Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

Current Biology

CellPress
OPEN ACCESS

Report

Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros

Edana Lord,[1,2,3,33,34,*] Nicolas Dussex,[1,2,3,33] Marcin Kierczak,[4] David Díez-del-Molino,[1,3] Oliver A. Ryder,[5] David W.G. Stanton,[1,2] M. Thomas P. Gilbert,[6,7] Fátima Sánchez-Barreiro,[6] Guojie Zhang,[8,9,10,11] Mikkel-Holger S. Sinding,[6,12] Eline D. Lorenzen,[6] Eske Willerslev,[6] Albert Protopopov,[13] Fedor Shidlovskiy,[14] Sergey Fedorov,[15] Hervé Bocherens,[16,17] Senthilvel K.S.S. Nathan,[18] Benoit Goossens,[18,19,20,21]

(Author list continued on next page)

[1]Centre for Palaeogenetics, Svante Arrhenius väg 20C, Stockholm 10691, Sweden
[2]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Box 50007, Stockholm 10405, Sweden
[3]Department of Zoology, Stockholm University, Stockholm 10691, Sweden
[4]Dept of Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Husargatan 3, Uppsala 752 37, Sweden
[5]San Diego Zoo Institute for Conservation Research, 15600 San Pasqual Valley Road, Escondido, CA 92027, USA
[6]GLOBE Institute, University of Copenhagen, Øster Farimagsgade 5A, Copenhagen 1352, Denmark
[7]Norwegian University of Science and Technology, University Museum, Trondheim 7491, Norway
[8]Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen 2100, Denmark
[9]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
[10]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China
[11]BGI-Shenzhen, Shenzhen 518083, China
[12]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland
[13]Academy of Sciences of Sakha (Yakutia), Yakutsk, Russia

(Affiliations continued on next page)

SUMMARY

Ancient DNA has significantly improved our understanding of the evolution and population history of extinct megafauna. However, few studies have used complete ancient genomes to examine species responses to climate change prior to extinction. The woolly rhinoceros (*Coelodonta antiquitatis*) was a cold-adapted megaherbivore widely distributed across northern Eurasia during the Late Pleistocene and became extinct approximately 14 thousand years before present (ka BP). While humans and climate change have been proposed as potential causes of extinction [1–3], knowledge is limited on how the woolly rhinoceros was impacted by human arrival and climatic fluctuations [2]. Here, we use one complete nuclear genome and 14 mitogenomes to investigate the demographic history of woolly rhinoceros leading up to its extinction. Unlike other northern megafauna, the effective population size of woolly rhinoceros likely increased at 29.7 ka BP and subsequently remained stable until close to the species' extinction. Analysis of the nuclear genome from a ~18.5-ka-old specimen did not indicate any increased inbreeding or reduced genetic diversity, suggesting that the population size remained steady for more than 13 ka following the arrival of humans [4]. The population contraction leading to extinction of the woolly rhinoceros may have thus been sudden and mostly driven by rapid warming in the Bølling-Allerød interstadial. Furthermore, we identify woolly rhinoceros-specific adaptations to arctic climate, similar to those of the woolly mammoth. This study highlights how species respond differently to climatic fluctuations and further illustrates the potential of palaeogenomics to study the evolutionary history of extinct species.

RESULTS AND DISCUSSION

Genome Sequencing

To investigate changes in genetic diversity that preceded the extinction of the woolly rhinoceros, and to obtain a glimpse of the species' genomic adaptation to the arctic environment, we sequenced a woolly rhinoceros nuclear genome and 14 mitochondrial genomes that ranged in age from >50 to 14.1 thousand calibrated years before present (ka cal BP). The specimen from which we recovered the nuclear genome was

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

CellPress
OPEN ACCESS

Current Biology
Report

Johannes van der Plicht,[22] Yvonne L. Chan,[1,2,3] Stefan Prost,[23,24] Olga Potapova,[13,25,26] Irina Kirillova,[14] Adrian M. Lister,[27] Peter D. Heintzman,[28] Joshua D. Kapp,[29] Beth Shapiro,[29,30] Sergey Vartanyan,[31] Anders Götherström,[1,32] and Love Dalén[1,2,3,*]

[14]Ice Age Museum, National Alliance of Shidlovskiy 'Ice Age', Moscow, Russia
[15]Mammoth museum of North-Eastern Federal University, Yakutsk, Russia
[16]Senckenberg Centre for Human Evolution and Palaeoenvironment (S-HEP), Sigwartstr. 10, Tübingen 72076, Germany
[17]Department of Geosciences, Biogeology, University of Tübingen, Hölderlinstr. 12, Tübingen 72074, Germany
[18]Sabah Wildlife Department, Wisma Muis, 88100 Kota Kinabalu, Sabah, Malaysia
[19]Organisms and Environment Division, Cardiff School of Biosciences, 33 Park Place, Cardiff CF10 3BA, UK
[20]Sustainable Places Research Institute, Cardiff University, 33 Park Place, Cardiff CF10 3BA, UK
[21]Danau Girang Field Centre, c/o Sabah Wildlife Department, Wisma Muis, 88100 Kota Kinabalu, Sabah, Malaysia
[22]Center for Isotope Research, Groningen University, Groningen, the Netherlands
[23]LOEWE-Centre for Translational Biodiversity Genomics, Senckenberg Museum, Frankfurt, Germany
[24]South African National Biodiversity Institute, National Zoological Garden, Pretoria, South Africa
[25]Pleistocene Park Foundation, Philadelphia, PA 19006, USA
[26]Mammoth Site of Hot Springs, SD, Inc., Hot Springs, SD 57747, USA
[27]Department of Earth Sciences, Natural History Museum, London SW7 5BD, UK
[28]The Arctic University Museum of Norway, UiT The Arctic University of Norway, Tromsø 9037, Norway
[29]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz CA 95064, USA
[30]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA 96050, USA
[31]N.A. Shilo North-East Interdisciplinary Scientific Research Institute, Far East Branch, Russian Academy of Sciences (NEISRI FEB RAS), Magadan 685000, Russia
[32]Archaeological Research Laboratory, Department of Archaeology and Classical Studies, Stockholm University, Stockholm 106 91, Sweden
[33]These authors contributed equally
[34]Lead Contact
*Correspondence: edana.lord@zoologi.su.se (E.L.), love.dalen@nrm.se (L.D.)
https://doi.org/10.1016/j.cub.2020.07.046

radiocarbon dated to 18,530 ± 170 cal BP (Data S1A) and had an endogenous DNA content of 56.7%. After mapping the raw data to a new and high-quality assembly of its closest extant relative, the Sumatran rhinoceros (*Dicerorhinus sumatrensis*) [5, 6], the average genome coverage was 13.6X, with 70% of the genome having a coverage ≥10X. The average DNA fragment length was 84 bp, and overall, we identified 28,180,718 high-quality SNPs after filtering out SNPs with low mapping and base quality and low coverage (see STAR Methods). Furthermore, we conducted low-coverage shotgun sequencing on 13 additional woolly rhinoceros specimens, recovering in total 14 full mitochondrial genomes from northeastern Siberia (Figure 1A) with an average depth ranging from 7.5X to 912.8X (Data S1A).

Demographic History

Among the 14 mitochondrial genomes, we identified 13 unique mitogenome haplotypes (Figure S1A). Based on both a Bayesian phylogeny and median-joining network, we identified two clades (1 and 2) that diverged ~205 ka BP (95% highest posterior density [HPD]: 440–116 ka BP) and persisted up until close to the extinction of the woolly rhinoceros (Figures 1B and S1A). There was no indication of geographic or temporal structuring between or within these clades, except for a single sample from Wrangel Island forming its own distinct lineage, diverging from clade 1 ~154 ka BP (95% HPD: 326–91 ka BP). This unique lineage may result from isolation on Wrangel Island because of less favorable habitat south of this locality, as steppe turned into forested and shrub lowlands during warm periods [9, 10]. Future studies that include additional samples from Wrangel will allow exploration of potential genetic structure between Wrangel and adjacent regions in northeastern Siberia.

The mitochondrial phylogeny had long, well-resolved branches between the three clades (Figure 1B), similar to the pattern observed in woolly mammoths (*Mammuthus primigenius*) [11]. It has been hypothesized that this mitogenome structure in mammoths potentially resulted from isolation in interglacial refugia [11]. Finding a similar structure in woolly rhinoceros and mammoths suggests that these species responded similarly to past climate warming. Within clades 1 and 2 of the mitochondrial phylogeny, the short unresolved branches suggest a recent diversification approximately 86–22 ka BP (95% HPDs for Node C and D, Figure 1B). In the demographic analyses, a model of constant population size obtained the highest support (Figure S1B; Data S1B), with a female effective population size (N_{ef}) of around 100,000 over the last 110 ka until the extinction at ca. 14 ka BP. However, an alternative model, albeit with a lower likelihood of support, indicated an expansion in N_{ef}, consistent with the recent diversification of lineages within each clade observed in the phylogeny (Figure S1C).

To further examine the demographic history of the woolly rhinoceros, we used a pairwise sequentially Markovian coalescent (PSMC) analysis based on the nuclear genome (Figure 2). Effective population size (N_e) increased gradually from ~1 Ma BP during the Early Pleistocene, reaching a peak of ~21,000 at around 152 ka BP (95% confidence interval [CI]: 274–111 ka BP), during the Marine Isotope Stage 6 (MIS 6) glaciation (130–191 ka BP). Subsequently, N_e decreased 10-fold from 127 ka BP (95% CI: 226–94 ka BP) until 29.7 ka BP (95% CI: 40–26.3 ka BP), at which point there was a rapid expansion in N_e. The effective population size then remained constant until the time of the death of the individual (18.5 ka cal BP), approximately 4.5 ka prior to the extinction of the species.

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

CellPress
OPEN ACCESS

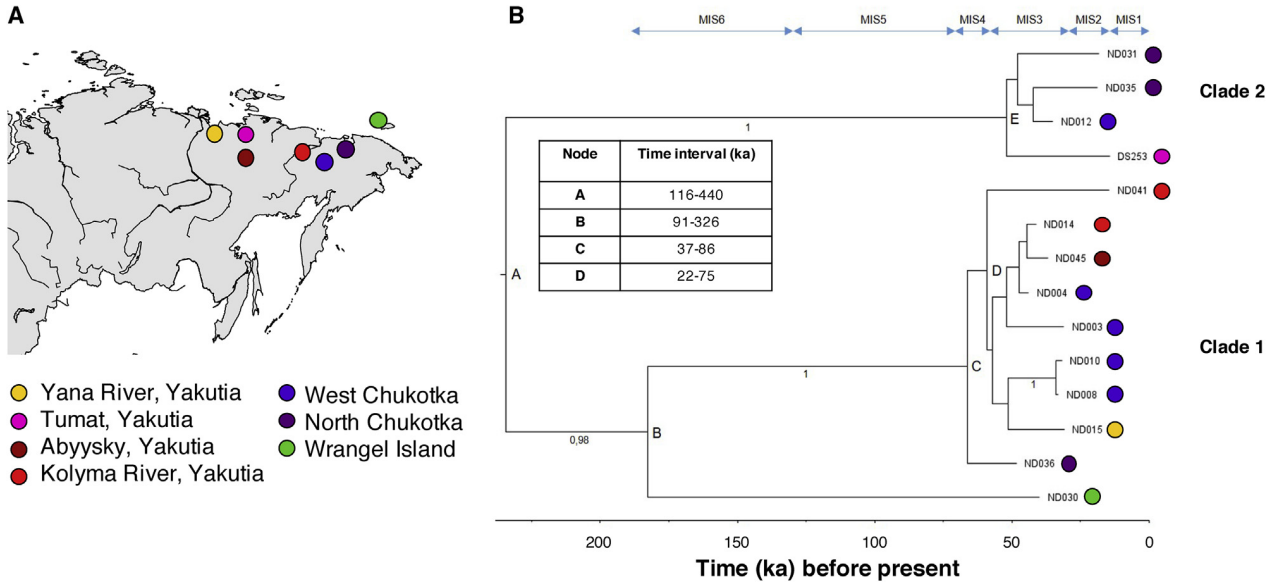

Figure 1. Sampling Locations and Bayesian Phylogeny from a Constant Size Model Inferred with BEAST

(A) Map showing sampling locations in Siberia. The map was created using R [7].

(B) Bayesian phylogeny for mitochondrial genomes (16,438 bp), where posterior probability support values above 0.9 are shown. MIS1–MIS6 corresponds to Marine Isotope Stages. The estimated ages of nodes A–D are shown as 95% HPD ranges. Mitochondrial sequences for ND008 and ND010 were identical; however, the position of all terminal nodes were adjusted to show the calibrated age of each specimen, using BEAUti v1.10.4 [8], with the median value of dates listed in Data S1A.

See also Figure S1.

The observation that N_{ef} was higher than N_e during the Late Pleistocene (Figures 2 and S1B) could potentially be explained by male-biased dispersal and female philopatry. However, there is little evidence for sex-biased dispersal in extant rhinoceros (e.g., black rhinoceros *Diceros bicornis* [14]; white rhinoceros *Ceratotherium simum* [15]), which makes this explanation unlikely for woolly rhinoceros. Instead, we hypothesize that the comparatively high N_{ef} in the woolly rhinoceros is a consequence of a high variance in male reproductive success, similar to what has been reported for white rhinoceros [16]. Future analysis of nuclear data from multiple male and female woolly rhinoceros will be necessary to further explore this question and to provide genomic insights into aspects of its behavior.

The observed increase in N_e leading up to MIS 6 may signify a demographic expansion but could alternatively be attributed to population subdivision and the divergence of the two clades identified in the mitochondrial analyses. It is plausible that these clades formed in allopatry, possibly during an interglacial period, and that these populations subsequently expanded and merged during or after MIS 6, leading to the lack of phylogeographic structure observed in the mitochondrial data. Thus, the peak observed at MIS 6 may be an artifact of a population subdivision rather than an increase in population size as population structure is known to affect PSMC [17–19]. Following MIS 6, the effective population size decreased through the Eemian interglacial (~130–115 ka) and the beginning of the last glacial period, reaching a minimum N_e at ~33 ka BP.

Although PSMC has reduced power in estimating N_e during the 20 ka prior to the age of the sample [17], we observed an increase in N_e at 29.7 ka BP. While this increase is consistent with previous estimates based on short mitochondrial DNA sequences from woolly rhinoceros [2] and the diversification within each clade observed in our mitogenomes, it is in contrast to data from the woolly mammoth, which did not indicate an expansion in N_e at that time [20]. We hypothesize that the observed increase in N_e in the woolly rhinoceros may have been related to the transition from the climatically unstable MIS 3 to the more stable cold period of MIS 2 around 29 ka BP [21], which was a period suggested to have suitable habitat in northeastern Siberia for cold-adapted species [22, 23]. However, as the woolly rhinoceros experienced an increase in N_e, other cold-adapted taxa such as the woolly mammoth's N_e remained stable. MIS 2 may thus have provided a particularly suitable habitat for the woolly rhinoceros with glacial tundra-steppe conditions prevailing and allowing population expansion [24]. An alternative explanation may be that the increase in N_e represents the merging of populations as the range of the highly specialized grazers such as woolly rhinoceros [24] contracted toward northeastern Siberia [18], while the mammoth, which may have been ecologically more flexible as exemplified by its wider distribution [2, 25, 26], maintained a constant N_e.

Although tentative because of the limitations of PSMC analysis, our results suggest that the woolly rhinoceros' population size may have remained constant after the expansion 29.7 ka BP and until the death of the sequenced individual. Our mitochondrial data further supports a scenario of population stability until close to the extinction of the species (Figures 1B and S1B), since the two lineages identified here persisted until within ~300 years of the estimated extinction event at ~14 ka BP [1]. In spite of a progressive range contraction toward northeastern Siberia

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

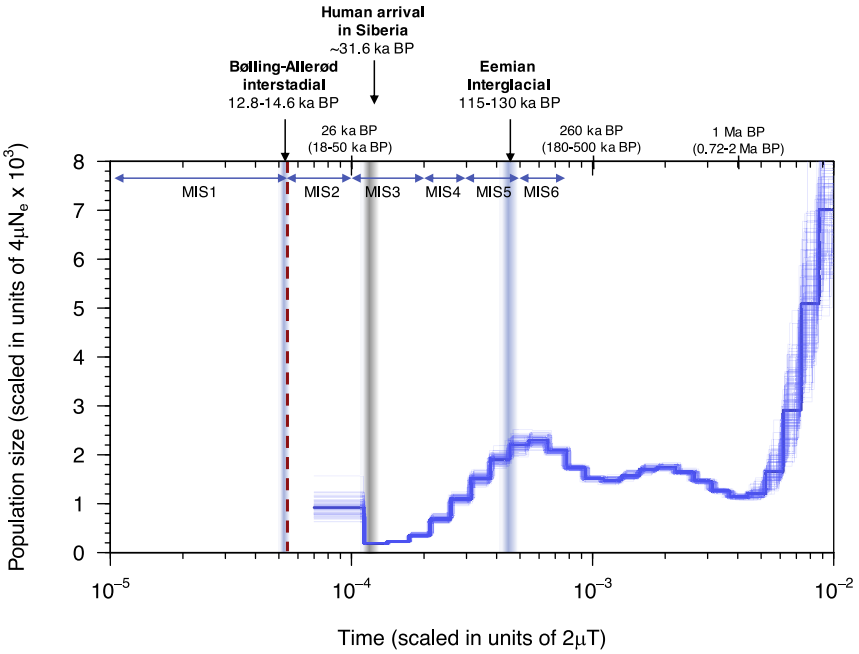

Figure 2. Temporal Changes in Woolly Rhinoceros Effective Population Size (N_e) Using the PSMC

Time is given in units of divergence per base pair on the lower x axis. The upper x axis corresponds to time in years BP assuming a substitution rate of 2.34×10^{-8} substitutions/site/generation [12] with the range given in parentheses taking into account the uncertainty of the rate estimate (see STAR Methods) and a generation time of 12 years [13]. Thin lines depict 100 bootstrap replicates for specimen ND035 (18,530 ± 170 cal BP). The y axis corresponds to the effective population size (N_e). MIS1–MIS6 corresponds to Marine Isotope Stages. The vertical red line depicts the approximate extinction of woolly rhinoceros at ~14 ka BP. The blue bars depict the Eemian interglacial and Bølling-Allerød interstadial. The gray bar represents approximate first human arrival in northeastern Siberia [4].

starting 35 ka BP, the fossil record indicates that the species was still widespread up until 18.5 ka BP [1], which may explain why our population size estimates remained constant. Interestingly, data from several other mammals highlight the importance of northeastern Siberia as a late glacial refugium. For example, recent analyses indicate that extant wolf lineages originated in northeastern Siberia [27], and it has been hypothesized that admixing of modern human populations occurred in the area, prior to the colonization of North America (e.g., [4, 28, 29]). Similarly, horse, bison, and collared lemming also have highly diverged mitochondrial lineages that survived in northeastern Siberia after the Last Glacial Maximum [30–32], suggesting long-term population continuity for several taxa in this region.

Genomic Diversity and Extinction

The woolly rhinoceros genome had an average heterozygosity of approximately 1.7 heterozygous sites per 1,000 bp (95% CI: 1.66–1.74). This is higher than the genomic diversity observed in a previously published mainland mammoth genome (1.25 heterozygous sites per 1,000 bp [20]), as well as the extant Sumatran rhinoceros (1.3 heterozygous sites per 1,000 bp [12]) and Northern and Southern White rhinoceros (1.1 and 0.9 heterozygous sites per 1,000 bp, respectively [33]). Based on identification of runs of homozygosity (ROH), we estimated the inbreeding coefficient (F_{ROH}) to be 5.9% when considering ROH regions >0.5 Mb. Furthermore, 96% of the ROH were <0.5 Mb in length, and the maximum ROH length was 2.5 Mb (Figures 3 and S2). This result was consistent when using a range of less stringent parameters, accounting for any remaining DNA damage after USER treatment (see STAR Methods; Figure S2). This observed level of inbreeding is comparatively low and, for example, on par with non-African human populations [34]. We note, however, that the level of inbreeding is higher than that observed in a Late Pleistocene mainland mammoth (F_{ROH}: 0.83%) and indicates some degree of background relatedness from mating between distant relatives [35], potentially because of higher population substructure and/or reduced local population size at the time compared with the mammoth. However, this result is in stark contrast to a 4.3 ka BP mammoth from Wrangel Island (F_{ROH}: 23.3%), which showed increased inbreeding associated with long-term small population size [20, 36].

Taken together, our analyses of nuclear and mitochondrial genomic diversity in the woolly rhinoceros provide no evidence for a decline in population size preceding the samples analyzed here, nor any indication of elevated inbreeding typical of small populations. While we cannot exclude the role of humans in woolly rhinoceros' extinction, our results imply that the arrival of anatomically modern humans in northeastern Siberia was not correlated with a demographic decline in the woolly rhinoceros. However, we caution that the earliest evidence of human presence in northeastern Siberia, dated to 31.6 ka BP [4], may represent temporary settlements [37] and that currently there is only evidence of sporadic human presence through MIS 3–2 [38], thus humans may have only had a limited negative impact on woolly rhinoceros populations.

Overall, our findings of a stable population size until at least 18.5 ka suggest that the final decline toward extinction was rapid and started within the 4,500 years prior to the extinction (i.e., after the death of the individual whose genome was sequenced here). This severe and rapid demographic decline, which based on radiocarbon evidence [1] likely coincided with the Bølling-Allerød interstadial (14.6–12.8 ka), could imply that the extinction of woolly rhinoceros was primarily driven by the changes in climate and vegetation characteristic of the period [22]. Across Eurasia, the Bølling-Allerød interstadial was characterized by an increase in forest habitats and woody plant cover [9]. Stuart and Lister [1] previously suggested that the replacement of low-growing vegetation by shrub-tundra and tree biomes (e.g., *Salix sp.*, *Betula sp.*) in Siberia during the warm Bølling-Allerød interstadial [9, 23], combined with increased snowfall [39], likely led to the

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

CellPress
OPEN ACCESS

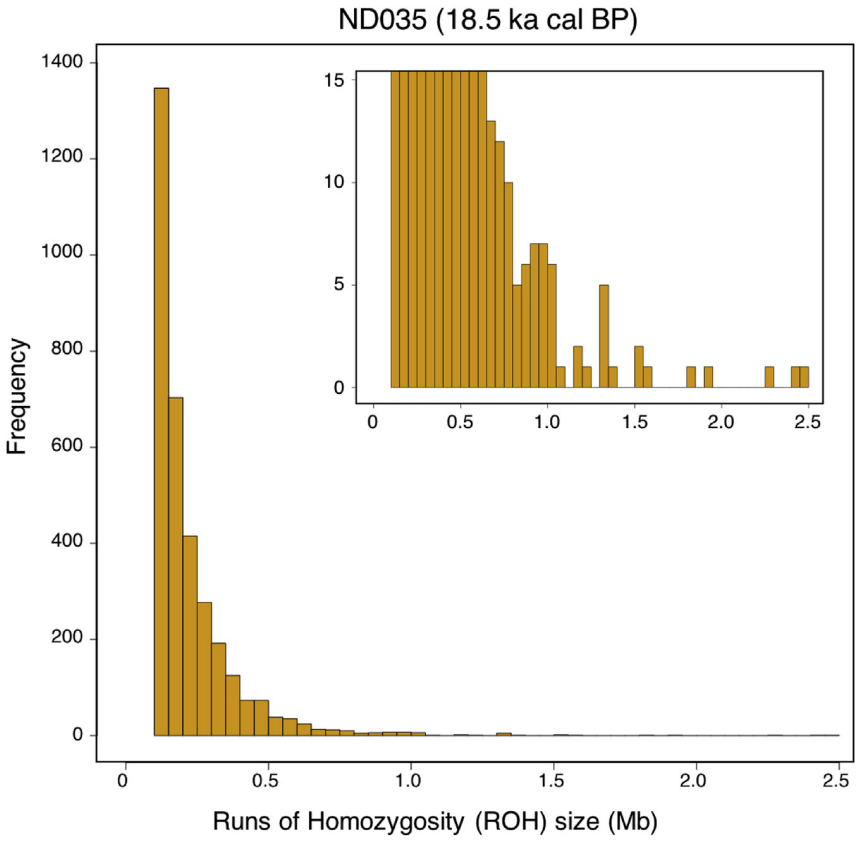

Figure 3. Frequency Distribution of Runs of Homozygosity (ROH) Size in One Woolly Rhinoceros

The specimen (ND035) was dated to 18,530 ± 170 cal BP. Only ROH ≥0.1 Mb are shown. Inset shows the magnification of ROH for clarity. See also Figure S2.

extinction of the woolly rhinoceros. Additional sequencing of individuals closer to the extinction event will be needed to gain a better understanding of the timing and rate of decline toward extinction.

Adaptation to Cold Environments

We undertook a preliminary evaluation of adaptations in woolly rhinoceros relative to Sumatran rhinoceros by examining non-synonymous mutations (i.e., missense; loss of function, LoF) across 19,556 coding genes. Overall, we found 1,524 identifiable genes with non-synonymous mutations (*n* missense = 1386, *n* LoF = 138; Data S1D) associated with biological processes including cellular component organization or biogenesis, cellular process, localization, reproduction, biological regulation, response to stimulus, developmental processes, and metabolic processes, several of which are significantly overrepresented (Data S1E and S1F). In contrast to previous analyses of another cold-adapted megafaunal species, the woolly mammoth [40], we did not observe non-synonymous variants in genes associated with fat deposition and changes to circadian rhythm thought to have played a role in mammoth adaptation to the arctic environment. However, in 89 genes, the woolly mammoth and the woolly rhinoceros both had non-synonymous variants potentially indicative of positive selection, including in TRPA1 (Transient Receptor Potential subfamily A; Data S1G and S1H), which is known to be involved in adaptation to cold tolerance [41, 42]. Variants undergoing positive selection in genes encoding TRP channels, including TRPA1, have recently been described in a range of cold-adapted taxa [40, 43]. Furthermore, there was one gene (KCNK17, potassium channel subfamily K) in which both species have a LoF mutation. KCNK17 is a paralog of KCNK4 (also known as TRAAK, TWIK-Related Arachidonic Acid-Stimulated Potassium Channel Protein), which under normal function has been shown to silence TRP proteins including TRPA1 and TRPM8 [42]. Thus, this gene is involved in cold temperature perception and, when knocked out, may play a role in cold adaptation [42].

To further identify genes that may have been of adaptive significance in the woolly rhinoceros, we ranked all identified missense mutations (n = 17,888) according to three indices (amino acid index [aal], experimental exchangeability, and Sneath's amino acid dissimilarity) in order to evaluate their impact on protein structure and physicochemical properties (Data S1C; see STAR Methods). The results showed that all three indices gave similar results (Data S1C). The distribution of aal was bi-modal, with the majority of mutations predicted to cause mostly weak to moderate changes in protein structure (Figure S3). However, there were 284 variants with an aal of 1, indicating maximal change of amino acid physicochemical properties. Of these variants, 83 were found across 41 different olfactory receptor genes (Data S1C), which is consistent with frequent gene gains and losses during the evolution of this gene family [44].

Conclusions

Our analyses of genomic diversity have several implications for understanding the population history and biology of the woolly rhinoceros. First, the finding of deep divergence among mitochondrial lineages hints at a dynamic history during the Middle Pleistocene, possibly characterized by the fragmentation and subsequent merger of populations. Second, analyses of mitogenomes and the nuclear genome both suggest that the species' final decline toward extinction was rapid and did not begin until after 18.5 ka BP. This implies that the woolly rhinoceros N_e did not start to decline until approximately 13 ka after the first arrival of humans in northeastern Siberia [4, 28, 29]. This does not exclude the possibility that humans later contributed to their extinction. For instance, hunting of woolly rhinoceros by humans could have reduced the population growth rate and, thus, may have accelerated the extinction of the species. However, given the data at hand, it appears likely that changes in the

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

environment, associated with the onset of the Bølling-Allerød interstadial, were the primary drivers of the woolly rhinoceros' extinction. It should be possible to further investigate the extent to which the final demographic decline coincided with the Bølling-Allerød by analyzing additional genomes from the time period 18–14 ka BP. Finally, our preliminary assessment of adaptive genetic variation in the woolly rhinoceros identified a range of non-synonymous changes in genes associated with several biological processes, including a gene (TRPA1) involved in temperature sensation. Taken together, these findings highlight the utility of genomic data in unraveling previously unknown evolutionary processes in extinct species and illustrate the need to investigate demographic trajectories in other megafauna to develop a better understanding of the timing and rate of demographic change during the Late Quaternary.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - DNA extraction
  - Library preparation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - De-novo reference assembly and annotation
  - Estimation of endogenous DNA content
  - Mitogenome data processing
  - Nuclear genome data processing
  - Mitogenome data analysis
  - Demographic reconstruction
  - Heterozygosity and inbreeding
  - Non-synonymous mutations

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.07.046.

ACKNOWLEDGMENTS

We thank Anthony Stuart for providing fossil samples. The authors are grateful to Alexander Banderov, Semen Ivanov, Klimovsky Aisen, and several others for help with collection of samples in the field. We acknowledge support from the Uppsala Multidisciplinary Centre for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. We also acknowledge Tom van der Valk for sharing a script for CpG sites filtering and Shyam Gopalakrishnan for sharing a demultiplex script of BGI raw data. Sequencing was performed by the Swedish National Genomics Infrastructure (NGI) at the Science for Life Laboratory (Illumina data), which is supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation, and at the China National Genebank (BGISeq data). We thank the Sabah Biodiversity Centre for allowing us to export samples from Sumatran rhinoceros individuals from Sabah (License Ref JKM/MBS.1000-2/3 JLD.2 (32)). We also thank Dr. Zainal Zainuddin and Dr. John Payne from Borneo Rhino Alliance (BORA) and the late Dr. Diana Angeles Ramirez Saldivar from the Wildlife Rescue Unit for providing samples from the Sumatran rhinoceros individual used for the reference genome assembly. This work was supported by Formas (2015-676 and 2018-01640) and the Bolin Centre for Climate Research to L.D., the Swiss National Science Foundation (P2SKP3_165031 and P300PA_177845) and the Carl Tryggers Foundation (CTS 19:257) to N.D., the Carl Tryggers Foundation (CTS 17:109) to D.D.-d.-M., and an ERC Consolidator Award 681396-Extinction Genomics to M.T.P.G. M.K. is financially supported by the Knut and Alice Wallenberg Foundation as part of the National Bioinformatics Infrastructure Sweden at SciLifeLab.

AUTHOR CONTRIBUTIONS

E.L., N.D., and L.D. designed the research. E.L., N.D., M.K., D.D.-d.-M., O.A.R., D.W.G.S., M.T.P.G., F.S.-B., G.Z., M.-H.S.S., E.D.L., E.W., A.P., F.S., S.F., H.B., S.K.S.S.N., B.G., J.v.d.P., Y.L.C., S.P., O.P., I.K., A.M.L, P.D.H., J.D.K., B.S., S.V., A.G., and L.D. performed research. E.L., N.D., D.W.G.S., Y.L.C., and F.S.-B. undertook laboratory work. E.L., N.D., M.K., and D.D.-d.-M. analyzed data. S.F., I.K., A.M.L., S.V., E.D.L., E.W., and B.S. provided samples. E.L. and N.D. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 14, 2020
Revised: June 18, 2020
Accepted: July 14, 2020
Published: August 13, 2020

REFERENCES

1. Stuart, A.J., and Lister, A.M. (2012). Extinction chronology of the woolly rhinoceros *Coelodonta antiquitatis* in the context of late Quaternary megafaunal extinctions in northern Eurasia. Quat. Sci. Rev. *51*, 1–17.
2. Lorenzen, E.D., Nogués-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K.A., Ugan, A., Borregaard, M.K., Gilbert, M.T.P., Nielsen, R., et al. (2011). Species-specific responses of Late Quaternary megafauna to climate and humans. Nature *479*, 359–364.
3. Kuzmin, Y.V. (2010). Extinction of the woolly mammoth (*Mammuthus primigenius*) and woolly rhinoceros (*Coelodonta antiquitatis*) in Eurasia: Review of chronological and environmental issues. Boreas *39*, 247–261.
4. Sikora, M., Pitulko, V.V., Sousa, V.C., Allentoft, M.E., Vinner, L., Rasmussen, S., Margaryan, A., de Barros Damgaard, P., de la Fuente, C., Renaud, G., et al. (2019). The population history of northeastern Siberia since the Pleistocene. Nature *570*, 182–188.
5. Orlando, L., Leonard, J.A., Thenot, A., Laudet, V., Guerin, C., and Hänni, C. (2003). Ancient DNA analysis reveals woolly rhino evolutionary relationships. Mol. Phylogenet. Evol. *28*, 485–499.
6. Willerslev, E., Gilbert, M.T.P., Binladen, J., Ho, S.Y.W., Campos, P.F., Ratan, A., Tomsho, L.P., da Fonseca, R.R., Sher, A., Kuznetsova, T.V., et al. (2009). Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. BMC Evol. Biol. *9*, 95.
7. R Core Team (2017). R: A Language and Environment for Statistical Computing. (R Found Statistical Computing).
8. Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. *4*, vey016.
9. Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J.O., Andreev, A., Bezrukova, E., Blyakharchuk, T., Jankovska, V., et al. (2017). Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. Quat. Sci. Rev. *157*, 80–97.
10. Hoffecker, J.F., Elias, S.A., and Potapova, O. (2020). Arctic Beringia and Native American Origins. PaleoAmerica *6*, 158–168.

11. Palkopoulou, E., Dalén, L., Lister, A.M., Vartanyan, S., Sablin, M., Sher, A., Edmark, V.N., Brandström, M.D., Germonpré, M., Barnes, I., and Thomas, J.A. (2013). Holarctic genetic structure and range dynamics in the woolly mammoth. Proc. Biol. Sci. *280*, 20131910.

12. Mays, H.L., Jr., Hung, C.-M.M., Shaner, P.-J.J., Denvir, J., Justice, M., Yang, S.-F.F., et al. (2018). Genomic Analysis of Demographic History and Ecological Niche Modeling in the Endangered Sumatran Rhinoceros *Dicerorhinus sumatrensis*. Curr. Biol. *28*, 70–76.

13. Roth, T.L., Reinhart, P.R., Romo, J.S., Candra, D., Suhaery, A., and Stoops, M.A. (2013). Sexual maturation in the Sumatran rhinoceros (*Dicerorhinus sumatrensis*). Zoo Biol. *32*, 549–555.

14. Law, P.R., Fike, B., and Lent, P.C. (2014). Birth sex in an expanding black rhinoceros (*Diceros bicornis minor*) population. J. Mammal. *95*, 349–356.

15. Shrader, A., and Owen-Smith, N. (2002). The role of companionship in the dispersal of white rhinoceroses (*Ceratotherium simum*). Behav. Ecol. Sociobiol. *52*, 255–261.

16. Kretzschmar, P., Auld, H., Boag, P., Ganslo\u00dfer, U., Scott, C., Van Coeverden de Groot, P.J., and Courtiol, A. (2019). Mate choice, reproductive success and inbreeding in white rhinoceros: New insights for conservation management. Evol. Appl. *13*, 699–714.

17. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496.

18. Mazet, O., Rodríguez, W., Grusea, S., Boitard, S., and Chikhi, L. (2016). On the importance of being structured: instantaneous coalescence rates and human evolution–lessons for ancestral population size inference? Heredity *116*, 362–371.

19. Mather, N., Traves, S.M., and Ho, S.Y.W. (2019). A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data. Ecol. Evol. *10*, 579–589.

20. Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., et al. (2015). Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. Curr. Biol. *25*, 1395–1400.

21. Lisiecki, L.E., and Raymo, M.E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic δ 18O records. Paleoceanography *20*, PA1003.

22. Allen, J.R.M., Hickler, T., Singarayer, J.S., Sykes, M.T., Valdes, P.J., and Huntley, B. (2010). Last glacial vegetation of northern Eurasia. Quat. Sci. Rev. *29*, 2604–2618.

23. Andreev, A.A., Schirrmeister, L., Tarasov, P.E., Ganopolski, A., Brovkin, V., Siegert, C., Wetterich, S., and Hubberten, H.-W. (2011). Vegetation and climate history in the Laptev Sea region (Arctic Siberia) during Late Quaternary inferred from pollen records. Quat. Sci. Rev. *30*, 2182–2199.

24. Kahlke, R.-D., and Lacombat, F. (2008). The earliest immigration of woolly rhinoceros (*Coelodonta tologoijensis*, Rhinocerotidae, Mammalia) into Europe and its adaptive evolution in Palaearctic cold stage mammal faunas. Quat. Sci. Rev. *27*, 1951–1961.

25. Lister, A.M., and Sher, A.V. (2015). Evolution and dispersal of mammoths across the Northern Hemisphere. Science *350*, 805–809.

26. Puzachenko, A.Y., Markova, A.K., Kosintsev, P.A., van Kolfschoten, T., van der Plicht, J., Kuznetsova, T.V., Tikhonov, A.N., Ponomarev, D.V., Kuitems, M., and Bachura, O.P. (2017). The Eurasian mammoth distribution during the second half of the Late Pleistocene and the Holocene: Regional aspects. Quat. Int. *445*, 71–88.

27. Loog, L., Thalmann, O., Sinding, M.S., Schuenemann, V.J., Perri, A., Germonpré, M., Bocherens, H., Witt, K.E., Samaniego Castruita, J.A., Velasco, M.S., et al. (2020). Ancient DNA suggests modern wolves trace their origin to a Late Pleistocene expansion from Beringia. Mol. Ecol. *29*, 1596–1610.

28. Graf, K.E., and Buvit, I. (2017). Human Dispersal from Siberia to Beringia: Assessing a Beringian Standstill in Light of the Archaeological Evidence. Curr. Anthropol. *58*, S583–S603.

29. Vachula, R.S., Huang, Y., Longo, W.M., Dee, S.G., Daniels, W.C., and Russell, J.M. (2019). Evidence of Ice Age humans in eastern Beringia suggests early migration to North America. Quat. Sci. Rev. *205*, 35–44.

30. Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K.A.S., Albizuri, S., et al. (2019). Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. Cell *177*, 1419–1435.

31. Kirillova, I.V., Zanina, O.G., Chernova, O.F., Lapteva, E.G., Trofimova, S.S., Lebedev, V.S., Tiunov, A.V., Soares, A.E.R., Shidlovskiy, F.K., and Shapiro, B. (2015). An ancient bison from the mouth of the Rauchua River (Chukotka, Russia). Quat. Res. *84*, 232–245.

32. Fedorov, V.B., Trucchi, E., Goropashnaya, A.V., Waltari, E., Whidden, S.E., and Stenseth, N.C. (2020). Impact of past climate warming on genomic diversity and demographic history of collared lemmings across the Eurasian Arctic. Proc. Natl. Acad. Sci. USA *117*, 3026–3033.

33. Tunstall, T., Kock, R., Vahala, J., Diekhans, M., Fiddes, I., Armstrong, J., Paten, B., Ryder, O.A., and Steiner, C.C. (2018). Evaluating recovery potential of the northern white rhinoceros from cryopreserved somatic cells. Genome Res. *28*, 780–788.

34. Kirin, M., McQuillan, R., Franklin, C.S., Campbell, H., McKeigue, P.M., and Wilson, J.F. (2010). Genomic runs of homozygosity record population history and consanguinity. PLoS ONE *5*, e13996.

35. Pemberton, T.J., Absher, D., Feldman, M.W., Myers, R.M., Rosenberg, N.A., and Li, J.Z. (2012). Genomic patterns of homozygosity in worldwide human populations. Am. J. Hum. Genet. *91*, 275–292.

36. Rogers, R.L., and Slatkin, M. (2017). Excess of genomic defects in a woolly mammoth on Wrangel island. PLoS Genet. *13*, e1006601.

37. Pitulko, V., Nikolskiy, P., Basilyan, A., and Pavlova, E. (2013). Human habitation in arctic western Beringia prior to the LGM. Paleoamerican odyssey, pp. 13–44.

38. Pitulko, V., Pavlova, E., and Nikolskiy, P. (2017). Revising the archaeological record of the Upper Pleistocene Arctic Siberia: Human dispersal and adaptations in MIS 3 and 2. Quat. Sci. Rev. *165*, 127–148.

39. Sher, A.V. (1997). Late-Quaternary extinction of large mammals in northern Eurasia:A new look at the Siberian contribution. Past and Future Rapid Environmental Changes (Springer), pp. 319–339.

40. Lynch, V.J., Bedoya-Reina, O.C., Ratan, A., Sulak, M., Drautz-Moses, D.I., Perry, G.H., Miller, W., and Schuster, S.C. (2015). Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. Cell Rep. *12*, 217–228.

41. Chen, J., Kang, D., Xu, J., Lake, M., Hogan, J.O., Sun, C., Walter, K., Yao, B., and Kim, D. (2013). Species differences and molecular determinant of TRPA1 cold sensitivity. Nat. Commun. *4*, 2501.

42. Karashima, Y., Talavera, K., Everaerts, W., Janssens, A., Kwan, K.Y., Vennekens, R., Nilius, B., and Voets, T. (2009). TRPA1 acts as a cold sensor in vitro and in vivo. Proc. Natl. Acad. Sci. USA *106*, 1273–1278.

43. Cardona, A., Pagani, L., Antao, T., Lawson, D.J., Eichstaedt, C.A., Yngvadottir, B., Shwe, M.T.T., Wee, J., Romero, I.G., Raj, S., et al. (2014). Genome-wide analysis of cold adaptation in indigenous Siberian populations. PLoS ONE *9*, e98076.

44. Niimura, Y., Matsui, A., and Touhara, K. (2014). Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. Genome Res. *24*, 1485–1496.

45. Meyer, M., and Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harbor Protoc. *2010*, db.prot5448.

46. Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S.S.T., Sinding, M.H.S., Samaniego, J.A., Wales, N., Sicheritz-Pontén, T., and Gilbert, M.T.P. (2018). Single-tube library preparation for degraded DNA. Methods Ecol. Evol. *9*, 410–419.

47. Mak, S.S.T., Gopalakrishnan, S., Carøe, C., Geng, C., Liu, S., Sinding, M.S., Kuderna, L.F.K., Zhang, W., Fu, S., Vieira, F.G., et al. (2017). Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

sequencing platforms for palaeogenomic sequencing. Gigascience *6*, 1–13.

48. Reimer, P.J., Bard, E., Bayliss, A., Warren Beck, J., Blackwell, P.G., Ramsey, C.B., Buck, C.E., Cheng, H., Lawrence Edwards, R., Friedrich, M., et al. (2013). IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. Radiocarbon *55*, 1869–1887.
49. Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I.A., Belmonte, M.K., Lander, E.S., Nusbaum, C., and Jaffe, D.B. (2008). ALLPATHS: de novo assembly of whole-genome shotgun microreads. Genome Res. *18*, 810–820.
50. Putnam, N.H., O'Connell, B.L., Stites, J.C., Rice, B.J., Blanchette, M., Calef, R., Troll, C.J., Fields, A., Hartley, P.D., Sugnet, C.W., et al. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. *26*, 342–350.
51. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.
52. Neethiraj, R., Hornett, E.A., Hill, J.A., and Wheat, C.W. (2017). Investigating the genomic basis of discrete phenotypes using a Pool-Seq-only approach: New insights into the genetics underlying colour variation in diverse taxa. Mol. Ecol. *26*, 4990–5002.
53. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.
54. Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. *44* (D1), D286–D293.
55. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.
56. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
57. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.
58. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
59. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684.
60. Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics *32*, 292–294.
61. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.
62. Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., and Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol. Biol. Evol. *34*, 3299–3302.
63. Leigh, J.W., and Bryant, D. (2015). PopART: Full-feature software for haplotype network construction. Methods Ecol. Evol. *6*, 1110–1116.
64. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods *9*, 772.
65. Rambaut, A. (2007). FigTree, a graphical viewer of phylogenetic trees. http://tree.bio.ed.ac.uk/software/figtree.
66. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. *67*, 901–904.
67. Haubold, B., Pfaffelhuber, P., and Lynch, M. (2010). mlRho - a program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. Mol. Ecol. *19* (*Suppl 1*), 277–284.
68. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
69. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.
70. Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nat. Protoc. *8*, 1551–1566.
71. Ramsey, C.B., and Lee, S. (2013). Recent and Planned Developments of the Program OxCal. Radiocarbon *55*, 720–730.
72. Yang, D.Y., Eng, B., Waye, J.S., Christopher Dudar, J., and Saunders, S.R. (1998). Improved DNA extraction from ancient bones using silica-based spin columns. Am J Phys Anthropol. *105*, 539–543.
73. Brace, S., Palkopoulou, E., Dalén, L., Lister, A.M., Miller, R., Otte, M., Germonpré, M., Blockley, S.P.E., Stewart, J.R., and Barnes, I. (2012). Serial population extinctions in a small mammal indicate Late Pleistocene ecosystem instability. Proc. Natl. Acad. Sci. USA *109*, 20532–20536.
74. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., et al. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. USA *110*, 15758–15763.
75. Gilbert, M.T.P., Tomsho, L.P., Rendulic, S., Packard, M., Drautz, D.I., Sher, A., Tikhonov, A., Dalén, L., Kuznetsova, T., Kosintsev, P., et al. (2007). Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. Science *317*, 1927–1930.
76. Knapp, M., Clarke, A.C., Horsburgh, K.A., and Matisoo-Smith, E.A. (2012). Setting the stage - building and working in an ancient DNA laboratory. Ann. Anat. *194*, 3–6.
77. Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res. *38*, e87.
78. Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. *40*, e3.
79. Heintzman, P.D., Zazula, G.D., Cahill, J.A., Reyes, A.V., MacPhee, R.D.E., and Shapiro, B. (2015). Genomic Data from Extinct North American Camelops Revise Camel Evolutionary History. Mol. Biol. Evol. *32*, 2433–2440.
80. Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., and Pachter, L. (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. *12*, R22.
81. Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.
82. Steiner, C.C., Houck, M.L., and Ryder, O.A. (2018). Genetic variation of complete mitochondrial genome sequences of the Sumatran rhinoceros (*Dicerorhinus sumatrensis*). Conserv. Genet. *19*, 397–408.
83. Goto, H., Ryder, O.A., Fisher, A.R., Schultz, B., Kosakovsky Pond, S.L., Nekrutenko, A., and Makova, K.D. (2011). A massively parallel sequencing approach uncovers ancient origins and high genetic variability of endangered Przewalski's horses. Genome Biol. Evol. *3*, 1096–1106.
84. Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., et al. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature *499*, 74–78.

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

85. Lynch, M. (2008). Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. Mol. Biol. Evol. *25*, 2409–2419.

86. Kawashima, S., Ogata, H., and Kanehisa, M. (1999). AAindex: Amino Acid Index Database. Nucleic Acids Res. *27*, 368–369.

87. Rudnicki, W.R., and Komorowski, J. (2004). Feature Synthesis and Extraction for the Construction of Generalized Properties of Amino Acids. Rough Sets and Current Trends in Computing (Springer), pp. 786–791.

88. Li, X., Kierczak, M., Shen, X., Ahsan, M., Carlborg, O., and Marklund, S. (2013). PASE: a novel method for functional prediction of amino acid substitutions based on physicochemical properties. Front. Genet. *4*, 21.

89. Sneath, P.H.A. (1966). Relations between chemical structure and biological activity in peptides. J. Theor. Biol. *12*, 157–195.

90. Yampolsky, L.Y., and Stoltzfus, A. (2005). The exchangeability of amino acids in proteins. Genetics *170*, 1459–1472.

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
EDTA	ThermoFisher Scientific	Cat#15575020
UREA	VWR	Cat#443874G
DTT	ThermoFisher Scientific	Cat#R0861
Tango Buffer (10X)	ThermoFisher Scientific	Cat#BY5
ATP (100mM)	ThermoFisher Scientific	Cat#R0441
T4 Polynucleotide Kinase (10U/ul)	ThermoFisher Scientific	Cat#EK0032
T4 DNA Polymerase 5U/ul	ThermoFisher Scientific	Cat#EP0062
USER Enzyme	NEB	Cat#M5505L
T4 DNA Ligase (5U/ul)	ThermoFisher Scientific	Cat#EL0011
Bst Polymerase	NEB	Cat#M0275S
AccuPrime Pfx	Life Technologies	Cat#12344-024
T4 DNA ligase (400U/ul)	NEB	Cat#M0202S
T4 DNA polymerase (3U/ul)	NEB	Cat#M0203S
Bst 2.0 Warmstart Polymerase (8U/ul)	NEB	Cat#M0538S
NaCl 5M solution	Sigma-Aldrich	Cat#59222C-500ML
PEG-8000	Sigma-Aldrich	Cat#89510-250G-F
BSA (20 mg/mL)	NEB	Cat#B9000S
Critical Commercial Assays		
High Sensitivity DNA kit	Agilent	Cat#5067-4626
KingFisher™ Cell and Tissue DNA Kit	ThermoFisher Scientific	Cat#97030196
Deposited Data		
Raw fastq reads (mitochondrial and nuclear data)	This study	ENA study accession number PRJEB35556
de-novo assembly for *Dicerorhinus sumatrensis*	This study	GenBank: JABWHU000000000
Oligonucleotides		
IS1 adaptor P5: 5′->3′ A*C*A*C*TCTTTCCCTACACGACGCTCTTC CG*A*T*C*T	[45]	Sigma-Aldrich
IS2 adaptor P7: 5′->3′ G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTC CG*A*T*C*T	[45]	Sigma-Aldrich
IS3 adaptor P5+P7: 5′->3′ A*G*A*T*CGGAA*G*A*G*C	[45]	Sigma-Aldrich
Illumina AmplifyingPrimer IS4: 5′->3′ AATGATACGGCGACCACCGAGATCTACACTC TTTCCCTACACGACGCTCTT	[45]	Sigma-Aldrich
Illumina Indexing Primer: 5′->3′ CAAGCAGAAGACGGCATACGAGATNNNNN NNGTGACTGGAGTTCAGACGTGT Ns represent indexes	[45]	Sigma-Aldrich
IS3 ATDC3 adaptor: 5′->3′ G*A*T*C*GGAA*G*A*G*C[C3spacer]	[46]	Sigma-Aldrich

(*Continued on next page*)

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

CellPress
OPEN ACCESS

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BGISEQ adaptor AD1_Long: 5′->3′ AAGCAGAAGACGGCATACGAGATGTT GTCTTCCTAAGACCGCTTGGCCTCCGACTT	[47]	Sigma-Aldrich
BGISEQ adaptor AD1_Short: 5′->3′ AAGTCGGAGGCC	[47]	Sigma-Aldrich
BGISEQ adaptor AD2_Long: 5′->3′ TTGTCTTCCTAAGGAACGACATGGCT ACGATCCGACTT	[46]	Sigma-Aldrich
BGISEQ adaptor AD2_Short: 5′->3′ AAGTCGGATCGT	[46]	Sigma-Aldrich
BGISEQ Indexing primer: 5′->3′ TGTGAGCCAAGGAGTTGNNNNNNNN NNTTGTCTTCCTAAGACCGC Ns represent indexes	[46]	Sigma-Aldrich
Common amplifying primer BGI forward: 5′->3′ GAACGACATGGCTACGA	[46]	Sigma-Aldrich
Software and Algorithms		
OxCal v4.3	[48]	https://c14.arch.ox.ac.uk/oxcal.html
Allpaths v.2.0	[49]	ftp://ftp.broadinstitute.org/pub/crd/ALLPATHS/Release-LG/
HiRise pipeline	[50]	Dovetail Genomics
BLAST+ 2.5.0	[51]	NCBI
MESPA pipeline	[52]	https://sourceforge.net/projects/mespa/
cufflinks v 2.2.1	[53]	http://cole-trapnell-lab.github.io/cufflinks/
eggNOG-mapper v4.5.1	[54]	http://eggnog-mapper.embl.de/
bcl2Fastq v1.8.3	Illumina	https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html
Custom BGI demultiplexing script	Shyam Gopalakrishnan	https://github.com/shyamsg/SantasHelpers/
SeqPrep	John St. John	https://github.com/jstjohn/SeqPrep
BWA v0.7.13	[55]	http://bio-bwa.sourceforge.net/
SAMtools v1.3	[56]	https://sourceforge.net/projects/samtools/files/samtools/1.3/
Geneious® v7.0.336	[57]	https://www.geneious.com/
Picard v1.141	Broad Institute	http://broadinstitute.github.io/picard
GATK v3.4.0	[58]	https://gatk.broadinstitute.org/hc/en-us
Mapdamage v2.0	[59]	https://ginolhac.github.io/mapDamage/
Qualimap v2.2.1	[60]	http://qualimap.bioinfo.cipf.es/
BEDtools v2.29.2	[61]	https://bedtools.readthedocs.io/en/latest/
DnaSP6 v6.12.03	[62]	http://www.ub.edu/dnasp/
PopArt	[63]	http://popart.otago.ac.nz/index.shtml
BEAST Software v1.10.4	[8]	https://beast.community/
jModelTest v2.1.9	[64]	https://github.com/ddarriba/jmodeltest2
Figtree v1.4.4	[65]	http://tree.bio.ed.ac.uk/software/figtree/
Tracer v1.7.1	[66]	https://github.com/beast-dev/tracer/releases/tag/v1.7.1
PSMC v0.6.5	[17]	https://github.com/lh3/psmc
mlRho v2.7	[67]	http://guanine.evolbio.mpg.de/mlRho/
PLINK v1.9	[68]	https://www.cog-genomics.org/plink2
SNPeff v4.3	[69]	http://snpeff.sourceforge.net/index.html

(*Continued on next page*)

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
simpred	NBIS	https://github.com/NBISweden/simpred
Panther	70	http://www.pantherdb.org/
Other		
Proteinase K	VWR	Cat#1.24568.0100
dNTPs	VWR	Cat#733-1854
QiaQuick PCR purification Kit	QIAGEN	Cat#28106
Min Elute PCR purification Kit	QIAGEN	Cat#28006
Agencourt AmPure XP 5mL Kit	Beckman Coulter	Cat#63880

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for reagents and data may be directed to and will be fulfilled by the Lead Contact, Edana Lord (edana.lord@zoologi.su.se).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

Raw fastq reads for mitogenome and nuclear data are deposited at the European Nucleotide Archive (ENA; study accession number PRJEB35556). The *de-novo* assembly for *Dicerorhinus sumatrensis* is deposited on GenBank (accession: JABWHU000000000).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We obtained 12 bones, one mummified tissue biopsy, and one hair sample of woolly rhinoceros, which were radiocarbon dated to between 14,100 and > 50,000 cal BP from North-eastern Siberia (Data S1A). Radiocarbon dating was performed at the Oxford Radiocarbon Accelerator Unit (ORAU, OxA), Beta Analytics (Miami, FL), ETH Zürich, and the Center for Isotope Research of Groningen University (GrA). We calibrated all radiocarbon dates using the IntCal13 calibration curve [48] in OxCal v4.3 [71]. For the *de-novo* assembly, we obtained tissue and cell lines from one male Sumatran Rhinoceros (*Dicerorhinus sumatrensis*), called Kertam, that originated from Borneo.

METHOD DETAILS

DNA extraction

We extracted DNA from bone samples according to protocol C in Yang et al. [72] as modified in Brace et al. [73]. For the mummified tissue biopsy and hair samples, we extracted DNA following Dabney et al. [74], but substituted the digestion buffer and incubation temperature with that described in Gilbert et al. [75]. Appropriate precautions were taken to minimize the risk of contamination during the processing of ancient samples [76].

For the *de-novo* assembly, we extracted DNA from blood and cell lines from Kertam, using a Kingfisher robot (Thermo Fisher Scientific) and following the Kingfisher blood & tissue extraction protocol according to the manufacturer's instructions. Concentrations were measured using QuBit® 2.0 Fluorometer (Invitrogen, USA) and the quality of the DNA was evaluated by running the samples through agarose gels with electrophoresis.

Library preparation

Double stranded Illumina libraries were built for 14 extracts according to Meyer & Kircher [45], along with 2 extraction blanks. 20 μl of DNA extract was used in a 40 μl blunt-end repair reaction with the following final concentration: 1 × buffer Tango, 100 μM of each dNTP, 1 mM ATP, 25 U T4 polynucleotide kinase (Thermo Scientific) and 3U USER enzyme (New England Biolabs). A USER enzyme treatment was performed to excise uracil residues resulting from post-mortem damage [77, 78]. Samples were incubated for 3 h at 37°C, followed by the addition of 1 μl T4 DNA polymerase (Thermo Scientific) and incubation at 25°C for 15 min and 12°C for 5 min. The samples were then purified using MinElute spin columns following the manufacturer's protocol and eluted in 20ul EB buffer. Next, an adaptor ligation step was performed where DNA fragments within each library were ligated to a combination of incomplete, partially double-stranded P5- and P7-adapters (10 μM each). This reaction was performed in a 40 μl reaction volume using 20 μl of blunt-ended DNA library and 1 μl P5-P7 adaptor mix per sample with a final concentration of 1 × T4 DNA ligase buffer, 5% PEG-4000, 5U T4 DNA ligase (Thermo Scientific). Samples were incubated for 30 min at room temperature and cleaned using MinElute spin columns as described above.

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

Next, we performed an adaptor fill-in reaction in 40 μl final volume using 20 μl adaptor ligated DNA with a final concentration of 1 × Thermopol Reaction Buffer, 250 μM of each dNTP, 8U *Bst* Polymerase, Long Fragments. The libraries were incubated at 37°C for 20 min, and then heat-inactivated at 80°C for 20 min. These libraries were then used as stock for indexing PCR amplification for screening (i.e., estimation of endogenous DNA content of each sample) and deep-sequencing.

PCR amplifications were performed in 25 μl volumes with 3μl of adaptor-ligated library as template, with the following final concentrations: 1x AccuPrime reaction mix, 0.3μM IS4 amplification primer, 0.3μM P7 indexing primer, 7 U AccuPrime Pfx (Thermo Scientific) and the following cycling protocol: 95°C for 2 min, 12 cycles at 95°C for 30 s, 55°C for 30 s and 72°C for 1 min and a final extension at 72°C for 5 min. We used dual unique indexes of 6 bp for each library.

Purification and size selection of the libraries were performed using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA), first using a 0.5X bead:DNA ratio and second with a 1.8X bead:DNA ratio to remove long and short (i.e., adapter dimers) fragments, respectively. Library concentration was measured with a high-sensitivity DNA chip on a Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA). Finally, multiplexed libraries were pooled into a single pool in equimolar concentrations and sequenced on Illumina HiSeq2500 in High Output mode with a 2 × 125bp setup at Sci*Life*Lab, Stockholm. The nuclear genome sample, ND035, was deep sequenced on an Illumina HiSeqX with a 2 × 150 bp setup at Sci*Life*Lab, Stockholm.

For the hair sample, ND045, a double stranded library was constructed following the Meyer & Kircher protocol [45] as modified by Heintzman et al. [79]. We also built additional libraries for three samples (ND012, ND036, ND045) using the BEST2 library build protocol, a blunt-end, single tube library preparation procedure suitable for degraded DNA samples [46] and using custom-design adaptor oligos specific for the BGISEQ-500 Sequencing Platform [47]. No USER treatment was performed for these three libraries.

QUANTIFICATION AND STATISTICAL ANALYSIS

De-novo reference assembly and annotation

We generated a *de-novo* reference genome for one male Sumatran rhinoceros (*D. sumatrensis*) by sequencing a combination of Chicago, Hi-C, mate-pair and short insert libraries from a high molecular weight DNA extract from one male from the Bornean population (Kertam). An initial assembly based on the short insert and mate-pairs was done using Allpaths v.2.0 [49]. The final genome assembly was done using the HiRise pipeline (Dovetail Genomics, [50]) The final assembly size was of 2.4 Gb and comprised 1,763 scaffolds with an N50 of 62 Mb, where 99% of the genome was comprised within 44 scaffolds >1 Mb.

We identified X-linked scaffolds in the Sumatran rhinoceros genome in BLAST+ 2.5.0 [51] using the horse X chromosome as subject sequence. The BLAST+ parameters were set as: -evalue = 1e-10; -word_size = 15; -max_target_seqs = 1000. For all downstream analyses, we excluded two X chromosome-linked scaffolds (Sc9M7eS_1319;HRSCAF = 1962 and Sc9M7eS_931;HRSCAF = 1475) from the assembled genome.

We annotated the assembly using the MESPA pipeline [52]. We collapsed reference protein sets for white rhinoceros (*Ceratotherium simum simum*; GenBank: GCF_000283155.1) to 90% coverage following Uniprot90 guidelines where each protein cluster is composed of sequences with at least 90% sequence identity to, and 80% overlap with, the longest sequence using a custom script. In that way, we discarded isoforms of the reference datasets. We then used MESPA to extract the gene models in Sumatran rhinoceros with 90% length coverage to each set of reference proteins and to generate an annotation in gff format. We extracted 99% (21,953 out of 22,054) high quality protein models (i.e., aligning to 90% of their expected length) using white rhinoceros as a reference protein set.

We extracted the CDSs and protein sequences of this annotation with cufflinks v 2.2.1 [53, 80] gffread command for downstream analyses using the -V option to remove gene models with in-frame STOP codons. We retained 19,556 gene models with a mean length of 1,724 bp (Median = 1,051; min = 34; max = 26,418).

Finally, we performed a functional annotation of these gene models using the eggNOG-mapper v4.5.1 [54]. We used 'Mammals' as a taxonomic scope and the 'Restrict to one-to-one' and the 'Use experimental terms only' to prioritize precision and quality of matches.

Estimation of endogenous DNA content

Raw Illumina sequence data were demultiplexed based on their unique indices from Bcl to Fastq using bcl2Fastq v1.8.3 (CASAVA software suite) while raw BGI data was converted and samples were demultiplexed using a custom script (https://github.com/shyamsg/SantasHelpers/). We used SeqPrep (https://github.com/jstjohn/SeqPrep) to trim adapters and merge paired-end reads using default settings, with a minor modification in the source code that allowed us to choose the best quality scores of bases in the merged region instead of aggregating the scores, following [20]. Raw Illumina sequencing reads were aligned to the *de-novo* genome of the Sumatran rhinoceros (*Dicerorhinus sumatrensis*), which is the closest extant relative to woolly rhinoceros, with BWA v0.7.13 [55] and then processed with SAMtools v1.3 [56]. We mapped the merged sequencing reads against the reference genome using the BWA aln algorithm and slightly modified default settings with deactivated seeding (-l 16,500), allowing more substitutions (-n 0.01) and allowing up to two gaps (-o 2). We then used the BWA samse command to generate alignments and subsequently converted reads mapping to the reference genome from SAM to BAM format, sorted and indexed using SAMtools. We estimated the endogenous DNA content for each sample as the proportion of reads mapping to the reference genome. Duplicate reads were

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

removed in order to avoid upward bias in the estimation of endogenous DNA content and to avoid inflation of length distribution for loci with deep coverage using a custom python script [20]. Endogenous DNA content ranged from 7.4% to 72.0% with a mean of 49.1% (Data S1A).

Mitogenome data processing

Reads were also mapped against the woolly rhinoceros mitochondrial reference genome (GenBank ID: FJ905813) using the above settings to generate mitochondrial BAM files for downstream processing. We imported the mitochondrial BAM files generated using SAMtools v1.3 as above to Geneious® v7.0.336 [57] where sequences were aligned using MUSCLE v3.8.31 [81]. We then called consensus sequences for positions with at least 5X coverage using a majority consensus rule, with ambiguous and low-coverage positions called as undetermined (N). Finally, we visually inspected the assembled sequences to assess overall coverage across the 16,438 base pairs (bp) of the mitogenome and quality of the SNPs identified.

Nuclear genome data processing

We selected one sample with 56.7% endogenous DNA and dated to 18,699-18,356 cal BP for deep-sequencing (ND035, Data S1A). We used the stock library described above as a template for PCR amplification and performed six indexing PCR reactions in order to increase library complexity. PCR amplification, purification and size selection were performed as described above and following Meyer & Kircher [45].

We used the same approach and parameters as described above (in ''Endogenous DNA content estimation'') to generate BAM files, including mapping the woolly rhinoceros reads to the *de-novo* Sumatran Rhinoceros reference and duplicate read removal. Next, we used Picard v1.141 (http://broadinstitute.github.io/picard) to assign read group information including library, lane and sample identity to each bam file. Reads were then re-aligned around indels using GATK v3.4.0 [58]. Only reads/alignments with mapping quality $\geq$30 were kept for subsequent analysis. We estimated damage patterns and then performed a base-recalibration step on the BAM files using Mapdamage v2.0 [59]. Finally, we estimated the depth of genome coverage using Qualimap v2.2.1 [60]. The average genome coverage was 13.6X with 70% of the genome with a genome coverage $\geq$10X (Data S1A). We then called variants using bcftools mpileup v1.3 [56] using a minimum depth of coverage of 1/3X of the average coverage and a maximum of 2X the average coverage, base quality $\geq$30 and removed SNPs within 5bp of indels. We also identified CpG sites using a custom script masking CG sites and removed them using BEDtools v2.29.2 [61]. Finally, for all downstream analyses, we excluded chromosome-linked scaffolds and masked repeat regions using BEDtools. Overall, we obtained 28,180,718 SNPs.

Mitogenome data analysis

Basic statistics including nucleotide diversity (π), number of haplotypes (n), haplotype diversity (*d*), and number of segregating sites (S) were performed using DnaSP6 v6.12.03 [62]. Nucleotide diversity (π) within the samples was 0.00268; the number of segregating sites (S) was 119; and haplotype diversity (d) was 0.989. Second, we created a median joining network in PopArt [63]. We added a traits block to the nexus alignment using a custom python script to visualize the samples based on geographic region (Figure S1A). Third, we performed demographic reconstruction of woolly rhinoceros over the last 125 ky BP in BEAST v1.10.4 [8]. The evolutionary model for the 14 mitogenome dataset was determined to be HKY+I using the Bayesian Inference Criterion in jModelTest v2.1.9 [64]. Calibrated tip dates were added in BEAUti v1.10.4 [8] using the median value of dates listed in Data S1A. For one sample (ND045) that was dated at > 45 Cal ka BP (Data S1A), we used a prior with a wide boundary (uniform, initial value: 45300, lower: 0, upper: 500,000) in order to estimate its age. Its date was estimated at 36,445 (95% HPD: 14,839-49,073) ka BP. Three tree models were analyzed: constant size, Bayesian skyline and Bayesian Skyride. For the Skyline model, we decreased the number of groups to five in order to avoid over-parameterisation of the model. A strict molecular clock was applied and the clock rate was set to a normal distribution with the initial value as $6.1\text{x}10^{-9}$ substitutions/site/year, the mean value at $6.1\text{x}10^{-9}$ and a standard deviation of 0.01. The initial value was taken from Steiner et al. who calculated a substitution rate of $6.1\text{x}10^{-3}$ per site per million years for Sumatran Rhinoceros, the closest extant relative of the woolly rhinoceros [82]. All models were run using Beast v1.10.4 for 10 million generations with sampling every 1000 generations. We calculated marginal likelihoods for each tree model using path and stepping stone models implemented in Beast v1.10.4 (Data S1B). All output log files were visualized in Tracer v1.7.1 [66] in order to ensure convergence had occurred. Demographic reconstructions based on female effective population size for each tree model were also performed using Tracer (Figure S1B-d). Tree Annotator v1.10.4 [8] was used to remove 10% burn-in from the tree files. The phylogenies were then visualized in Figtree v1.4.4 [65].

Demographic reconstruction

We used the Pairwise Sequentially Markovian Coalescent (PSMC v0.6.5 [17]) model to infer the effective population sizes (N_e) of the woolly rhinoceros over time. This approach infers the distribution of the time to the most recent common ancestor (TMRCA) between the two alleles across all chromosomes using the density of heterozygous sites across the diploid genome of a single individual. Regions of low heterozygosity reflect recent coalescent events while regions of high heterozygosity reflect more ancient coalescent events. The rate of coalescent events in each segment is then informative about changes in effective population size through time since the rate of coalescence is inversely proportional to effective population size. We generated consensus sequences for all

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

CellPress
OPEN ACCESS

autosomes of the ancient genome using the SAMtools v1.3 mpileup command and the 'vcf2fq' command from vcfutils.pl. We used filters for base quality, mapping quality and root-mean-squared mapping quality below 30, and depth below 1/3 and higher than 2-times the average coverage estimated for each specimen. In order to infer the distribution of the time to the most recent common ancestor (TMRCA) between the two copies of each chromosome from each individual across all autosomes, we set N (the number of iterations) = 25, t (Tmax) = 15 and p (atomic time interval) = 64 (4+25*2+4+6, for each of which parameters are estimated with 28 free interval parameters). We used a rate of 2.34×10^{-8} substitutions/site/generation [12] and a generation time of 12 years [13]. Due to the uncertainty in substitution rates in Rhinocerotidae, we also tested rates of 1.2×10^{-8} [83] and 3.3×10^{-8} substitutions/site/generation based on horse [84].

Heterozygosity and inbreeding

We first estimated the overall autosomal heterozygosity using mlRho v2.7 [67] to estimate the population mutation rate (θ), which approximates expected heterozygosity under the infinite sites model. We filtered out bases with quality below 30, reads with mapping quality below 30 and positions with root-mean-squared mapping quality below 30. Because high or low coverage in some regions resulting from structural variation can create erroneous mapping to the reference genome and false heterozygous sites, we filtered out sites with depth lower than 1/3X and higher than 2X the estimated average coverage. The maximum likelihood approach implemented in mlRho has been shown to provide unbiased estimates of average within-individual heterozygosity at high coverage [67, 85].

Second, we estimated inbreeding by identifying runs of homozygosity (ROH) and calculating the individual inbreeding coefficients (F_{ROH}) with the sliding-window approach implemented in PLINK v1.9 [68]. We converted the filtered vcf file into a ped file and identified ROHs in all autosomal scaffolds. We used a sliding window size of 100 SNPs (*–homozyg-window-snp 100*). A window was then defined as homozygous if there were not more than 15 missing sites (*homozyg-window-missing 15*) and not more than 5 heterozygous sites per window (*homozyg-window-het 5*). If at least 5% of all windows that included a given SNP were defined as homozygous, the SNP was defined as being in a homozygous segment of a chromosome (*homozyg-window-threshold 0.05*). This threshold was chosen to ensure that the edges of a ROH are properly delimited. A homozygous segment was defined as a ROH if all of the following conditions were met: the segment included $\geq$25 SNPs (*homozyg-snp 25*) and covered $\geq$100 kb (*homozyg-kb* 100). Furthermore, the minimum SNP density was one SNP per 50 kb (*homozyg-density 50*) and the maximum distance between two neighboring SNPs was $\leq$1,000 kb (*homozyg-gap 1,000*). For the number of heterozygous sites within ROHs, we set the value at 750 (*homozyg-het 750*) in order to prevent sequencing errors to cut ROHs. We found that the majority of ROH were $<$0.5 Mb in length with a maximum ROH length of 2.5 Mb.

We also used a less stringent number of heterozygous sites per window of 10 (*homozyg-window-het 10*) to account for potential ancient DNA damage. However, results were very similar with the majority of ROH $<$0.5 Mb in length and maximum ROH of 3.4 Mb (Figure S2).

Non-synonymous mutations

We used SNPeff v4.3 [69] to annotate non-synonymous nucleotide substitutions in coding regions for the woolly rhinoceros. First, we generated a database for the Sumatran rhinoceros reference genome using the protein sequences extracted from our annotation. Second, we identified non-synonymous variants in two different SNPeff impact categories: (a) *Moderate*, non-disruptive variants that might change protein function and effectiveness, hereafter referred to as missense variants; and (b) *High*, variants assumed to have high (disruptive) impact in the protein, probably causing protein truncation or triggering nonsense mediated decay (e.g., stop codons, splice donor variant and splice acceptor), hereafter referred to as loss of function (LoF) variants [69].

Next, we ranked all missense variants reported by SNPEff using relative change of selected physicochemical properties of substituted amino acids and using a custom script (https://github.com/NBISweden/simpred) (Data S1C). For every non-synonymous amino acid substitution, we computed three values reflecting the substitution-induced change in amino acid properties:

1. *aaIndex score.* Seven aaIndex [86] descriptors of amino acid properties selected in [87, 88]. Every such descriptor represents a certain property of amino acid (e.g., its polarity as a numerical value corresponding to the magnitude of the property). These descriptors were chosen with biological interpretability in mind and span a 7-dimensional space (with every descriptor represented by one axis) that preserves full discernibility between any pair of amino acids. Now, we represented every amino acid as a point in this 7-dimensional space and following this step, we used multi-dimensional scaling (MDS) to find a low dimensional (2D) representation of the data points. In this low-dimensional representation, the simple Euclidean distance between any pair of points representing particular amino acids corresponds to the overall magnitude of change induced by substituting one amino acid from the pair to the other. This relation is naturally symmetric. We then constructed a relative substitution scores matrix where the score for a given pair of amino acids is relative to the maximal possible pairwise distance between a pair of amino acids.
2. *Sneath index* which takes into account various chemical properties of substituting amino acids [89]. Similarly to *aaIndex*, we use scores relative to maximal possible change.
3. *Non-exchangeability index*, which is an inverse of the exchangeability index proposed in Yampolsky and Stoltzfus [90]. The exchangeability index is based on a number of experimental studies and thus provides another perspective on potential amino

Please cite this article in press as: Lord et al., Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.07.046

acid substitution impact. It is a non-symmetric index, i.e non_exchgb(aa1 → aa2) ≠ non_exchgb(aa2 → aa1) and is not defined for all amino acid substitutions. We are also using relative non-exchangeability for consistency with the previously mentioned scores. For every type of score, value of 1.0 corresponds to the maximal possible change (impact) while low values indicate likely mild substitution effects.

Finally, we retained genes affected by missense and LoF variants and identified orthologs, and assessed the functional enrichment of these LoF variants using Panther with horse as the reference set [70] (Data S1D-H).

Current Biology, Volume *30*

Supplemental Information

Pre-extinction Demographic Stability and Genomic Signatures of Adaptation in the Woolly Rhinoceros

Edana Lord, Nicolas Dussex, Marcin Kierczak, David Díez-del-Molino, Oliver A. Ryder, David W.G. Stanton, M. Thomas P. Gilbert, Fátima Sánchez-Barreiro, Guojie Zhang, Mikkel-Holger S. Sinding, Eline D. Lorenzen, Eske Willerslev, Albert Protopopov, Fedor Shidlovskiy, Sergey Fedorov, Hervé Bocherens, Senthilvel K.S.S. Nathan, Benoit Goossens, Johannes van der Plicht, Yvonne L. Chan, Stefan Prost, Olga Potapova, Irina Kirillova, Adrian M. Lister, Peter D. Heintzman, Joshua D. Kapp, Beth Shapiro, Sergey Vartanyan, Anders Götherström, and Love Dalén

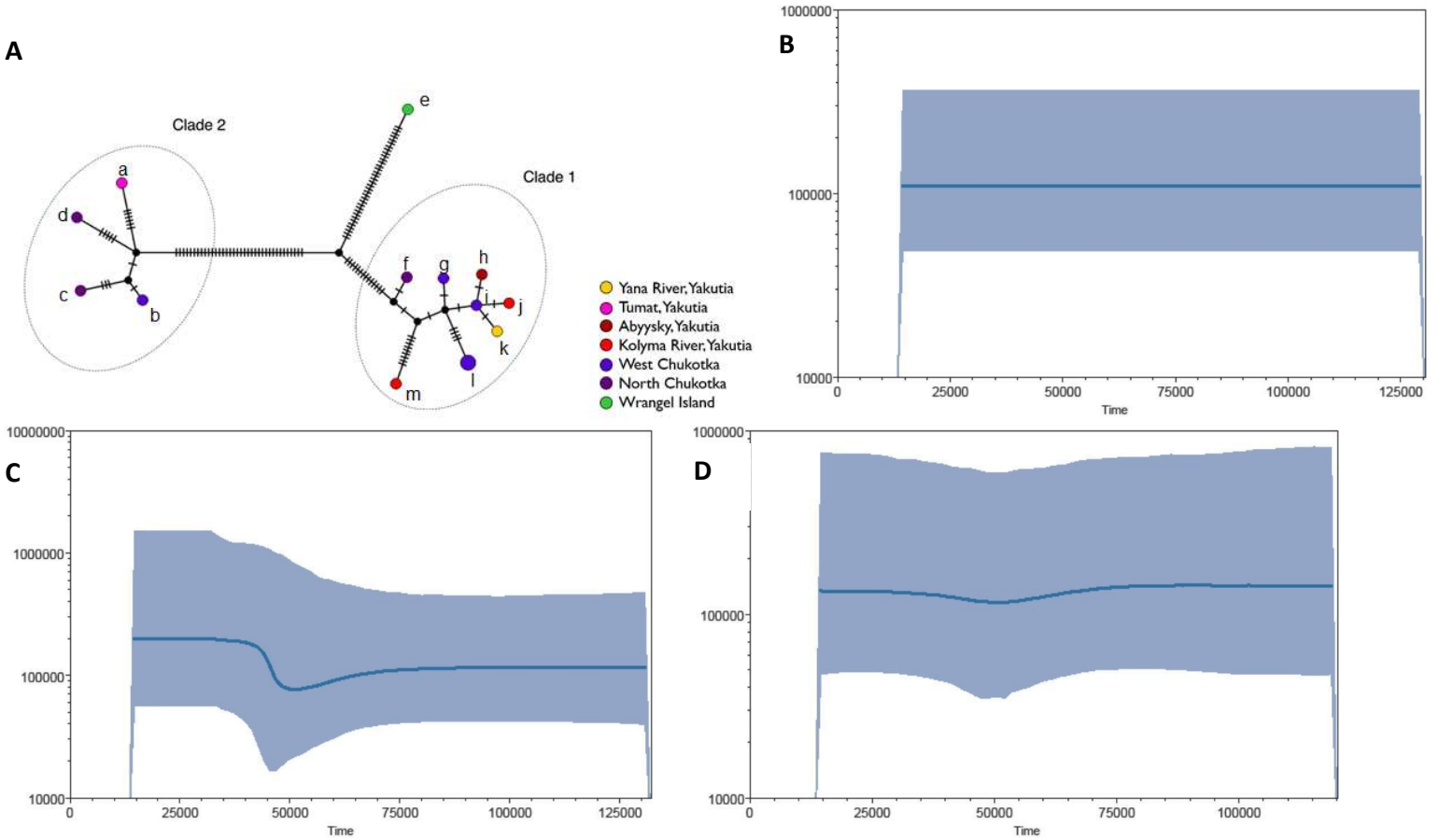

Figure S1. **Mitogenome network and demographic analyses. Related to Figure 1. A)** Mitogenome median joining network for the 14 woolly rhinoceros mitogenomes. Hatch marks represent one mutational step and black circles are missing haplotypes. The size of the circles is proportional to the number of samples with a given haplotype. Haplotypes are listed in Data S1A. **B-D)** Bayesian demographic analyses of the mitochondrial genomes in order of best-supported model as per marginal likelihood estimation (Data S1B). B. Constant Size model (best supported model), C. Bayesian skyline, D. Bayesian Skyride.

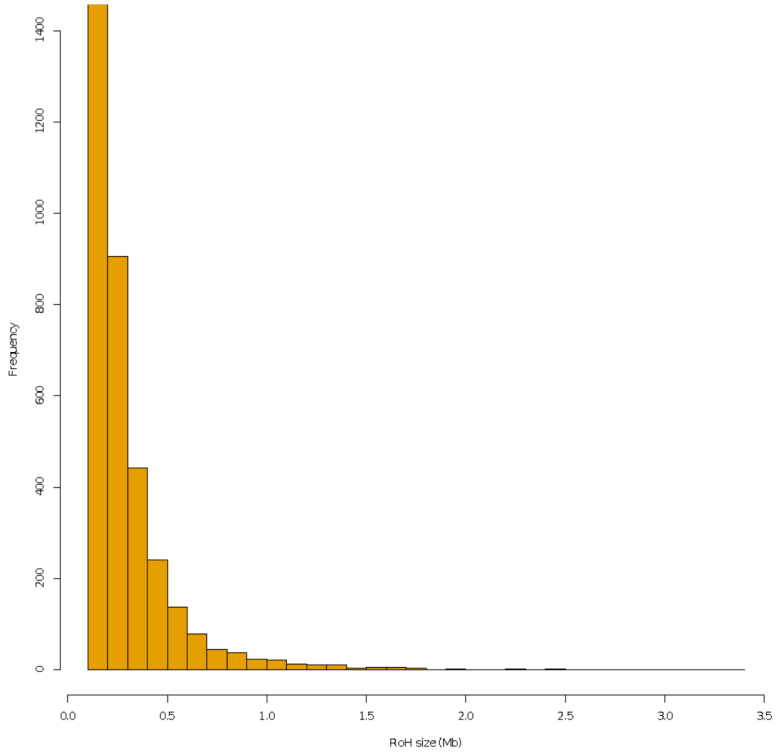

Figure S2. Frequency distribution of runs of homozygosity (ROHs) size in one woolly rhinoceros. Related to Figure 3. The specimen (ND035) was dated to 18,530 ±170 cal BP. Only ROH ≥0.1 Mb are shown. Results are shown here for comparison with main results and were obtained using less stringent parameters with *homozyg-window-het*=10 instead of *homozyg-window-het*=5.

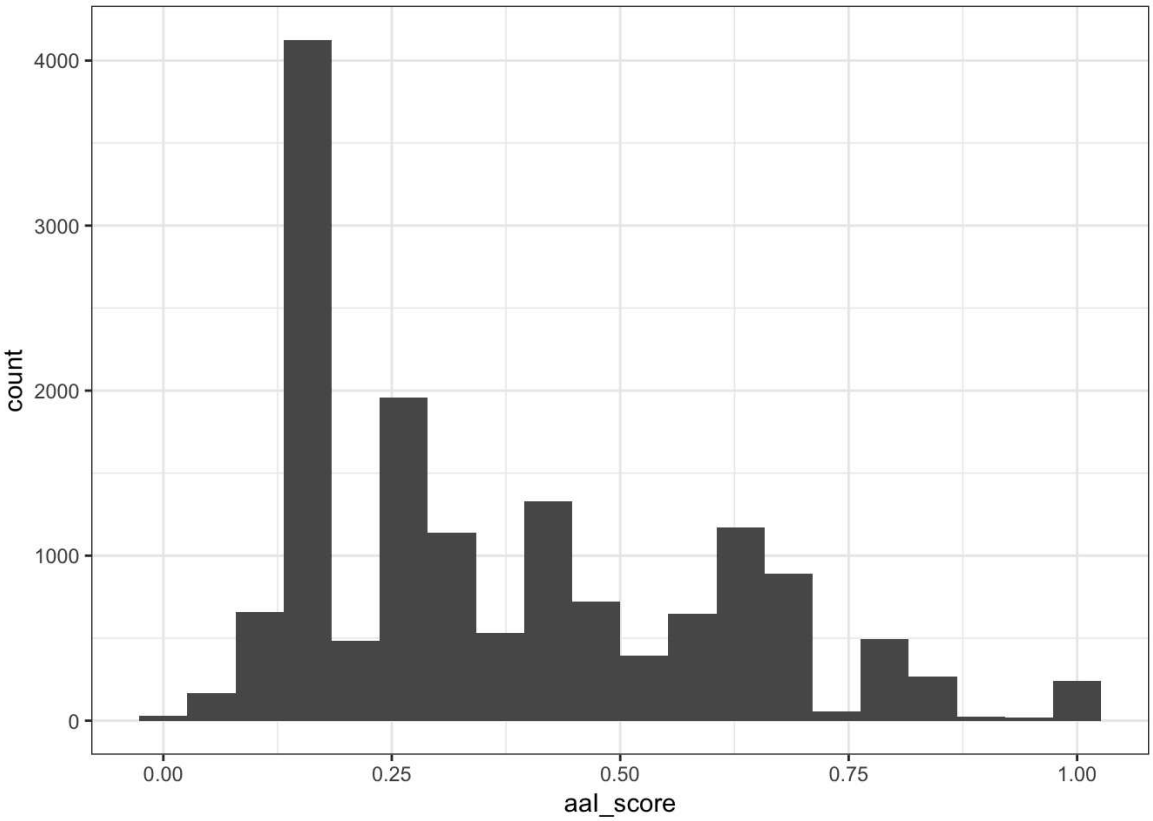

Figure S3. **Distribution of amino acid index (aaI) scores describing the change in physico-chemical properties of proteins. Related to Star Methods.** aaI scores for each variant are listed in Data S1C.