Genome-based sexing provides clues about behavior and social structure in the woolly mammoth

Patrícia Pečnerová¹²*, David Díez-del-Molino¹, Nicolas Dussex¹, Tatiana Feuerborn¹³, Johanna von Seth¹², Johannes van der Plicht⁴⁵, Pavel Nikolskiy⁶, Alexei Tikhonov⁷⁸, Sergey Vartanyan⁹, Love Dalén¹⁰*

1. Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Box 50007, SE-10405 Stockholm, Sweden
2. Department of Zoology, Stockholm University, SE-10691 Stockholm, Sweden
3. Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5–7, 1350 Copenhagen K, Denmark
4. Centre for Isotope Research, Groningen University, Nijenborgh 4, 9747 AG Groningen, The Netherlands
5. Faculty of Archaeology, Leiden University, Postbus 9514, 2300 RA Leiden, The Netherlands
6. Geological Institute of the Russian Academy of Sciences, Pyzhevsky pereulok 7, Moscow 119017, Russia
7. Zoological Institute of Russian Academy of Sciences, Universitetskaya nab. 1, 199034 Saint-Petersburg, Russia
8. Institute of the Applied Ecology of the North, North-Eastern federal University, Lenina 1, 677000 Yakutsk, Russia
9. North-East Interdisciplinary Scientific Research Institute N.A.N.A. Shilo, Far East Branch, Russian Academy of Sciences (NEISRI FEB RAS), Magadan, Russia
10. Lead Contact

* Corresponding authors

Summary

While present day taxa are valuable proxies for understanding the biology of extinct species, it is also crucial to examine physical remains in order to obtain a more comprehensive view of their behaviour, social structure, and life histories [1, 2]. For example, information on
demographic parameters such as age distribution and sex-ratios in fossil assemblages can be used to accurately infer socioecological patterns (e.g., [3]). Here we use genomic data to determine the sex of 98 woolly mammoth (*Mammuthus primigenius*) specimens in order to infer social and behavioural patterns in the last 60,000 years of the species’ existence. We report a significant excess of males among the identified samples (69% vs 31%; P<0.0002). We argue that this male bias among mammoth remains is best explained by males more often being caught in natural traps that favour preservation. We hypothesize that this is a consequence of social structure in proboscideans, which is characterized by matriarchal hierarchy and sex segregation. Without the experience associated with living in a matriarchal family group, or a bachelor group with an experienced bull, young or solitary males may have been more prone to die in natural traps where good preservation is more likely.

### Keywords

The woolly mammoth, *Mammuthus primigenius*, sex ratio, sex bias, behavior

### Results and Discussion

#### Sampling and sexing

To investigate the sex of the mammoth remains, we generated low-coverage genomic data from 83 bone, tooth, and tusk samples collected at various locations throughout Siberia (Figure 1, Table S1). The samples mostly comprise of individual fragments found in river basins, along coastlines and lake shores where they have been redeposited after erosion from permafrost sediments. DNA was extracted using a silica-based method [5, 6], converted into indexed libraries [7] and sequenced on an Illumina HiSeq 2500 platform. Additionally, we included previously published whole-genome shotgun data from mammoth hair shafts [8, 9] to generate a final dataset of 98 mammoth samples. Sequence reads were mapped against the genome assembly of the African savannah elephant (*Loxodonta africana*). The number of reads mapping to chromosome X and 8, respectively, were used to determine the sex of each specimen (for details, see Methods). In total, 66 specimens were identified as males and 29 as females (Figure 2).
Causes for a biased sex ratio

All samples were collected opportunistically and do not originate from fossil assemblages, and can thus be considered a random sample of the available fossil record. In the absence of other factors, this sampling scheme would be expected to yield a sex ratio equal to the natal sex ratio, which is usually balanced in mammal populations [10]. Furthermore, the natal sex ratios in both the wild Asian elephant (Elephas maximus) and the African savannah elephant are close to 1:1 [11], suggesting that the natal sex ratio was likely balanced also in the woolly mammoth.

We find a role of sexual dimorphism unlikely in explaining the observed skew in sex ratio. In sexually dimorphic species, taphonomic processes such as scavenging, decomposition, and erosion can lead to differential preservation of male and female remains. Indeed, the size of skeletal elements affects degradation processes, with large elements disappearing more slowly than smaller ones [12]. However, in large megafaunal species such as mammoths, fossil preservation sex-biases are not common [13], and are especially unlikely when the remains, as in our case, have been recovered from permafrost where preservation should be facilitated.

Several of the best preserved woolly mammoths discovered so far are believed to have died in natural traps such as falling through thin ice (e.g., the Berezovka mammoth; [14]) or getting caught in a mudflow or drowning in pools (e.g., Lyuba and Khroma calves, respectively; [15]). Similarly, characteristic accumulations of mammoth and mastodon remains have been observed at sites representing natural traps such as a kettle-hole in Condover, England [17, 18] and Hot Springs, South Dakota [Agenbroad, 1984], in addition to non-natural trap-sites which have multi-individual accumulations such as Big Bone Lick in Kentucky [16] and the catastrophic accumulation in Waco, Texas [19]. Morphological studies of the mammoth remains found at Hot Springs have shown that the sex ratio was heavily skewed, with 13 young adult males (10 to 30 years of age) and a single female [20]. In fact, across Eurasia the remains from isolated individuals found in natural traps, such as sinkholes and crevasses, largely represent those of males rather than females [21].

Even though the samples analysed in this study do not have a direct association with natural traps, the vast plains of northeastern Siberia are known to have been full of “taphonomic traps” such as gullies, crevices, and sinkholes that formed in the permafrost [22]. Passing over
fragile ice, landslides on river banks, mud flows, and sinkholes in walls formed by ice veins were some of the traps in the mammoth steppe landscape [22]. Therefore, one possible explanation for the skewed sex ratio we observe in our samples could be that samples preserved for thousands of years in permafrost represent, to a disproportionate extent, male mammoths that have died in these types of natural traps. Taphonomic processes may subsequently have facilitated preservation of remains in these traps. However, why males would die more frequently than females in such traps remains to be answered.

Male-biased dispersal is considered the norm in mammals [23], including extant elephant species [24, 25]. Dispersal is stimulated by various factors, such as reduction in competition for mates and resources [26] or inbreeding avoidance [27], and is usually highly related to the social structure of a species [28]. In ungulates, various hypotheses have been proposed to explain the spatial segregation of sexes [29], and a study on free-ranging African savannah elephants in Botswana suggests that differences in habitat use between elephant bull groups and family units are responsible for the segregation [30]. While female movement is limited by the presence of offspring, males are able to move further and explore more remote patches of vegetation. Moreover, ranging behavior of elephant males is influenced by musth. Males in musth can travel over long distances seeking out receptive females [31, 32]. Males that are not in musth might disperse to considerable distance to avoid bulls in musth that express aggressive behavior due to high hormone levels [30]. Dispersal represents a considerable cost in fitness as it is sometimes associated with a higher mortality risk [33]. Mortality resulting from sex-biased dispersal should therefore be reduced when dispersal possibilities are limited.

To test this hypothesis, we compared the sex ratios in samples collected from the Siberian mainland (n=46), and from Wrangel Island (n=49), where a population of woolly mammoths survived in isolation for over 6,000 years after the rise of sea levels at the end of the last glaciation. While long-distance dispersal likely occurred on the Siberian mainland [34], dispersal must have been more limited on the comparatively small Wrangel Island (i.e., 7,600 km²), where most of the area consists of mountains, rocks, ice and snow fields, i.e., habitat not suitable for mammoths. The ranging behavior of extant elephants is influenced by various environmental and social factors and data from present elephant populations show that proboscidean home range sizes can vary by orders of magnitude, e.g., insular Asian elephant populations from Sri Lanka have a home range of ~60 km², while Asian elephant populations from mainland southern India have home range ten times larger [35], suggesting that the woolly mammoth home ranges on Wrangel Island might have been smaller than those on the
Siberian mainland. Considering that dispersal distance is proportional to the home range size [36], we assume that male dispersal on Wrangel Island was more restricted. Thus, if sex-biased dispersal alone led males to be more often caught in natural traps, we would have expected a less skewed sex ratio in the Wrangel Island remains. However, we did not observe a difference in the sex ratios between these two groups of samples (Figure 3; p>0.05), suggesting that sex-biased mortality during dispersal cannot solely explain the skewed sex ratio in mammoth fossil deposits.

Similar to other proboscideans, woolly mammoths are thought to have lived in sex-segregated herds centred around a matriarchal group consisting of a dominant female and her offspring and solitary or loosely associated males [37]. Sex-segregation has been observed in fossil trackways from the Miocene, suggesting that this social structure may be an ancestral feature of proboscideans [38]. Moreover, evidence from fossil deposits [14, 39] supports the assumption that mammoth social structure was very similar to the structure of extant elephant social groups and a mammoth herd likely comprised of a small number of adult females and juveniles. Upon reaching maturity at around 13-15 years of age, males dispersed from their natal family unit. Depending on their age and sexual state, adult males probably spent time alone or in small groups of other males in particular bull areas [40]. These bachelor groups typically included individuals of multiple age ranges that could utilize habitats too marginal or poor in resources to be used by family groups [41]. In 1980s, Agenbroad and Mead [42, 43] formulated a hypothesis that the male-biased sex ratio in the Hot Springs assemblage can be explained by the lack of experience in young males and lack of assistance from conspecifics in solitary males of all age groups. Without the experience associated with the matriarchal family group or more experienced bulls within a bachelor group, young or solitary males unfamiliar with their environment may have been especially vulnerable and likely to enter unfamiliar terrain or take higher risks when dispersing [21, 44]. As a consequence, males may have had an increased risk of falling into sinkholes and through the ice of lakes and rivers, as well as ending up in mud flows or landslides [20]. Quick deposition of such remains would have led to exceptional preservation of these ill-fated individuals until the present day. While this hypothesis has been proposed to explain the male bias in fossil assemblages like Hot Springs, we also hypothesize that most fossil remains found opportunistically as re-deposited elements originate from individuals that, because of their behavior, died in a way that ensured good preservation.
These results might have wider ramifications for other studies of mammoth biology. For example, previous estimates of body size based on the size of long bones or molar teeth could be biased if even sex ratios were assumed. Similarly, diet analyses, for example using stable isotopes, might need to take into account that most samples originate from male specimens. Importantly, the genome-based sexing method presented here provides new opportunities for more detailed studies in the future, such as exploring sex-specific differences in body size, diet and other life history parameters.

**Sex bias in other extinct megafauna**

Assuming that social structure can lead to this type of sex bias in fossil remains, we predict that other Pleistocene fauna that lived in equivalent female-dominated social groups would show a similar pattern. For example, palaeontological sites such as Rancho La Brea and McKittrick tar pits contain various species of megafauna that were accumulated over thousands of years as individuals became trapped in the tar [45]. Within the deposits at La Brea, remains from the now extinct wild horse, *Equus occidentalis*, follow the predicted pattern and are predominantly comprised of subadult males [46].

On the other hand, the La Brea Tar Pits also contain approximately 300 bison (*Bison antiquus*) individuals with females twice as abundant as males [46]. One possible explanation for this seemingly contradictory pattern is offered by bison social structure and migration patterns. Bison (*Bison bison*) form “mixed” groups consisting of calves, young males and females of all ages, which cluster during spring forming larger herds [47]. Individual age identification of the La Brea *Bison antiquus* remains suggests that groups of adult females and calves passed through the region in late spring when the asphalt became sticky [48]. Although speculative, it is therefore possible that the female-biased skew in sex ratio of La Brea bison remains reflects the abundance of female-dominated social groups that visited the region during seasonal migrations. This example implies that despite the expectation of strong male-biased sex ratios in remains from taxa with matriarchal social structure, other factors of a species’ biology need to be taken into consideration. Moreover, this result may be a consequence of La Brea being a single site. However, among randomly collected steppe bison (*Bison priscus*) samples from the permafrost region, we predict that future studies will identify a male-biased sex ratio due to the species’ social structure, where females and young stay together whereas adult males are more solitary or live in small temporary groups [49].
Examining the effect of social behavior on sex ratio in fossil remains can become particularly illustrative when two related species with different social structure are compared. For instance, the even sex ratio in fossil assemblages of the extinct *Aphelops* rhinoceros [50] is consistent with the solitary behavior of all extant Rhinocerotidae [51]. However, the extinct rhinoceros *Teleoceras*, which may have lived in female-dominated herds, and may have formed bachelor groups [50], shows a strong male bias in fossil assemblages [50, 52].

Our results demonstrate the utility of isolated and fragmented fossil remains for reconstructing the socioecology and behavior of extinct taxa. This approach makes use of easily accessible data and it has wider application in palaeontology. Although our data warrants a cautious interpretation since only tentative conclusions can be drawn, combining fossil and genomic data marks an important step in the study of sociobiology of extinct megafauna.
Author Contributions

Acknowledgments

The authors would like to thank Pontus Skoglund for advice on molecular sexing and Tom Gilbert and Fátima Sánchez Barreiro from the Natural History Museum of Denmark in Copenhagen for helpful discussions during the writing of the manuscript. The authors would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, and UPPMAX (project number: b2015028) for providing assistance in massive parallel sequencing and computational infrastructure. P.P. would like to acknowledge bioinformatic training within the Swedish Bioinformatic Advisory Programme. Genetic analyses were funded through a grant from the Swedish Research Council (VR grant 2012-3869). J.v.S. and T.F. acknowledge support from FORMAS (project 2015-676) and the EU-funded ITN project ArchSci2020, respectively. N.D. acknowledges support from the Swiss Research Council.
References


Figure legends

Figure 1. Map showing locations of sample localities. Numbers within circles show the number of samples collected in more densely sampled regions, from left to right: Taimyr Peninsula, New Siberian Islands, Chaun Bay area, and Wrangel Island. Three mainland samples with unknown locations are not shown. See also Table S1.

Figure 2. Determination of sex based on a comparison of the number of reads mapping to chromosome X and to chromosome 8. Gray areas depict the ranges where most of samples clustered within the male and female categories. Dotted lines show confidence intervals. Orange dots represent samples assigned as males while blue dots represent females. Labels on the left side are sample IDs and labels on the right side show sample radiocarbon dates. See also Table S1.

Figure 3. Number of males and females in the two compared localities – Pleistocene mainland Siberia and Wrangel Island, where the latter includes only samples radiocarbon dated to after the island was formed (i.e., less than 10,500 years before present). Sex ratio deviation from parity was significant for the overall sample (P<0.0002). When tested separately, the deviation from parity was significant for the Wrangel samples (P<0.002), and was very close to being significant for the mainland Siberian samples (P=0.054). The difference in sex ratios between Wrangel Island and mainland Siberia was not significant (P=0.504). See also Table S1.
STAR Methods

Contact for Reagent and Resource Sharing

Further information and requests for reagents may be directed to, and will be fulfilled by the Lead Contact, Love Dalén (love.dalen@nrm.se).

Experimental Model and Subject Details

Woolly mammoth samples analyzed in this study consist of fragments of bones, teeth, and tusks that were collected opportunistically in various locations throughout Siberia (Table S1). Majority of samples represents isolated elements found in river beds after they had been eroded from permafrost. The character of the data does not allow morphological identification of sex or age-at-death. Samples were radiocarbon-dated and are reported in conventional radiocarbon years (BP), which includes correction for isotopic fractionation and usage of the conventional half-life [53]. The 14C dates are calibrated into calendar ages using the recommended calibration curve IntCal13 [54] using the program OxCal 4.2 [55]. Medians of the calibrated dates are reported in calBP, i.e. calendar years relative to 1950 AD.

Method Details

DNA Extraction

All pre-amplification steps were carried out in a clean ancient DNA facility at the Swedish Museum of Natural History. Contamination was prevented by using protective suits, gloves, and face masks; by regular bleaching of surfaces and UV-irradiation of tools; and by using negative controls during all extraction and library-building steps. Bone powder was obtained using a hand-held Dremel drill and DNA was extracted from the powder following a modified version of a silica-based protocol [5, 6]. Approximately 50 mg of bone powder were incubated overnight under motion in 715 µl of extraction buffer (0.45M EDTA, 0.1M UREA, 150 µg proteinase K). Following digestion, the DNA is concentrated on a silica membrane of a 30K MWCO Vivaspin filter (Sartorius) by centrifugation at 2,300 rpm. Purification and elution of extracted DNA is performed using standard QIAquick PCR Purification Kit (Qiagen).
Library Preparation

Multiplexed, paired-end, Illumina libraries were prepared from 20 μl of DNA extract following an established protocol [7] using uracil-treatment with the USER enzyme (New England Biolabs).

The first step of the library build is the blunt-end repair with a reaction mix consisting of following: 1x Buffer Tango, dNTP (100 μM each), 1 mM ATP, 0.15 U/μl USER enzyme, and 0.5 U/μl T4 PNK. After 3-hour incubation in a thermocycler at 37°C, 0.1 U/μl T4 Polymerase was added and the library was further incubated for 15 min at 25°C followed by 5 min at 12°C. After purification with the MinElute purification kit (Qiagen) and elution in 22 μl of EB buffer, adapter ligation was performed using a following reaction mix: 1x T4 ligation buffer, 5% PEG-4000, 0.125 U/μL T4 ligase, and an adapter mix of P7 and P5 adapters 2.5 μM each [7]. Libraries were incubated for 30 minutes at 22°C and again cleaned using MinElute purification kit (Qiagen). Finally, adapter fill-in was performed using a reaction mix that consisted of: 1x Thermopol buffer, dNTP (250 μM each), and 0.3 U/μl Bst polymerase. After incubation at 37°C for 20 minutes, the final heatkill was performed by incubation at 80°C for 20 minutes.

Each library was indexed and amplified from 3 μl of library template using a following reaction mix: 0.05U/μl AccuPrime™ Pfx DNA Polymerase (Life Technologies), 2.5 μl of AccuPrime™ reaction mix, 200 nM of IS4 primer [7], and 200 nM of indexing primer [7]. Libraries were amplified under following conditions: 95°C for 2 min; between 8 and 14 cycles (depending on quality) of: 95°C for 15 seconds, 60°C for 30 seconds. Amplified libraries were purified along with size selection using Agencourt AMPure XP beads (Beckman Coulter). Library concentrations were measured with a high-sensitivity DNA chip on a Bioanalyzer 2100 (Agilent). Multiplexed libraries were pooled in several separate pools in equimolar concentrations and sequenced using the Illumina HiSeq2500 technology.

Quantification and Statistical Analysis

Data Processing

SeqPrep 1.1 (available at https://github.com/jstjohn/SeqPrep) was used to trim adapters and to merge paired-end reads, using default settings and a minor modification to the source code allowing choosing the best quality scores of bases in the merged region instead of aggregating the scores. Mapping was performed using BWA 0.7.8 [56] and the alignments were processed.
using Samtools 0.1.19 [57]. Sequencing reads were mapped to a merged nuclear-
mitochondrial reference consisting of the African savanna elephant nuclear genome
(LoxAfr4) generated by the Broad Institute, and a mammoth mitogenome (Krause; GenBank
accession no. DQ188829). BWA aln algorithm designed for short Illumina reads was used for
the mapping, applying slightly modified default settings with deactivated seeding (-l 16500),
allowing more substitutions (-n 0.01) and allowing up to two gaps (-o 2). Alignments were
processed in SAMtools 0.1.19, including converting the alignments in SAM format to BAM
format, coordinate sorting, indexing, and removing duplicates. A mapping quality filter of
MQ=30 was applied and both types of files, before and after filtering, were used in the sexing.

Sexing

Since the African savannah elephant genome originates from a female individual, reference
for the chromosome Y was not available. Instead, we made use of the chromosome-level
LoxAfr4 assembly and compared the number of reads mapping to an autosome compared to
sex chromosome. Specifically, we compared the number of reads mapping to chromosome 8
and chromosome X, which are of comparable sizes. The number of mapped reads was
normalized by the length of the chromosome sequence.

Specimens belonging to males were expected to have about 50% of reads mapping to
chromosome X compared to chromosome 8, because while female mammoths have two
copies of chromosome X, male mammoths only have a single copy. Female specimens were
expected to have a comparable number of reads mapping to both chromosomes.

Confidence Intervals

We estimated upper and lower confidence intervals to identify sample sex from the
normalized chromosome X/chromosome 8 ratios by calculating the standard deviation (SD)
on the ratios of all the samples that could be unambiguously assigned as males (ratio <0.6)
and females (ratio >0.8) (Figure 2, grey areas). Then, the upper limit of the ratio to identify a
sample as male was obtained by adding 3 times the male SD to the male sample with the
highest ratio, and the lower ratio for females, by subtracting 3 times the female SD to the
female sample with lowest ratio (Figure 2, dotted lines). Out of the total dataset of 98
specimens, 95 fell into one of the two categories - male or female - while three samples could
not be determined with confidence.
In order to test if the observed sex ratio differed from the null expectation of a 1:1 sex ratio, we used a two-tailed binomial test applied to the dataset of sex-determined samples (N=95) as well as to Wrangel Island (N=49) and to mainland Siberia (N=46) datasets separately. Sex ratio deviation from parity was significant for the overall sample (P<0.0002). When tested separately, the deviation from parity was significant for the Wrangel samples (P<0.002), and marginally significant for the mainland Siberian samples (P=0.054).

To test for differences in sex ratios between Wrangel Island and mainland Siberia, we used a two-tailed Fisher’s exact test (available at https://www.graphpad.com/quickcalcs/contingency1.cfm), which showed that sex ratios on Wrangel and the mainland did not significantly differ from each other (P=0.504).

The accession number for the bam files containing chromosomes 8 and X of 83 new woolly mammoth samples reported in this paper is ENA: PRJEB22575.