possible reference biases) (*SI Appendix*, Figs. S12.3 and S12.4) for the woolly mammoth-related and forest elephant-related components, respectively. (*B*) *D*-statistics testing for asymmetric genetic affinity between each of the two forest elephants and another elephantid (*X*). Positive values indicate excess genetic affinity between *L. cyclotis_A* 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.

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 each of the individuals within the woolly mammoth cluster. Instead, the Columbian mammoth shares significantly more alleles

with sympatric North American woolly mammoths than it does with any of the Eurasian woolly mammoths in our dataset (all $|Z|$-scores > 9.4) (*SI Appendix*, Table S11.3). We used an f_4-ratio test (18) to estimate the Columbian mammoth ancestry proportion to 8.8 to 11.7% (95.4% confidence interval) in *Mammuthus_V* from Wyoming and 4.4 to 8.7% in *M. primigenius_H* from Alaska (*SI Appendix*, Fig. S11.1 and Table S11.7). These data suggest a north–south cline in the proportion of Columbian admixture, with the Alaskan mammoth having less Columbian ancestry, consistent with the fact that the range of the Columbian mammoth was limited to more southern temperate regions within North America.

Lastly, we tested for evidence of admixture between the ancestors of forest and savanna elephants. Despite their high average pairwise nuclear sequence divergence (0.74%; which is higher than that between Asian elephants and mammoths) (*SI Appendix*, Table S8.1), the mitochondrial phylogeography of the two African elephant species indicates that hybridization between them must have occurred (22, 23). However, according to *D*-statistics, we found that the pairs of forest and savanna individuals in our study are mutually symmetrically related. This suggests that little, if any, gene flow has occurred subsequent to the splits of the pairs of sampled elephants from each species (609,000 y ago based on the oldest intraspecific split time estimated for the two forest elephants). Alternatively, gene flow from an unknown ancestral forest elephant lineage into both savanna elephant lineages and in equal proportions (or into the common ancestor of savanna elephants), or vice versa, could have occurred more recently. Hybridization in fact still occurs locally where the two species' ranges overlap (24–26). Recent work by Mondol et al. (27) shows that gene flow is bidirectional and that hybrids are fertile but appears to have not resulted in detectable introgression of nuclear alleles beyond these hybrid regions. The finding of deep population structure between the two subgroups of forest elephants (see *Within-Species Analyses: Diversity, Population Size Change, and Population Substructure* and Fig. 4*A*) and of isolation between forest and savanna elephants has implications for elephant conservation biology. While hybridization occurs between forest and savanna elephants along their current contact zone (24–27), which has long hindered their recognition as distinct species (28), our genome-wide analysis shows that this process has not left detectable traces on the genomes of representative members of the two species across their range. Thus, for conservation purposes, forest elephants and savanna elephants are appropriately viewed as reproductively distinct units, meeting the definition of the Biological Species Concept (29).

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.

Table 2. Additional *D*-statistics supporting the admixture graph in Fig. 2*A*

D-statistic	*D*	SE	*Z*	No. of transversions	Interpretation
Straight-tusked, forest; Asian, mastodon	0.076	0.004	17.94	371,372	Asian elephants share more alleles with straight-tusked elephants than with African elephants
Straight-tusked, savanna; Asian, mastodon	0.021	0.004	5.04	336,514	Asian elephants share more alleles with straight-tusked elephants than with African elephants
Straight-tusked, forest; woolly, mastodon	0.135	0.005	29.69	354,235	Mammoths share more alleles with straight-tusked elephants than with African elephants
Straight-tusked, savanna; woolly, mastodon	0.054	0.004	12.32	335,375	Mammoths share more alleles with straight-tusked elephants than with African elephants
Woolly, Asian; straight-tusked, mastodon	0.04	0.004	9.25	275,766	Straight-tusked elephants share more alleles with mammoths than with Asian elephants

15). 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_4-ratio tests (Fig. 2A 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^{-7}), and between straight-tusked and Asian elephants (1.32×10^{-7} to 5.71×10^{-7}), 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
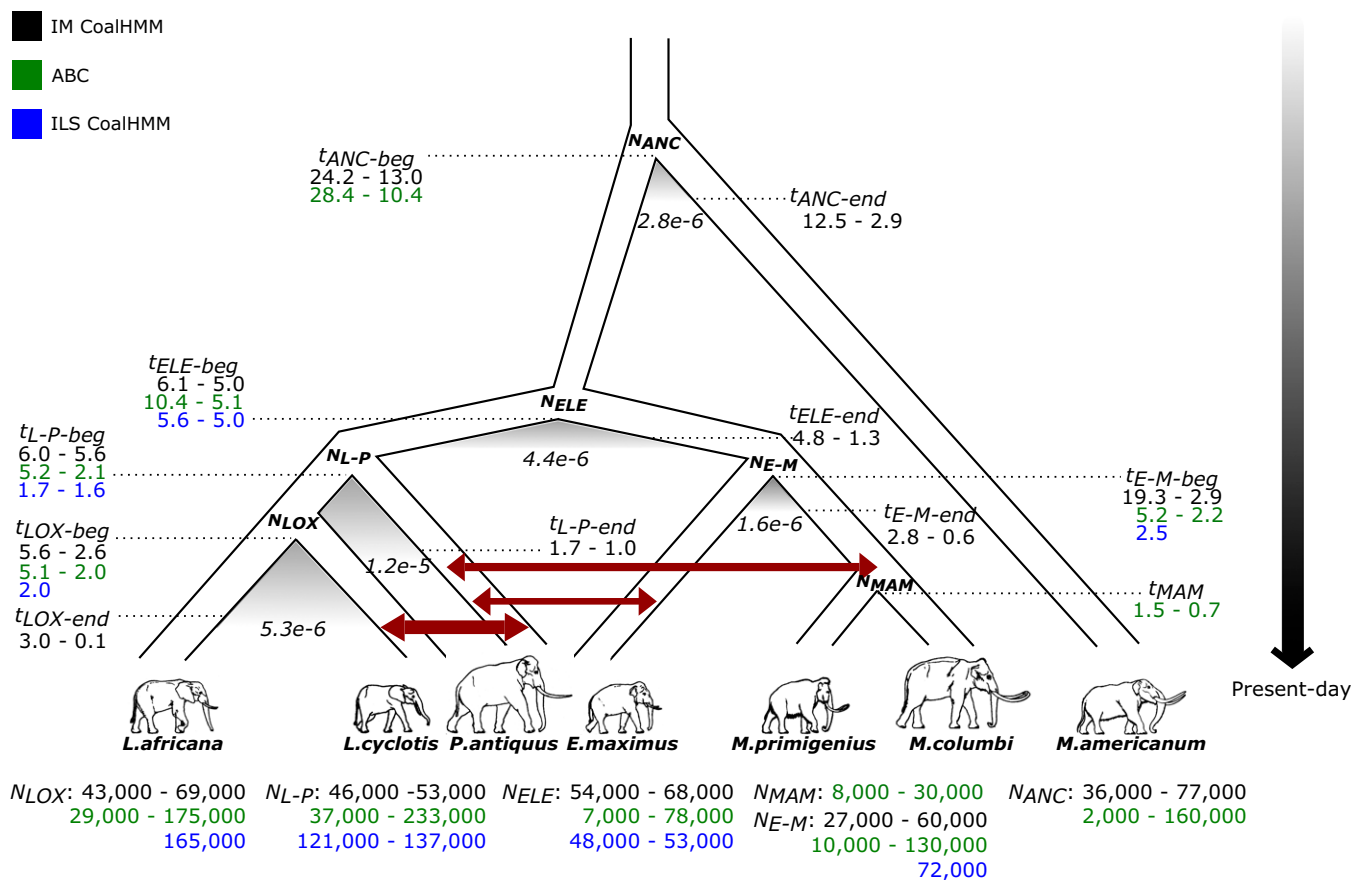

Fig. 3. A consensus demographic model for the history of elephantids. Inferred parameters from three modeling approaches are shown: (*i*) coalescent simulations with approximate Bayesian computation (ABC), (*ii*) incomplete lineage sorting analysis (ILS CoalHMM), and (*iii*) isolation-and-migration models (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_x) and split times (t_x) 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 $t_{x\text{-}beg}$ referring to the initial split time and $t_{x\text{-}end}$ 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 elephantids 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 N_e 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 N_e 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. 2*A* 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. 2*A*).

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 (CI: 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 (CI: 5.6 to 2.6 Mya) until 1.3 Mya (CI: 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 N_e 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 (0.00285 to 0.00364) (Fig. 4*B*) and sequence divergence (*SI Appendix*, Table S8.1) among extant and extinct elephantids (15, 40–42). Mammoths, straight-tusked elephants, and Asian elephants displayed intermediate levels of heterozygosity (0.00093 to 0.00167) (Fig. 4*B*), except for *E. maximus_E* from Malaysian Borneo, which had extremely low heterozygosity (0.00032). Savanna

elephants exhibited the lowest heterozygosity among all elephantids (0.00085 to 0.00088) (Fig. 4*B*).

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 (N_e) 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. 4*C*), in line with the ~21% lower heterozygosity in the former. The two savanna elephants had lower N_e relative to forest elephants for hundreds of thousands of years (Fig. 4*D*), 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. 4*C*), 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. 2*A*). Asian elephants are inferred to have gone through a phase of population growth, succeeded by decline ~120,000 y ago, resulting in a current N_e estimated to be about half that of savanna elephants (Fig. 4*E*). 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. 4*F*), 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 $F(A|B)$ 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. 4*A*). This is consistent with a hypothesis of deep population structure with limited gene flow, as well as with the high ancestral N_e 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 N_e curves (Fig. 4*D*), 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_E* and other individuals (190,000 to 103,000 y ago) (Fig. 4*A*), 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
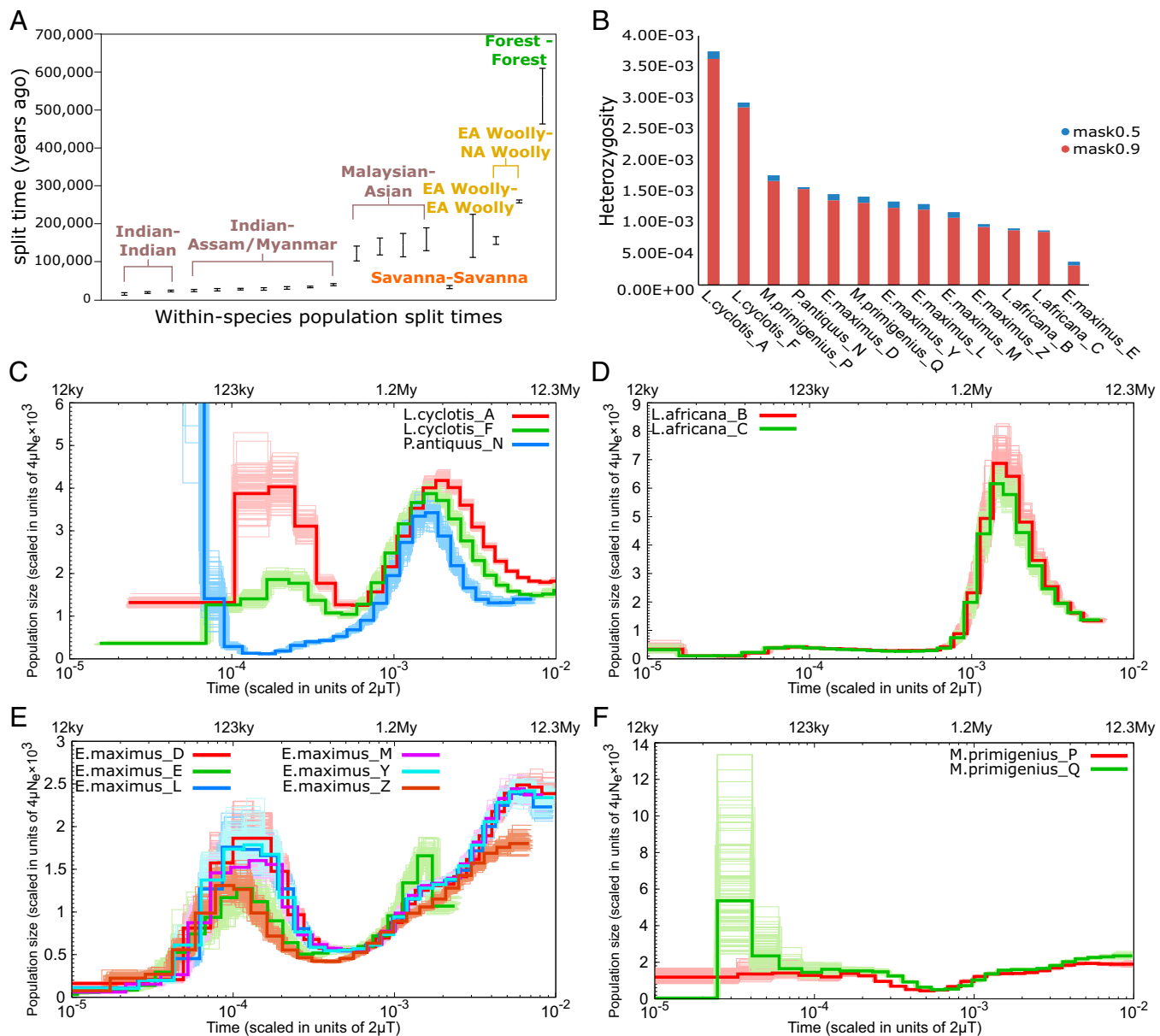
Fig. 4. Population size history, heterozygosity and within-species population split times. (*A*) Within-species population split time ranges (95.4% confidence intervals) as estimated from the $F(A|B)$ analysis, assuming a mutation rate (μ) of 0.406×10^{-9} per year and a generation time of 31 y. Confidence intervals of split times from reciprocal elephantid-pairs are combined and shown as a single interval. EA Woolly indicates the two Eurasian woolly mammoths (*M. primigenius_P* and *M. primigenius_Q*) and NA Woolly the North American (*Mammuthus_V*) woolly mammoth. (*B*) Individual autosomal heterozygosity estimated with the 90% mappability filter and the less stringent 50% mappability filter (see *SI Appendix*, Note 13 for more details). (*C–F*) PSMC inference of effective population size changes through time (bold curves) from high-coverage individual genomes of (*C*) forest and straight-tusked elephants, (*D*) savanna elephants, (*E*) Asian elephants, and (*F*) woolly mammoths. Bootstrap replicates are indicated by the soft-colored curves. Time is given in units of divergence per base on the lower *x* axis and in years before present on the upper *x* axis, assuming the mutation rate and generation time mentioned above. Population size is given in units of $4\mu N_e \times 10^3$ on the *y* axis.

woolly mammoths (*M. primigenius_P* and *M. primigenius_Q*; 225,000 to 112,000 y ago) (Fig. 4*A*).

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]; 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 elephantids as well. Our results in elephantids 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 elephantids: 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 elephantids, 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 elephantid 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 transversions. A neighbor-joining tree with support values from 100 bootstrap analyses was built from the resulting matrix with PHYLIP (58) version 3.696.

Admixture Analyses. To test for signals of gene flow within and between species, we computed *D*-statistics (18) with the population genomics program POPSTATS (https://github.com/pontussk/popstats), which performs computations as in refs. 17 and 18, and estimates SEs using a block jackknife procedure by splitting chromosomes into 5-Mb blocks and weighting blocks by the number of polymorphic positions. Admixture signals detected from *D*-statistics were further integrated into a single admixture graph (phylogenetic tree augmented with admixture events) using *qpGraph* (18), estimating branch lengths and mixture proportions. Mixture proportions were also inferred from f_4-ratios (18) computed with the software POPSTATS.

These analyses were based on transversion SNPs (called from randomly sampled alleles per site) to alleviate biases from residual postmortem damage in CpG context and recurrent mutations.

Interspecies Demographic Inference. Three modeling approaches were implemented to infer species ancestral effective population sizes, split times, and migration rates: (*i*) coalescent simulations with approximate Bayesian computation (ABC), (*ii*) incomplete lineage sorting (ILS CoalHMM), and (*iii*) isolation with migration CoalHMM models (IM CoalHMM). For the first approach, demographic scenarios of three elephantid lineages and the mastodon (outgroup) were modeled in *scrm* (59), using prior distributions for all demographic parameters. The ABC package (60) in R (R Development Core Team 2011) was used to fit parameters based on the following summary statistics: allelic states of pairs of adjacent variable sites (30), *D*-statistics, and pairwise divergence per base pair. For the second approach, CoalHMM isolation models (32) (without gene flow) were used to estimate proportions of ILS along alignments of three elephantid lineages, and to infer in parallel unbiased estimates of effective population size and split time parameters, as described in ref. 32. For the third approach, the isolation and the isolation-with-initial-migration (39) CoalHMM models were fitted to pairwise interspecies sequence alignments. The Akaike information criterion (AIC) was used to choose the preferred model and maximum likelihood estimates of ancestral effective population sizes, split times, start and end of migration period, and migration rates were obtained. Parameter estimates were converted to years, assuming a mutation rate of 0.406×10^{-9} per base per year (as calculated in *SI Appendix, Note 16*; but we caution there is substantial uncertainty in this estimate) and a generation interval of 31 y (as in ref. 15). For more details, see *SI Appendix, Notes 16–18*.

Within-Species Demographic Analyses. Individual heterozygosity was estimated for high-coverage genome sequences with mlRho (61) v.2.7. The PSMC (43) was used to reconstruct changes in effective population size through time by examining patterns of heterozygosity across the diploid genome of single individuals. Within-species population split times were estimated using the $F(A|B)$ 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.

1. Shoshani J (1998) Understanding proboscidean evolution: A formidable task. *Trends Ecol Evol* 13:480–487.
2. Maglio V (1973) Origin and evolution of the Elephantidae. *Trans Am Philos Soc* 63:1–149.
3. Todd NE (2010) New phylogenetic analysis of the family Elephantidae based on cranial-dental morphology. *Anat Rec (Hoboken)* 293:74–90.
4. Vartanov SL, Arslanov KA, Karhu JA, Possnert G, Sulerzhitsky LD (2008) Collection of radiocarbon dates on the mammoths (*Mammuthus primigenius*) and other genera of Wrangel Island, northeast Siberia, Russia. *Quat Res* 70:51–59.
5. Veltre DW, Yesner DR, Crossen KJ, Graham RW, Coltrain JB (2008) Patterns of faunal extinction and paleoclimatic change from mid-Holocene mammoth and polar bear remains, Pribilof Islands, Alaska. *Quat Res* 70:40–50.
6. Enk J, et al. (2011) Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. *Genome Biol* 12:R51.
7. Lister AM, Sher AV (2015) Evolution and dispersal of mammoths across the Northern Hemisphere. *Science* 350:805–809.
8. Stuart AJ (2005) The extinction of woolly mammoth (Mammuthus primigenius) and straight-tusked elephant (Palaeoloxodon antiquus) in Europe. *Quat Int* 126–128:171–177.
9. Shoshani J, et al. (2007) Relationships within the Elephantinae using hyoid characters. *Quat Int* 169–170:174–185.
10. Meyer M, et al. (2017) Palaeogenomes of Eurasian straight-tusked elephants challenge the current view of elephant evolution. *eLife* 6:e25413.
11. Shoshani J, et al. (2006) A proboscidean from the late Oligocene of Eritrea, a "missing link" between early Elephantiformes and Elephantimorpha, and biogeographic implications. *Proc Natl Acad Sci USA* 103:17296–17301.
12. Palkopoulou E, et al. (2015) Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr Biol* 25:1395–1400.

13. Lynch VJ, et al. (2015) Elephantid genomes reveal the molecular bases of woolly mammoth adaptations to the Arctic. *Cell Rep* 12:217–228.
14. Reddy PC, et al. (2015) Comparative sequence analyses of genome and transcriptome reveal novel transcripts and variants in the Asian elephant *Elephas maximus. J Biosci* 40:891–907.
15. Rohland N, et al. (2010) Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. *PLoS Biol* 8:e1000564.
16. Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature* 461:489–494.
17. Green RE, et al. (2010) A draft sequence of the Neandertal genome. *Science* 328: 710–722.
18. Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192:1065–1093.
19. Sanders WJGE, Harris JM, Saegusa H, Delmer C (2010) Proboscidea. *Cenozoic Mammals of Africa*, eds Werdelin L, Sanders WJ (Univ of California Press, Berkeley, CA).
20. Parmentier I, et al. (2007) The odd man out? Might climate explain the lower tree α-diversity of African rain forests relative to Amazonian rain forests? *J Ecol* 95: 1058–1071.
21. Enk J, et al. (2016) *Mammuthus* population dynamics in Late Pleistocene North America: Divergence, phylogeography, and introgression. *Front Ecol Evol* 4:42.
22. Roca AL, Georgiadis N, O'Brien SJ (2005) Cytonuclear genomic dissociation in African elephant species. *Nat Genet* 37:96–100.
23. Ishida Y, Georgiadis NJ, Hondo T, Roca AL (2013) Triangulating the provenance of African elephants using mitochondrial DNA. *Evol Appl* 6:253–265.
24. Backhaus D (1958) Zur Variabilität der äusseren systematischen Merkmale des afrikanischen Elefanten (*Loxodonta* Cuvier, 1825), Säugetierkd. *Mitt* 6:166–173.
25. Debruyne R (2005) A case study of apparent conflict between molecular phylogenies: The interrelationships of African elephants. *Cladistics* 21:31–50.
26. Johnson MB, et al. (2007) Complex phylogeographic history of central African forest elephants and its implications for taxonomy. *BMC Evol Biol* 7:244.
27. Mondol S, et al. (2015) New evidence for hybrid zones of forest and savanna elephants in Central and West Africa. *Mol Ecol* 24:6134–6147.
28. Blanc J (2008) *Loxodonta africana. The IUCN Red List of Threatened Species 2008* (International Union for Conservation of Nature, Gland, Switzerland), e. T12392A3339343.
29. Mayr E (1942) *Systematics and the Origin of Species, from the Viewpoint of a Zoologist* (Harvard Univ Press, Cambridge, MA).
30. Rasmussen M, et al. (2011) An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334:94–98.
31. Agenbroad LD (2005) North American Proboscideans: Mammoths: The state of knowledge, 2003. *Quat Int* 126–128:73–92.
32. Dutheil JY, et al. (2009) Ancestral population genomics: The coalescent hidden Markov model approach. *Genetics* 183:259–274.
33. Rohland N, et al. (2007) Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biol* 5:e207.
34. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2:e190.
35. Roca AL, et al. (2015) Elephant natural history: A genomic perspective. *Annu Rev Anim Biosci* 3:139–167.
36. Brandt AL, Ishida Y, Georgiadis NJ, Roca AL (2012) Forest elephant mitochondrial genomes reveal that elephantid diversification in Africa tracked climate transitions. *Mol Ecol* 21:1175–1189.
37. Aquadro CF, Begun DJ, Kindahl EC (1994) Selection, recombination, and DNA polymorphism in Drosophila. *Non-Neutral Evolution: Theories and Molecular Data*, ed Golding B (Springer, Boston), pp 46–56.
38. Laporte V, Charlesworth B (2002) Effective population size and population subdivision in demographically structured populations. *Genetics* 162:501–519.
39. Mailund T, et al. (2012) A new isolation with migration model along complete genomes infers very different divergence processes among closely related great ape species. *PLoS Genet* 8:e1003125.
40. Roca AL, Georgiadis N, Pecon-Slattery J, O'Brien SJ (2001) Genetic evidence for two species of elephant in Africa. *Science* 293:1473–1477.
41. Comstock KE, et al. (2002) Patterns of molecular genetic variation among African elephant populations. *Mol Ecol* 11:2489–2498.
42. Ishida Y, et al. (2011) Reconciling apparent conflicts between mitochondrial and nuclear phylogenies in African elephants. *PLoS One* 6:e20642.
43. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.
44. Sukumar R (2003) *The Living Elephants: Evolutionary Ecology, Behavior, and Conservation* (Oxford Univ Press, Oxford).
45. Fleischer RC, Perry EA, Muralidharan K, Stevens EE, Wemmer CM (2001) Phylogeography of the Asian elephant (*Elephas maximus*) based on mitochondrial DNA. *Evolution* 55: 1882–1892.
46. Vidya TNC, Sukumar R, Melnick DJ (2009) Range-wide mtDNA phylogeography yields insights into the origins of Asian elephants. *Proc Biol Sci* 276:893–902.
47. Cahill JA, et al. (2015) Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Mol Ecol* 24:1205–1217.
48. vonHoldt BM, Kays R, Pollinger JP, Wayne RK (2016) Admixture mapping identifies introgressed genomic regions in North American canids. *Mol Ecol* 25:2443–2453.
49. Koepfli K-P, et al. (2015) Genome-wide evidence reveals that African and Eurasian golden jackals are distinct species. *Curr Biol* 25:2158–2165.
50. de Manuel M, et al. (2016) Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* 354:477–481.
51. Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D (2006) Genetic evidence for complex speciation of humans and chimpanzees. *Nature* 441:1103–1108.
52. Figueiró HV, et al. (2017) Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Sci Adv* 3:e1700299.
53. Lister AM (2013) The role of behaviour in adaptive morphological evolution of African proboscideans. *Nature* 500:331–334.
54. Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010:pdb. prot5448.
55. Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 40:e3.
56. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
57. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
58. Felsenstein J (1989) PHYLIP–Phylogeny Inference Package (version 3.2). *Cladistics* 5: 164–166.
59. Staab PR, Zhu S, Metzler D, Lunter G (2015) scrm: Efficiently simulating long sequences using the approximated coalescent with recombination. *Bioinformatics* 31: 1680–1682.
60. Csillery K, Francois O, Blum MGB (2012) abc: An R package for approximate Bayesian computation (ABC). *Methods Ecol Evol* 3:475–479.
61. Haubold B, Pfaffelhuber P, Lynch M (2010) mlRho–A program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. *Mol Ecol* 19:277–284.