# Excess of genomic defects in a woolly mammoth on Wrangel island

Rebekah L. Rogers¤*, Montgomery Slatkin

Department of Integrative Biology, University of California, Berkeley, Berkeley, California, United States of America

¤ Current Address: Department of Bioinformatics, University of North Carolina, Charlotte, Charlotte, North Carolina, United States of America
* rebekah.rogers@uncc.edu
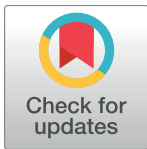
Abstract

Woolly mammoths (*Mammuthus primigenius*) populated Siberia, Beringia, and North America during the Pleistocene and early Holocene. Recent breakthroughs in ancient DNA sequencing have allowed for complete genome sequencing for two specimens of woolly mammoths (Palkopoulou et al. 2015). One mammoth specimen is from a mainland population 45,000 years ago when mammoths were plentiful. The second, a 4300 yr old specimen, is derived from an isolated population on Wrangel island where mammoths subsisted with small effective population size more than 43-fold lower than previous populations. These extreme differences in effective population size offer a rare opportunity to test nearly neutral models of genome architecture evolution within a single species. Using these previously published mammoth sequences, we identify deletions, retrogenes, and non-functionalizing point mutations. In the Wrangel island mammoth, we identify a greater number of deletions, a larger proportion of deletions affecting gene sequences, a greater number of candidate retrogenes, and an increased number of premature stop codons. This accumulation of detrimental mutations is consistent with genomic meltdown in response to low effective population sizes in the dwindling mammoth population on Wrangel island. In addition, we observe high rates of loss of olfactory receptors and urinary proteins, either because these loci are non-essential or because they were favored by divergent selective pressures in island environments. Finally, at the locus of *FOXQ1* we observe two independent loss-of-function mutations, which would confer a satin coat phenotype in this island woolly mammoth.

Author summary

We observe an excess of detrimental mutations, consistent with genomic meltdown in woolly mammoths on Wrangel Island just prior to extinction. We observe an excess of deletions, an increase in the proportion of deletions affecting gene sequences, and an excess of premature stop codons in response to evolution under low effective population sizes. Large numbers of olfactory receptors appear to have loss of function mutations in the island mammoth. These results offer genetic support within a single species for nearly-neutral theories of genome evolution. We also observe two independent loss of

function mutations at the *FOXQ1* locus, likely conferring a satin coat in this unusual woolly mammoth.

Introduction

Woolly mammoths (*Mammuthus primigenius*) were among the most populous large herbivores in North America, Siberia, and Beringia during the Pleistocene and early Holocene [1]. However warming climates and human predation led to extinction on the mainland roughly 10,000 years ago [2]. Lone isolated island populations persisted out of human reach until roughly 3,700 years ago when the species finally went extinct [3]. Recently, two complete high-quality high-coverage genomes were produced for two woolly mammoths [4]. One specimen is derived from the Siberian mainland at Oimyakon, dated to 45,000 years ago [4]. This sample comes from a time when mammoth populations were plentiful, with estimated effective population size of $N_e = 13,000$ individuals [4]. The second specimen is from Wrangel Island off the north Siberian coast [4]. This sample from 4,300 years ago represents one of the last known mammoth specimens. This individual comes from a small population estimated to contain roughly 300 individuals [4]. These two specimens offer the rare chance to explore the ways the genome responds to pre-extinction population dynamics.

Nearly neutral theories of genome evolution predict that small population sizes will lead to an accumulation of detrimental variation in the genome [5]. Such explanations have previously been invoked to explain genome content and genome size differences across multiple species [6]. Yet, within-species comparisons of how genomes are changed by small effective population sizes remain necessarily rare. These mammoth specimens offer the unique opportunity for within-species comparative genomics under a 43-fold reduction in population size. This comparison offers a major advantage as it will be free from confounding biological variables that are present in cross species comparisons. If nearly neutral dynamics lead to an excess of detrimental variation, we should observe an excess of harmful mutations in pre-extinction mammoths from Wrangel Island.

We use these two ancient DNA sequences to identify retrogenes, deletions, premature stop codons, and point mutations found in the Wrangel Island and Oimyakon mammoths. We identify an excess of putatively detrimental mutations, with an excess of stop codons, an excess of deletions, an increase in the proportion of deletions affecting gene sequences, an increase in non-synonymous substitutions relative to synonymous substitutions, and an excess of retrogenes, reflecting increased transposable element activity. These data bear the signature of genomic meltdown in small populations, consistent with nearly-neutral genome evolution. They furthermore suggest large numbers of detrimental variants collecting in pre-extinction genomes, a warning for continued efforts to protect current endangered species with small population sizes.

Results

Excess of amino acid substitutions and stop codons

We identified all SNPs in each mammoth genome as well as one Indian elephant specimen, Maya, using GATK [7]. We identified all non-synonymous and synonymous changes relative to the *L. africana* reference genome (https://www.broadinstitute.org/scientific-community/science/projects/mammals-models/elephant/elephant-genome-project) using r3.7 annotations lifted over to *L. africana* 4.0 genome sequences. We observe a significant increase in the
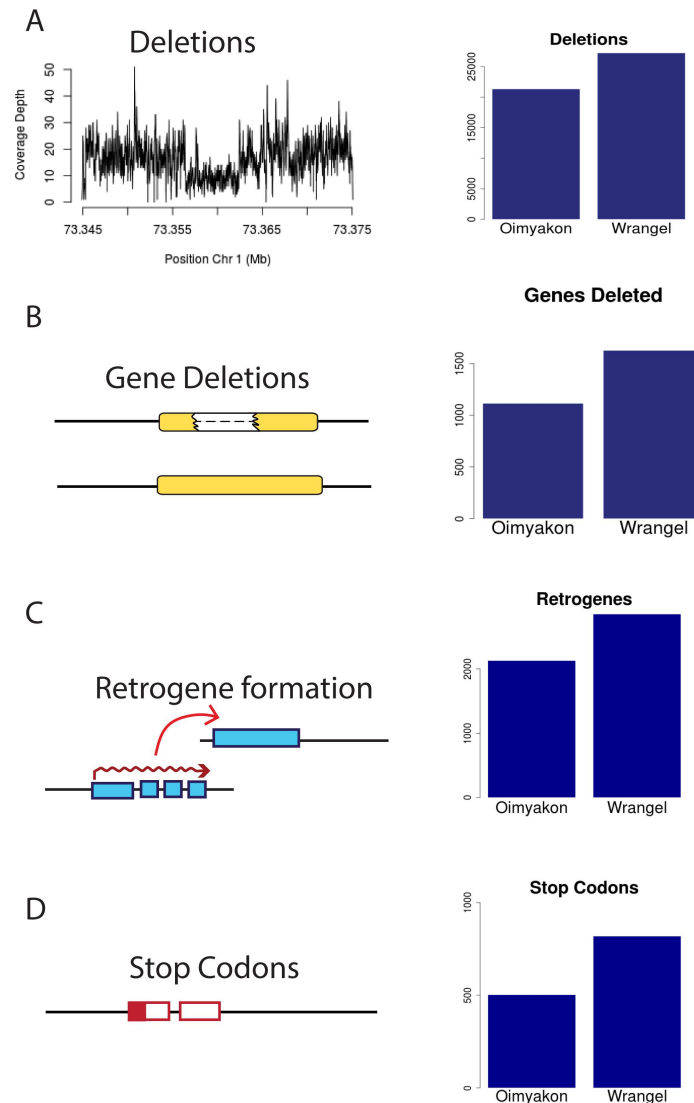
Fig 1. Excess of putatively detrimental mutations in the Wrangel island genome. A) Deletions B) Genes deleted C) Retrogenes D) Premature stop codons. Numbers shown are corrected for false negative rates of 30% for heterozygous SNPs and 0.5% for deletions in the lower coverage Oimyakon mammoth.

number of heterozygous non-synonymous changes relative to synonymous changes in the Wrangel island genome compared with Oimyakon ($\chi^2 = 68.799$, $df = 1$, $P < 2.2 \times 10^{-16}$; S1 Table). There is also a significant increase in the number of homozygous mutations at non-synonymous sites relative to synonymous sites ($\chi^2 = 9.96$, $df = 1$, $P < 0.0016$; S1 Table). We further observe an excess of premature stop codons in the genome of the Wrangel Island mammoth, with 1.8X as many genes affected. There are 503 premature stop codons in the Oimyakon genome (adjusting for a 30% false negative rate at heterozygous sites) compared with 819 in the Wrangel island genome (Fig 1, Table 1). There are 318 genes that have premature stop codons that are shared across the two mammoths, and 357 genes that are truncated in both mammoths, including mutations that form at independent sites. A total of 120 of these genes have stop codons in the two mammoths as well as in Maya the Indian elephant, suggesting read through in the *L. africana* reference. Among truncated genes, there is a significant

Table 1. Mutations identified in mammoth genomes.

Mutation	Oimyakon	Wrangel	Maya
Deletions	21346[1]	27228	28095
Retrogenes	2130	2853	1575
Genes with exons deleted	1115[1]	1628	3427
Stop Codons	503[2]	819	450
Stop Codons, excluding damage	377	645	401

[1] Corrected for false negative rate of 0.5% in heterozygotes

[2] Corrected for false negative rate of 30% at heterozygous sites established by Palkopoulou et al 2015.
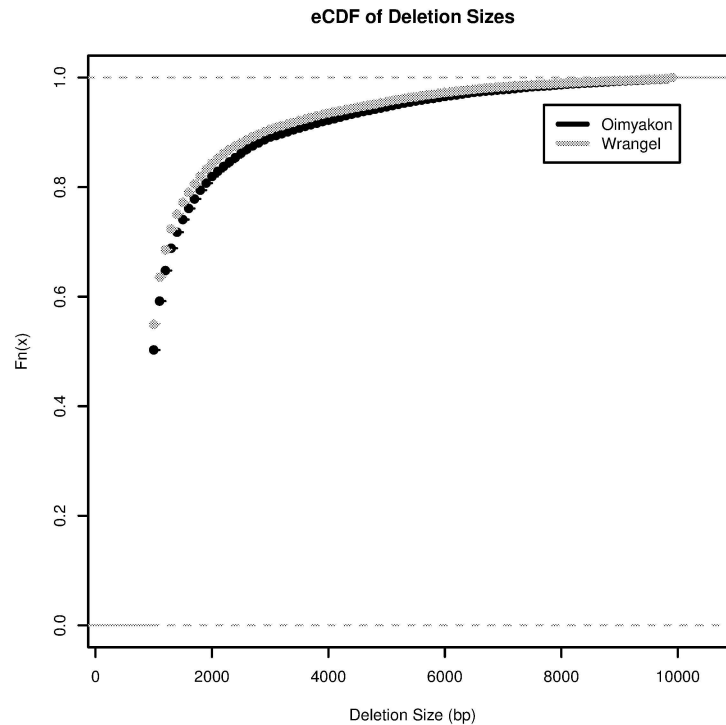
doi:10.1371/journal.pgen.1006601.t001

excess of olfactory genes and oderant binding receptors that appear to be pseudogenized with an EASE enrichment score of 9.1 (S2 Table) [8, 9]. We observe 85 truncated olfactory receptors and 3 vomeronasal receptors as well as multiple signal transduction peptides compared with 44 olfactory receptors and 2 vomeronasal receptors pseudogenized in the mainland mammoth.

It is possible that DNA damage in the archaic specimens could contribute to a portion of the observed stop codons. When we exclude A/G and C/T mutations, there is still a gross excess of premature stop codons, with 645 genes truncated in the Wrangel Island mammoth compared with 377 in the Oimyakon mammoth. Hence, the patterns are not explained solely by differential DNA damage in the two mammoths. Maya, the Indian Elephant specimen shows 450 premature stop codons, but 401 when A/G and T/C mutations are excluded. When putative damage to ancient DNA is excluded, Maya appears to house an intermediate number of premature stop codons, with a 6% increase compared to the Oimyakon mammoth.

Deletions

We identify 27228 deletions over 1 kb long in the Wrangel island genome, and 21346 (correcting for a 0.5% false negative rate at heterozygous sites) in the Oimyakon genome (Table 1). There are 6147 deletions (23%) identified in the Wrangel Island mammoth that are homozygous ($\leq$ 10% coverage) compared with 5035 (24%) in the Oimyakon mammoth. (S3 Table). A total of 13,459 deletions are identified in both mammoth genomes (S4 Table). Some 4813 deletions in the Wrangel Island mammoth and 4598 in the Oimyakon mammoth appear hemizygous but have stretches of zero coverage for at least 50% of their length. These sites may represent multiple independent homozygous deletions that cannot be differentiated via change point statistics. Alternatively, they might indicate smaller secondary deletions that appear on hemizygous haplotypes. Such secondary deletions are common when large loop mismatch repair attacks unpaired, hemizygous stretches of DNA [10, 11]. The Wrangel Island Mammoth has sharply increased heterozygosity for deletions in comparison with the Oimyakon mammoth (S3 Table). Some portion of the inflated heterozygosity for deletions in the Wrangel Island mammoth could be due to this difficulty in inferring genotypes in a high throughput setting. Alternatively, the effective mutation rate may have increased as fewer deletions were removed from the population via purifying selection, inflating θ_{del}. It is also possible that there was an increase in the rate of deletions in the Wrangel Island lineage due to defective DNA repair mechanisms. An increase in non-homologous end joining after DNA breaks rather than double stranded break repair could putatively induce such a change in the deletion rate.

Maya the Indian elephant shows a larger number of deletions than the Oimyakon mammoth, but with different character from the Wrangel Island mammoth. The bulk of these are derived from 22,954 hemizygous deletions (S3 Table). Maya houses only 5141 homozygous

eCDF of Deletion Sizes

Fig 2. eCDF for the size distribution of deletions in the Oimyakon and Wrangel island genomes. There is a significant reduction in the size of deletions identified in the Wrangel Island Genome.

doi:10.1371/journal.pgen.1006601.g002

deletions, similar to the mainland mammoth (S3 Table). There is an increase in the number of hemizygous deletions that affect gene sequences, but only a modest increase in the number of homozygous deletions that affect gene sequences (S3 Table). Competing pressures of higher N_e, longer time frames to accumulate mutations toward equilibrium frequencies, differences in mutation rates between the mammoths and elephants, differences in selective pressures, differences in the distribution of selective coefficients for deletions, different effective mutation rates due to different selective constraints, or differences in dominance coefficients might all contribute to differences in the number of deletions observed in elephants and mammoths. Additional samples would be necessary to determine the extent to which genetic declines may be influencing the diversity of deletions in modern Indian elephants. We currently have no basis for conclusions given this single sample, with no prior comparison.

There is a significant difference in the size distribution of deletions identified in the two mammoth samples, with a mean of 1707 bp in Oimyakon and 1606 bp in the Wrangel mammoth (Wilcox $W = 304430000$, $P < 2.2e - 16$; Fig 2). This difference could reflect either differences in DNA replication or repair mechanisms in the two mammoths, or altered selective constraints for different types of duplications. No significant difference is observed between the Wrangel island mammoth down sampled sequence data ($W = 2004400$, $P = 0.3917$) suggesting that the observed decrease in size is not due to differences in coverage. Some 1628 genes have deleted exons in the Wrangel Island mammoth compared to 1110 in Oimyakon (Table 1), a significant excess of genes deleted compared to expectations based on the number of deletions ($\chi^2 = 12.717$, $df = 1$, $P = 0.0003623$). Among these deleted genes, 112 in the mainland mammoth are homozygous compared to 133 homozygous exon deletions in the Wrangel Island Mammoth. Gene functions for affected genes in the Oimyakon mammoth include synapse functions, PHD domains, zinc fingers, aldo-keto metabolism, calcium dependent

membrane targeting, DNA repair, transcription regulation, and development (S5 Table). Gene functions overrepresented among deletions in the Wrangel Island mammoth include major urinary proteins, lipocalins, and pheromones, pleckstrins, transcription regulation, cell transport, DNA repair, chromatin regulation, hox domains, and development (S5 Table).

Among the genes deleted in the Wrangel Island mammoth, several have phenotypes of interest in other organisms. We observe a hemizygous deletion in riboflavin kinase *RFK* in the Wrangel Island mammoth, but normal coverage in the Oimyakon mainland mammoth (S1 Fig). Homozygous knockouts of riboflavin kinase, essential for B2 utilization/FAD synthesis, are embryonic lethal in mice [12]. Finally, we identify a hemizygous deletion in the Wrangel island mammoth that would remove the entire gene sequence at the *FOXQ1* locus (S2 Fig). The alternative haplotype carries a frameshift mutation that disrupts the FOXQ1 functional domain. *FOXQ1* knock-outs in mice are associated with the satin coat phenotype, which results in translucent fur but normal pigmentation due to abnormal development of the inner medulla of hairs [13], with two independent mutations producing this phenotype [13]. *FOXQ1* also regulates mucin secretion in the GI tract, a case of pleiotropic functions from a single gene [14]. If the phenotype in elephantids matches the phenotype exhibited in mice, this mammoth would have translucent hairs and a shiny satin coat, caused by two independently formed knock-out alleles at the same locus. These genes each have functions that are conserved across mammals, though there is no guarantee that they would produce identical phenotypes in other species.

Retrogene formation

Retrogene formation can serve as a proxy for retrotransposon activity. We identify retrogenes that display exon-exon junction reads in genomic DNA. We observe 1.3X more retrogenes formed in the Wrangel island mammoth. The Wrangel Island mammoth has 2853 candidate retrogenes, in comparison with 2130 in the Oimyakon mammoth and 1575 in Maya (Table 1). There are 436 retrogenes that are shared between the two mammoths, though some of these could arise via independent mutations. This excess of retrogenes is consistent with increased retroelement activity in the Wrangel Island lineage. During retrogene formation, highly expressed genes, especially those expressed in the germline, are expected to contribute to new retrogenes. To determine the types of loci that had been copied by retrotransposons, we performed a gene ontology analysis using DAVID [8, 9]. Functional categories overrepresented among candidate retrogenes include genes involved in transcription, translation, cell division/cytoskeleton, post translational modification, ubiquitination, and chaperones for protein folding (S6 and S7 Tables). All of these are expected to be highly expressed during cell divisions or constitutively expressed, consistent with expectations that highly expressed genes will be overrepresented. Gene ontologies represented are similar for both mammoths (S6 and S7 Tables). Although these retrogenes are unlikely to be detrimental in and of themselves, they may point to a burst of transposable element activity in the lineage that led to the Wrangel island individual. Such a burst of TE activity would be expected to have detrimental consequences, additionally contributing to genomic decline.

Genomic effects of demography

Under nearly-neutral theory of genome evolution, detrimental mutations should accumulate in small populations as selection becomes less efficient [5]. This increase in non-neutral amino acid changes and premature stop codons is consistent with reduced efficacy of selection in small populations. We attempted to determine whether the data is consistent with this nearly-neutral theory at silent and amino acid replacement substitutions whose mutation rates and

Table 2. Non-synonymous and synonymous heterozygosity.

	Wrangel	Oimyakon[1]
$H_S \pm 2\sigma$	0.00130 ± 0.00002	0.00161 ± 0.00002
$H_N \pm 2\sigma$	0.000490 ± 0.000012	0.000506 ± 0.000012
H_N/H_S	0.370	0.314

[1] Oimyakon corrected for false negative rate of 30% established by Palkopoulou et al 2015.

doi:10.1371/journal.pgen.1006601.t002

selection coefficients are well estimated in the literature. Under nearly neutral theory, population level variation for non-synonymous amino acid changes should accelerate toward parity with population level variation at synonymous sites.

Given the decreased population size on Wrangel Island, we expect to observe an accumulation of detrimental changes that would increase heterozygosity at non-synonymous sites (H_N) relative to synonymous sites (H_S) in the island mammoth. Heterozygosity depends directly on effective population sizes. We observe H_S = 0.00130 ± 0.00002 in the Wrangel Island mammoth, which is 80% of H_S = 0.00161 ± 0.00002 observed in the Oimyakon mammoth (Table 2). The magnitude of the difference between H_S in these two mammoths is 28 standard deviations apart, suggesting that these two mammoths could not have come from populations with the same effective population sizes. The specimens are well beyond the limits of expected segregating variation for a single population. To determine whether such results are consistent with theory, we fitted a model using PSMC [42] inferred population sizes for the Wrangel island mammoth, based on decay of heterozygosity of $(1 - 1/2N)^t H_0$. The observed reduction in heterozygosity is directly consistent theoretical expectations that decreased effective population sizes would lower heterozygosity to H_S = 0.00131.

At non-synonymous sites, however, there are no closed-form solutions for how H_N would decay under reduced population sizes. We observe H_N = 0.000490 in the Wrangel Island Mammoth, 95% of H_N = 0.000506 in the Oimyakon mammoth (Table 2). To determine whether such results could be caused by accumulation of nearly-neutral variation, we simulated population trajectories estimated using PSMC [42]. This trajectory shows ancient populations with $N_e \approx 10^4$, followed by a population decline prior to extinction. These numbers are slightly lower than previous estimates of ancestral N_e based on mitochondrial DNA [43]. We were able to qualitatively confirm results that population trajectories from PSMC with previously described mutation rates and selection coefficients can lead to an accumulation of detrimental alleles in populations. However, the magnitude of the effects is difficult to fit precisely. The simulations show a mean H_S = 0.00148 and H_N = 0.000339 in Oimyakon and H_S = 0.00126 and H_N = 0.000295 for the Wrangel Island Mammoth (S3 Fig). In simulations, we estimate H_N/H_S = 0.229 both for the Oimyakon mammoth and directly after the bottleneck, but H_N/H_S = 0.233 in the Wrangel Island Mammoth at the time of the Wrangel Island mammoth. These numbers are less than empirical observations of H_N/H_S = 0.370 (Table 2). Several possibilities might explain the observed disparity between precise estimates from simulations versus the data. The simulations may be particularly sensitive to perturbations from PSMC population levels or time intervals. Similarly, selection coefficients that differ from the gamma distribution previously estimated for humans might lead to greater or lesser changes in small populations. Additionally, an acceleration in generation time on Wrangel Island is conceivable, especially given the reduced size of Wrangel Island mammoths [15]. Finally, positive selection altering nucleotide variation on the island or the mainland could influence diversity levels.

Founder effects during island invasion sometimes alter genetic diversity in populations. However, it is unlikely that a bottleneck alone could cause an increase in H_N/H_S. There is no evidence in effective population sizes inferred using PSMC to suggest a strong bottleneck during Island colonization [4]. The power of such genetic analyses may be limited, but these results are in agreement with paleontological evidence showing no phenotypic differentiation from the mainland around 12,000 years ago followed by island dwarfism much later [15]. During glacial maxima, the island was fully connected to the mainland, becoming cut off as ice melted and sea levels rose. The timing of separation between the island and mainland lies between 10,000 years and 14,000 years before present [3, 15–17], but strontium isotope data for mammoth fossils suggests full isolation of island populations was not complete until 10,000-10,500 years ago [18]. Forward simulations suggest that hundreds of generations at small N_e are required for detrimental mutations to appear and accumulate in the population. These results are consistent with recent theory suggesting extended bottlenecks are required to diminish population fitness [19]. Thus, we suggest that a bottleneck alone could not produce the accumulation of H_N/H_S that we observe.

E. maximus indicus specimen, Maya shows an independent population decline in the past 100,000 years, with current estimates of N_e = 1000 individuals (S4 Fig). This specimen shows a parallel case of declining population sizes in a similar species of elephantid. Maya houses hemizygous deletions in similar numbers with the Wrangel Island Mammoth. However, the number of stop codons and homozygous deletions is intermediate in comparison with the Oimyakon and Wrangel mammoths (Table 1). It is possible that Indian elephants, with their recently reduced population sizes may be subject to similar accumulation of detrimental mutations, a prospect that would need to be more fully addressed in the future using population genomic samples for multiple individuals or timepoints and more thorough analyses.

Discussion

Nearly neutral theories of genome evolution

Nearly-neutral theories of genome evolution have attempted to explain the accumulation of genome architecture changes across taxa [5]. Under such models, mutations with selection coefficients less than the nearly neutral threshold will accumulate in genomes over time. Here, we test this hypothesis using data from a woolly mammoth sample from just prior to extinction. We observe an excess of retrogenes, deletions, amino acid substitutions, and premature stop codons in woolly mammoths on Wrangel Island. Given the long period of isolation and extreme population sizes observed in pre-extinction mammoths on Wrangel Island, it is expected that genomes would deteriorate over time. These results offer genetic support for the nearly-neutral theory of genome evolution, that under small effective population sizes, detrimental mutations can accumulate in genomes. Independent analysis supporting a reduction in nucleotide diversity across multiple individuals at MHC loci suggests a loss of balancing selection, further support the hypothesis that detrimental variants accumulated in small populations [20].

We observe two independent loss-of-function mutations in the Wrangel Island mammoth at the locus of *FOXQ1*. One mutation removes the entire gene sequence via a deletion, while the other produces a frameshift in the CDS. Based on phenotypes observed in mouse models, these two independent mutations would result in a satin fur coat, as well as gastric irritation [14]. Many phenotypic screens search for homozygous mutations as causative genetic variants that could produce disease. More recently, it has been proposed that the causative genetic variation for disease phenotypes may be heterozygous non-complementing detrimental mutations

[21]. These data offer one case study of independent non-functionalizing mutations in a single individual, genetic support for independent non-functionalizing mutations at a single locus. Woolly mammoth outer hairs house multiple medullae, creating a stiff outer coat that may have protected animals from cold climates [22] (though see [23] for alternative interpretations). Putative loss of these medullae through loss of *FOXQ1* could compromise this adaptation, leading to lower fitness.

Island specific changes

One of the two specimens comes from Wrangel Island, off the northern coast of Siberia. This mammoth population had been separated from the mainland population for at least 6000 years after all mainland mammoths had died off. Prior to extinction, some level of geographic differentiation combined with differing selective pressures led to phenotypic differentiation on Wrangel island [15]. Island mammoths had diminished size, but not until 12,000 years ago when mainland populations had reduced and ice sheets melted [15]. One possible explanation for the poor fit of simulations is that generation time may have decreased. Previous work suggested a very high mutation rate for woolly mammoths based on comparisons between island and mainland mammoths. It is possible that an acceleration in generation times could cause the accumulation of more mutations over time, and that the real mutation rate is similar to humans ($1 - 2 \times 10^{-8}$ [24] rather than 3.8×10^{-8} [4]). Such changes would be consistent with island dwarfism being correlated with shorter generation times, and would explain the unusually high mutation rate estimate for mammoths based on branch shortening observed in [4].

We observe large numbers of pseudogenized olfactory receptors in the Island mammoth. Olfactory receptors evolve rapidly in many mammals, with high rates of gain and loss [25]. The Wrangel island mammoth has massive excess even compared to the mainland mammoth. Wrangel island had different flora compared to the mainland, with peat and sedges rather than grasslands that characterized the mainland [17]. The island also lacked large predators present on the mainland. It is possible that island habitats created new selective pressures that resulted in selection against some olfactory receptors. Such evolutionary change would echo gain and loss of olfactory receptors in island *Drosophila* [26]. In parallel, we observe a large number of deletions in major urinary proteins in the island mammoth. In Indian elephants *E. maximus indicus*, urinary proteins and pheromones ellicit behavioral responses including mate choice and social status [27]. It is possible that coevolution between urinary proteins, olfactory receptors, and vomeronasal receptors led to a feedback loop, allowing for rapid loss in these related genes. It is equally possible that urinary peptides and olfactory receptors are not essential and as such they are more likely to fall within the nearly neutral range [25]. Either of these hypotheses could explain the current data.

Implications for conservation genetics

Many factors contributed to the demise of woolly mammoths in prehistoric times. Climate change led to receding grasslands as forests grew in Beringia and North America and human predation placed a strain on already struggling populations [2]. Unlike many cases of island invasion, Wrangel Island mammoths would not have continuous migration to replenish variation after mainland populations went extinct. Under such circumstances, detrimental variation would quickly accumulate on the island. The putatively detrimental variation observed in these island mammoths, with the excess of deletions, especially recessive lethals may also have limited survival of these struggling pre-extinction populations. Climate change created major limitations for mammoths on other islands [28], and these mammoths may have struggled to overcome similar selective pressures.

Many modern day species, including elephants, are threatened or endangered. Asiatic chee-tahs are estimated to have fewer than 100 individuals in the wild [29]. Pandas are estimated to have 1600 individuals living in highly fragmented territories [30]. Mountain Gorilla popula-tion census sizes have been estimated as roughly 300 individuals, similar to effective popula-tion sizes for pre-extinction mammoths [31]. If nearly neutral dynamics of genome evolution affect contemporary endangered species, detrimental variation would be expected in these genomes. With single nucleotide changes, recovered populations can purge detrimental varia-tion in hundreds to thousands of generations, returning to normal genetic loads [19]. How-ever, with deletions that become fixed in populations, it is difficult to see how genomes could recover quickly. The realm of back mutations to reproduce deleted gene sequences will be lim-ited or impossible. Although compensatory mutations might conceivably correct for some det-rimental mutations, with small effective population sizes, adaptation through both new mutation and standing variation may be severely limited [32]. Thus we might expect genomes affected by genomic meltdown to show lasting repercussions that will impede population recovery.

Ethics statement

All sequences are taken from publicly available sequence data in the ENA or SRA. Indian ele-phant specimens for previously published sequence data were handled by the San Diego Zoo.

Materials and methods

Genome sequence alignments

We used previously aligned bam files from ERR852028 (Oimyakon, 11X) and ERR855944 (Wrangel, 17X) (S8 Table) [4] aligned against the *L. africana* 4.0 reference genome (available on request from the Broad Institute—vertebrategenomes@broadinstitute.org; https://www. broadinstitute.org/scientific-community/science/projects/mammals-models/elephant/ elephant-genome-project). We also aligned 33X coverage of sequencing reads for one modern *E. maximus indicus* genome Maya (previously described as "Uno") using bwa 0.7.12-r1044 [33], with parameters set according to [4] bwa aln -l 16500 -o 2 -n 0.01. The *E. maxi-mus indicus* sample, previously labeled in the SRA as "Uno", is from Maya, a former resident of the San Diego Zoo wild-born in Assam, India, North American Studbook Number 223, Local ID #141002 (O. Ryder, personal communication). We were not able to use two other mammoth sequences are publicly available, M4 and M25 from Lynch et al. [34]. These sequences display abnormal PSMC results (S4 Fig), high heterozygosity (S5 Fig), and many SNPs with asymmetrical read support (S6 Fig). The unrealistically high heterozygosity as well as abnormal heterozygote calls raise concerns with respect to sequence quality. For further description, please see Supporting Information.

Synonymous and nonsynonymous substitutions

We used the GATK pipeleine [7] v3.4-0-g7e26428 to identify SNPs in the aligned sequence files for the Oimyakon and Wrangel Island mammoths. We identified and realigned all indel span-ning reads according to the standard GATK pipeline. We then identified all SNPs using the Unified Genotyper, with output mode set to emit all sites. We used all CDS annotations from cDNA annotations from *L. africana* r3.7 with liftover coordinates provided for *L. africana* 4.0 to identify SNPs within coding sequences. We identified all stop codons, synonymous substi-tutions, and non-synonymous substitutions for the Wrangel Island and Oimyakon mammoths at heterozygous and homozygous sites.

Retrogenes

We aligned all reads from the mammoth genome sequencing projects ERR852028 (Oimyakon) and ERR855944 (Wrangel) (S8 Table) against elephant cDNA annotations from *L. africana* r3.7. Sequences were aligned using bwa 0.7.12-r1044 [33], with parameters set according to [4] `bwa aln -l 16500 -o 2 -n 0.01` in order to account for alignments of damaged ancient DNA. We then collected all reads that map to exon-exon boundaries with at least 10 bp of overhang. Reads were then filtered against aligned genomic bam files produced by Palkopoulou et al [4], discarding all exon-exon junction reads that have an alignment with equal or better alignments in the genomic DNA file. We then retained all putative retrogenes that showed signs of loss for two or more introns, using only cases with 3 or more exon-exon junction reads.

Deletions

We calculated coverage depth using samtools [35] with a quality cutoff of -q 20. We then implemented change point analysis [36] in 20 kb windows. Change point methods have been commonly used to analyze microarray data and single read data for CNVs [37–39] The method seeks compares the difference in the log of sum of the squares of the residuals with one regression line vs. two regression lines [36]. The test statistic follows a chi-squared distribution with a number of degrees of freedom determined by the number of change-points in the data, in this case $df = 1$. We required significance at a Bonferroni corrected p-value of 0.05 or less. We allowed for a maximum of one CNV tract per window, with minimum of 1 kb and maximum of 10 kb (half the window size) with a 100 bp step size. We did not attempt to identify deletions smaller than 1 kb due to general concerns of ancient DNA sequence quality, limitations to assess small deletions in the face of stochastic coverage variation, and concerns that genotype calls for smaller deletions might not be as robust to differences in coverage between the two mammoths. Sequences with 'N's in the reference genome did not contribute to change point detection. We excluded all deletions that were identified as homozygous mutations in both mammoths and in *E. maximus indicus* specimen Maya, as these suggest insertion in the *L. africana* reference rather than deletion in other elephantids. To determine the effects that coverage differences would have on deletions, we downsampled the sequence file for the Wrangel Island mammoth using samtools to 11X coverage, using chromosome 1 as a test set. We observe a reduction in the number of deletions for chromosome 1 from 1035 deletions to 999 deletions, resulting in an estimated false negative rate of 0.5% at reduced coverage for deletions greater than 1 kb. Highly diverged haplotypes with greater than 2% divergence might prevent read mapping and mimic effects of deletions, but this would require divergence times within a species that are greater than the divergence between mammoths and *L. africana*. Mutations were considered homozygous if mean coverage for the region was less than 10% of the background coverage level. Otherwise it was considered to be heterozygous. These methods are high-throughput, and it is possible that multiple small homozygous deletions interspersed with full coverage sequences might mimic heterozygote calls. Whether such mutations might meet the conditions for significant change-point detection would depend on the deletion length, placement, and background coverage level.

Demography

We identified SNPs that differentiate Mammoth genomes from the reference using samtools mpileup (options -C50 -q30 -Q30), and bcftools 1.2 consenus caller (bcftools call -c). The resulting vcf was converted to fastq file using bcftools vcf2fq.pl with a mimimum depth of 3 reads and a maximum depth of twice the mean coverage for each genome. Sequences were

then converted to psmc fasta format using fq2psmcfa provided by psmc 0.6.5-r67. We then ran psmc with 25 iterations (-N25), an initial ratio of θ/ρ of 5 (-r5), and parameters 64 atomic time intervals and 28 free parameters (−p "4+25*2+4+6") as was done in previous analysis of woolly mammoths [4]. Effective population sizes and coalescence times were rescaled using previously estimated mutation rates of 3.8×10^{-8}. Using the population size estimates from PSMC, we calculated the expected reduction in heterozygosity at synonymous sites according to $\left(1 - \frac{1}{2N}\right)^t$ for each time period in PSMC output. We compared the number of deletions, number of premature stop codons, proportion affecting gene sequences, and number of putative retrogenes between the two mammoth genomes using chi squared tests.

Simulations

To determine expectations of sequence evolution at non-synonymous sites under population crash, we ran simulations using SLiM v. 2.0 population genetic software [40]. We modeled two classes of sites: neutral and detrimental. For detrimental mutations we used a gamma distributed DFE with a mean of -0.043 and a shape parameter of 0.23 as estimated for humans [41], assuming a dominance coefficient of 0.5 and free recombination across sites. Mutation rates were set as 3.8×10^{-8} based on previously published estimates [4]. The trajectory of population sizes was simulated according to estimates from PSMC, omitting the initial and final time points from PSMC, which are often subject to runaway behavior. We then simulated the accumulation of H_N/H_S in the Wrangel Island Mammoths. Simulations were run with a burn-in of 100,000 generations. We simulated 460 replicates of haplotypes with 100 sites for each mutation class.

Gene ontology

To gather a portrait of functional categories captured by deletions, retrogenes, and stop codons, we identified all mouse orthologs based on ENSEMBL annotations for *L. africana* 3.7 for affected gene sequences. We then used DAVID gene ontology analysis with the clustering threshold set to 'Low' (http://david.nicfcrf.gov/; Accessed April 2016) [8, 9]. S2–S7 Tables include all functions overrepresented at an EASE enrichment cutoff of 2.0. Full gene ontology data is included in Supplementary Information.

Supporting information

S1 Text. Supporting text.
(PDF)

S1 Table. Non synonymous and synonymous changes in Wrangel and Oimyakon mammoths.
(PDF)

S2 Table. Gene ontology categories for premature stop codons in mammoths.
(PDF)

S3 Table. Deletions identified in mammoth genomes.
(PDF)

S4 Table. Shared deletions in mammoth genomes.
(PDF)

S5 Table. Gene ontology for deleted exons.
(PDF)

S6 Table. Gene ontology for retrogenes in the Oimyakon mammoth.
(PDF)

S7 Table. Gene ontology for retrogenes in the Wrangel island mammoth.
(PDF)

S8 Table. SRA and ENA identifiers for mammoth and Elephant sequence data.
(PDF)

S9 Table. Heterozygous sites per 10 kb.
(PDF)

S10 Table. Asymmetrical support.
(PDF)

S1 Fig. Coverage depth at the RFK locus.
(PDF)

S2 Fig. Coverage depth at the FOXQ1 locus.
(PDF)

S3 Fig. Simulations for heterozygosity at synonymous and non-synonymous sites.
(PDF)

S4 Fig. PSMC results.
(PDF)

S5 Fig. Heterozygosity for mammoth and elephant samples.
(PDF)

S6 Fig. Asymmetric SNPs.
(PDF)

S7 Fig. Asymmetric SNPs, excluding damage.
(PDF)

S1 Data. SuppFiles.zip—Data release archive.
(ZIP)

Acknowledgments

Author Contributions

Conceptualization: RLR.

Data curation: RLR.

Formal analysis: RLR MS.

Funding acquisition: MS.

Investigation: RLR.

Methodology: RLR MS.

Project administration: RLR MS.

Software: RLR.

Supervision: MS.

Validation: RLR MS.

Visualization: RLR MS.

Writing – original draft: RLR MS.

Writing – review & editing: RLR MS.

References

1. Stuart AJ, Kosintsev P, Higham T, Lister AM (2004) Pleistocene to holocene extinction dynamics in giant deer and woolly mammoth. Nature 431: 684–689. doi: 10.1038/nature02890 PMID: 15470427

2. Nogués-Bravo D, Rodríguez J, Hortal J, Batra P, Araújo MB (2008) Climate change, humans, and the extinction of the woolly mammoth. PLoS Biol 6: e79. doi: 10.1371/journal.pbio.0060079 PMID: 18384234

3. Vartanyan 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. Quaternary Research 70: 51–59. doi: 10.1016/j.yqres.2008.03.005

4. Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, et al. (2015) Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. Current Biology 25: 1395–1400. doi: 10.1016/j.cub.2015.04.007 PMID: 25913407

5. Lynch M (2007) The origins of genome architecture, volume 98. Sinauer Associates Sunderland.

6. Lynch M (2006) The origins of eukaryotic gene structure. Molecular Biology and Evolution 23: 450–468. doi: 10.1093/molbev/msj050 PMID: 16280547

7. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. (2010) The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. Genome research 20: 1297–1303. doi: 10.1101/gr.107524.110 PMID: 20644199

8. Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research 37: 1–13. doi: 10.1093/nar/gkn923 PMID: 19033363

9. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using david bioinformatics resources. Nature protocols 4: 44–57. doi: 10.1038/nprot.2008.211 PMID: 19131956

10. Rogers RL, Cridland JM, Shao L, Hu TT, Andolfatto P, et al. (2014) Landscape of standing variation for tandem duplications in drosophila yakuba and drosophila simulans. Molecular biology and evolution: msu124. doi: 10.1093/molbev/msu124 PMID: 24710518

11. Kearney HM, Kirkpatrick DT, Gerton JL, Petes TD (2001) Meiotic recombination involving heterozygous large insertions in saccharomyces cerevisiae: formation and repair of large, unpaired dna loops. Genetics 158: 1457–1476. PMID: 11514439

12. Yazdanpanah B, Wiegmann K, Tchikov V, Krut O, Pongratz C, et al. (2009) Riboflavin kinase couples tnf receptor 1 to nadph oxidase. Nature 460: 1159–1163. doi: 10.1038/nature08206 PMID: 19641494

13. Hong HK, Noveroske JK, Headon DJ, Liu T, Sy MS, et al. (2001) The winged helix/forkhead transcription factor foxq1 regulates differentiation of hair in satin mice. Genesis 29: 163–171. doi: 10.1002/gene.1020 PMID: 11309849

14. Verzi MP, Khan AH, Ito S, Shivdasani RA (2008) Transcription factor foxq1 controls mucin gene expression and granule content in mouse stomach surface mucous cells. Gastroenterology 135: 591–600. doi: 10.1053/j.gastro.2008.04.019 PMID: 18558092

15. Vartanyan S, Garutt V, Sher AV (1993) Holocene dwarf mammoths from wrangel island in the siberian arctic. Nature 362. doi: 10.1038/362337a0

16. Elias SA, Short SK, Nelson CH, Birks HH (1996) Life and times of the bering land bridge. Nature 382: 60–63. doi: 10.1038/382060a0

17. Lozhkin A, Anderson P, Vartanyan S, Brown T, Belaya B, et al. (2001) Late quaternary paleoenvironments and modern pollen data from wrangel island (northern chukotka). Quaternary Science Reviews 20: 217–233. doi: 10.1016/S0277-3791(00)00121-9

18. Arppe L, Karhu JA, Vartanyan SL (2009) Bioapatite 87sr/86sr of the last woolly mammoths-implications for the isolation of wrangel island. Geology 37: 347–350. doi: 10.1130/G25467A.1

19. Balick DJ, Do R, Cassa CA, Reich D, Sunyaev SR (2015) Dominance of deleterious alleles controls the response to a population bottleneck. PLoS Genet 11: e1005436. doi: 10.1371/journal.pgen.1005436 PMID: 26317225

20. Pečnerová P, Díez-del Molino D, Vartanyan S, Dalén L (2016) Changes in variation at the mhc class ii dqa locus during the final demise of the woolly mammoth. Scientific reports 6. doi: 10.1038/srep25274 PMID: 27143688

21. Thornton KR, Foran AJ, Long AD (2013) Properties and modeling of gwas when complex disease risk is due to non-complementing, deleterious mutations in genes of large effect. PLoS Genet 9: e1003258. doi: 10.1371/journal.pgen.1003258 PMID: 23437004

22. Tridico SR, Rigby P, Kirkbride KP, Haile J, Bunce M (2014) Megafaunal split ends: microscopical characterisation of hair structure and function in extinct woolly mammoth and woolly rhino. Quaternary Science Reviews 83: 68–75. doi: 10.1016/j.quascirev.2013.10.032

23. Chernova O, Kirillova I, Boeskorov G, Shidlovskiy F, Kabilov M (2015) Architectonics of the hairs of the woolly mammoth and woolly rhino. Proceeding of the Zoological Institute RAS 319: 441–460.

24. Scally A, Durbin R (2012) Revising the human mutation rate: implications for understanding human evolution. Nature Reviews Genetics 13: 745–753. doi: 10.1038/nrg3295 PMID: 22965354

25. Nei M, Niimura Y, Nozawa M (2008) The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nature Reviews Genetics 9: 951–963. doi: 10.1038/nrg2480 PMID: 19002141

26. Stensmyr MC, Dekker T, Hansson BS (2003) Evolution of the olfactory code in the drosophila melanogaster subgroup. Proceedings of the Royal Society of London B: Biological Sciences 270: 2333–2340. doi: 10.1098/rspb.2003.2512 PMID: 14667348

27. Rasmussen L, Lazar J, Greenwood D (2003) Olfactory adventures of elephantine pheromones. Biochemical Society Transactions 31: 137–141. doi: 10.1042/bst0310137 PMID: 12546671

28. Graham RW, Belmecheri S, Choy K, Culleton BJ, Davies LJ, et al. (2016) Timing and causes of mid-holocene mammoth extinction on st. paul island, alaska. Proceedings of the National Academy of Sciences: 201604903. doi: 10.1073/pnas.1604903113 PMID: 27482085

29. Hunter L, Jowkar H, Ziaie H, Schaller G, Balme G, et al. (2007) Conserving the asiatic cheetah in iran: launching the first radio-telemetry study. Cat News 46: 8–11.

30. Wang X, Xu W, Ouyang Z (2009) Integrating population size analysis into habitat suitability assessment: implications for giant panda conservation in the minshan mountains, china. Ecological research 24: 1101–1109. doi: 10.1007/s11284-009-0589-2

31. Guschanski K, Vigilant L, McNeilage A, Gray M, Kagoda E, et al. (2009) Counting elusive animals: comparing field and genetic census of the entire mountain gorilla population of bwindi impenetrable national park, uganda. Biological Conservation 142: 290–300. doi: 10.1016/j.biocon.2008.10.024

32. Pennings PS, Hermisson J (2006) Soft sweeps ii-molecular population genetics of adaptation from recurrent mutation or migration. Molecular biology and evolution 23: 1076–1084. doi: 10.1093/molbev/msj117 PMID: 16520336

33. Li H, Durbin R (2009) Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics 25: 1754–1760. doi: 10.1093/bioinformatics/btp324 PMID: 19451168

34. Lynch VJ, Bedoya-Reina OC, Ratan A, Sulak M, Drautz-Moses DI, et al. (2015) Elephantid genomes reveal the molecular bases of woolly mammoth adaptations to the arctic. Cell reports 12: 217–228. doi: 10.1016/j.celrep.2015.06.027 PMID: 26146078

35. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The sequence alignment/map format and samtools. Bioinformatics 25: 2078–2079. doi: 10.1093/bioinformatics/btp352 PMID: 19505943

36. Yao YC (1988) Estimating the number of change-points via schwarz'criterion. Statistics & Probability Letters 6: 181–189. doi: 10.1016/0167-7152(88)90118-6

37. Olshen AB, Venkatraman E, Lucito R, Wigler M (2004) Circular binary segmentation for the analysis of array-based dna copy number data. Biostatistics 5: 557–572. doi: 10.1093/biostatistics/kxh008 PMID: 15475419

38. Chiang DY, Getz G, Jaffe DB, O'Kelly MJ, Zhao X, et al. (2009) High-resolution mapping of copy-number alterations with massively parallel sequencing. Nature methods 6: 99–103. doi: 10.1038/nmeth. 1276 PMID: 19043412

39. Niu YS, Zhang H (2012) The screening and ranking algorithm to detect dna copy number variations. The annals of applied statistics 6: 1306. doi: 10.1214/12-AOAS539SUPP PMID: 24069112

40. Messer PW (2013) Slim: simulating evolution with selection and linkage. Genetics 194: 1037–1039. doi: 10.1534/genetics.113.152181 PMID: 23709637

41. Eyre-Walker A, Woolfit M, Phelps T (2006) The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics 173: 891–900. doi: 10.1534/genetics.106.057570 PMID: 16547091

42. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. Nature 475: 493–496. doi: 10.1038/nature10231 PMID: 21753753

43. Barnes I, Shapiro B, Lister A, Kuznetsova T, Sher A, et al. (2007) Genetic structure and extinction of the woolly mammoth, mammuthus primigenius. Current Biology 17: 1072–1075. doi: 10.1016/j.cub.2007. 05.035 PMID: 17555965