Recent mitochondrial lineage extinction in the critically endangered Javan rhinoceros

Article *in* Zoological Journal of the Linnean Society · March 2020

DOI: 10.1093/zoolinnean/zlaa004/5802322

10 authors, including:

Michael W Bruford

Cardiff University

845 PUBLICATIONS **18,356** CITATIONS

SEE PROFILE

Zoological Journal of the Linnean Society, 2020, **XX**, 1–12. With 4 figures.

Recent mitochondrial lineage extinction in the critically endangered Javan rhinoceros

ASHOT MARGARYAN[1,2*], MIKKEL-HOLGER S. SINDING[1,3], SHANLIN LIU[1,4], FILIPE GARRETT VIEIRA[1], YVONNE L. CHAN[5], SENTHILVEL K. S. S. NATHAN[6], YOSHAN MOODLEY[7], MICHAEL W. BRUFORD[8] and M. THOMAS P. GILBERT[1,9]

[1]*Section for Evolutionary Genomics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark*
[2]*Institute of Molecular Biology, National Academy of Sciences, 7 Hasratian Street, Yerevan 0014, Armenia*
[3]*Greenland Institute of Natural Resources, Kivioq 2, PO Box 570, 3900 Nuuk, Greenland*
[4]*BGI-Shenzhen, Shenzhen, 518083, China*
[5]*Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Box 50007, SE-104 05, Sweden*
[6]*Sabah Wildlife Department, 5th Floor, Block B, Muis Tower, 88100 Kota Kinabalu, Sabah, Malaysia*
[7]*Department of Zoology, University of Venda, Private Bag X5050, Thohoyandou 0950, Republic of South Africa*
[8]*School of Biosciences and Sustainable Places Institute, Cardiff University, Cathays Park, Cardiff CF10 3AX, UK*
[9]*Norwegian University of Science and Technology, University Museum, 7491 Trondheim, Norway*

Received 1 October 2018; revised 14 January 2020; accepted for publication 14 January 2020

The Javan rhinoceros (*Rhinoceros sondaicus*) is one of five extant rhinoceros species and among the rarest large mammals on Earth. Once widespread across Southeast Asia, it is now on the verge of extinction, with only one wild population remaining (estimated at ~60 individuals) on the island of Java, Indonesia. To assess the past genetic diversity of the female lineage of *R. sondaicus*, we generated mitochondrial genome data from eight museum specimens dating back to the 19th century, before the range of the Javan rhinoceros was dramatically reduced, for comparison against mitochondrial DNA (mtDNA) sequences of current *R. sondaicus* and other rhinoceros species. We succeeded in reconstructing five full and three partial ancient mitogenomes from the eight samples. We used BEAST to assess the phylogenetic relationship of the five extant rhinoceros species and the historical samples. The results show that the oldest and most diverse mtDNA lineages of *R. sondaicus* are found in historical samples, indicating a significant reduction of mtDNA diversity in modern Javan rhinos. We anticipate that the newly sequenced data will represent a useful resource for improving our understanding of evolutionary history of this species, should future studies be able to increase the available dataset. We hope this information may help in conservation efforts for this species.

ADDITIONAL KEYWORDS: ancient DNA – diversity – DNA – endangered – Javan rhino – mitochondrial – *Rhinoceros sondaicus* – rhinos.

INTRODUCTION

The Javan rhinoceros (*Rhinoceros sondaicus* Desmarest, 1822) is the rarest of the five extant rhinoceros (rhino) species, today consisting of a single population in Ujung Kulon National Park (UKNP) on the western tip of the Indonesian island of Java (Fig. 1). This critically endangered species is estimated to have only ~60 animals remaining in the wild (Haryono *et al.*, 2016), and it is thus one of the rarest mammals in the world. Before the end of the 19th century, its range included much of Southeast Asia, from India

*Corresponding author. E-mail: ashotmarg2004@gmail.com

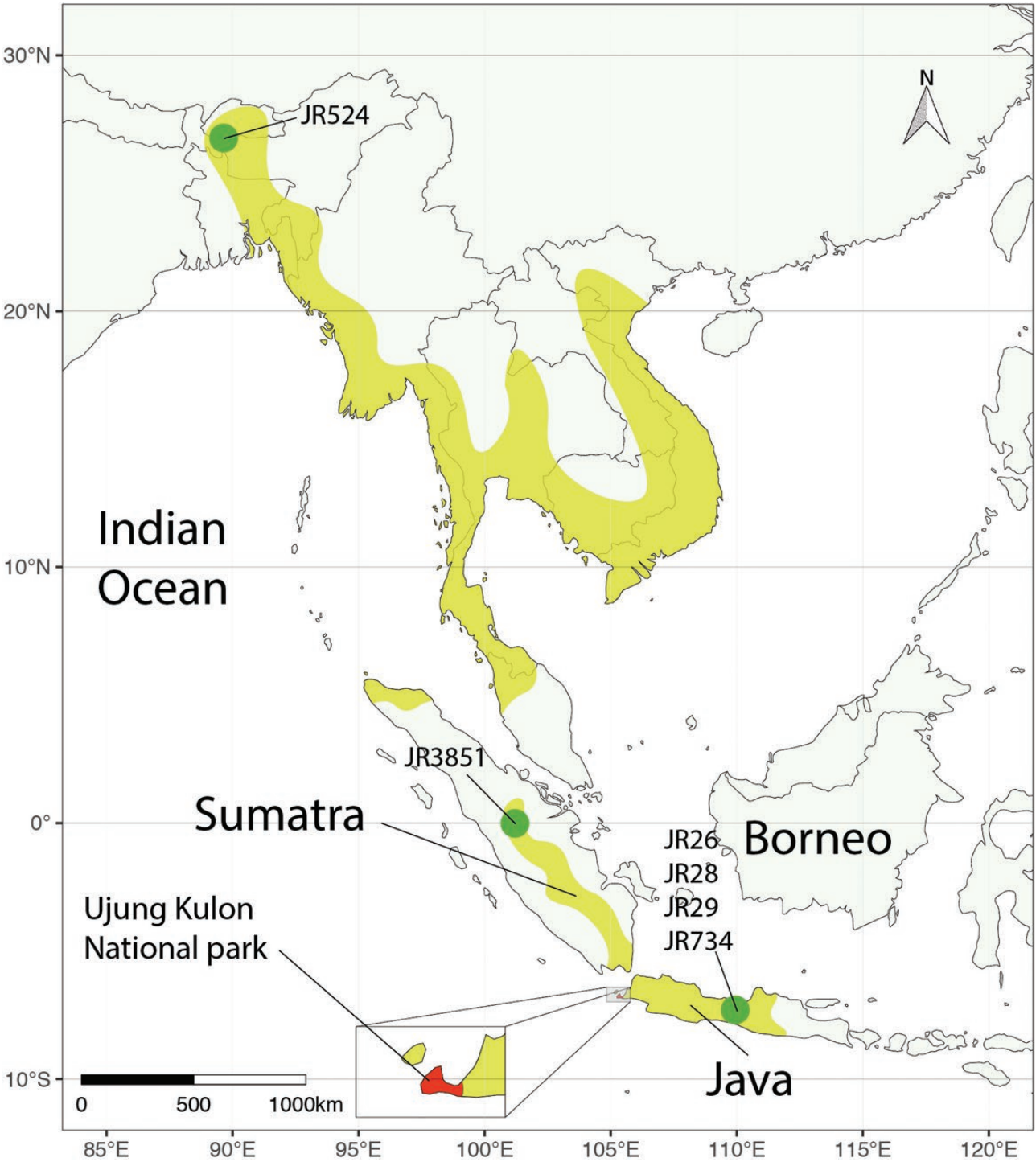

Figure 1. Map of Javan rhino distribution in Southeast Asia, showing their approximate historical (yellow) and current (red) distribution (Fernando *et al.*, 2006; Groves & Leslie, 2011). The green dots indicate the approximate locations of historical samples used in this study. Two samples with unknown locations are not mentioned on the plot (Table 1).

and China to the islands of Java and Sumatra (Groves & Leslie, 2011), putatively comprising three subspecies: *Rhinoceros sondaicus inermis* (Lesson, 1838) in the northern part of its range (Rookmaaker, 1997) where it is now extinct, *Rhinoceros sondaicus annamiticus* (Heude, 1892) mainly in Vietnam (Brook

et al., 2014), extinct since 2010, and the nominal subspecies *Rhinoceros sondaicus sondaicus* on Java and Sumatra (Groves & Leslie, 2011). However, the distribution and numbers of *R. sondaicus* since the end of the 19th century have decreased dramatically as a result of poaching, loss of habitat and diseases transmitted by cattle (Khairani *et al.*, 2018). With the last continental Asian Javan rhino killed in Vietnam in 2010, subspecies *R. sondaicus annamiticus* was declared extinct (Brook *et al.*, 2014), leaving only the Javan population (*R. sondaicus sondaicus*) in the wild. With the entire population of the world today restricted to the UKNP, the urgency of its conservation cannot be overstated.

Both fossil and genetic analyses indicate that the Indian rhinoceros (*Rhinoceros unicornis* Linnaeus, 1758) is the closest extant relative of the Javan rhinoceros, although there is disagreement on when they diverged. Fossil evidence suggests a divergence in Asia ~3 Mya (Carroll, 1988), whereas molecular estimates suggest a much earlier split (~12–13 Mya; Tougard *et al.*, 2001; Willerslev *et al.*, 2009). Similar phylogenetic relationships between the Indian and Javan rhinos were observed by Welker *et al.* (2017) based on protein sequences.

Previous genetic studies of *R. sondaicus* have been restricted to either a whole mitochondrial genome (mitogenome) sequence of a single historical specimen (Willerslev *et al.*, 2009; Mohd Salleh *et al.*, 2017; Moodley *et al.*, 2018), thus representing only a single point in its geographical range, or at the population scale restricted to short mitochondrial fragments, such as 12S rRNA gene and D-loop (Fernando *et al.*, 2006) and 12S rRNA and cytochrome *b* genes (Tougard *et al.*, 2001). Thus, there is a need to improve the dataset of complete mitochondrial DNA (mtDNA) genome sequences to be more representative of the geographical range that the species once covered. This can be resolved only by analysis of historical samples, such as those held in natural history collections. We therefore aim to contribute to this, by using Illumina sequencing technology to generate mtDNA genome-scale data from eight historical (100- to 200-year-old) specimens that span its historical geographical range. These data were also combined with new and existing rhino mtDNA sequences, in order to: (1) estimate the divergence times among Javan rhinos; and (2) reveal the loss of mtDNA diversity in Javan rhinos over the last century. Ultimately, although restricted to a relatively small sample size, we hope that these mtDNA sequences from this rare and understudied species will be valuable both for providing insights into its former diversity and for future studies, in which new mtDNA-based assays might be designed for its monitoring.

MATERIAL AND METHODS

Samples and DNA extraction

Eight historical *R. sondaicus* specimens, kept in the collections of the Natural History Museum at the University of Oslo or the Natural History Museum of Denmark, were sampled for DNA extraction. Most of the samples were collected or registered in the 19th century and originated from Java, Sumatra and Bhutan (only one sample) according to the museum records (for details, see Table 1). All laboratory work was performed in the ancient DNA (aDNA) laboratories at the Centre for GeoGenetics, Natural History Museum of Denmark, following standard clean laboratory procedures (Orlando *et al.*, 2011; Cappellini *et al.*, 2012). Samples of bone were digested in a urea–proteinase K buffer as described by Ersmark *et al.* (2015), and nail, cartilage or dried soft tissue was digested in an EDTA–proteinase K buffer as described by Gilbert *et al.* (2007). Digests were purified following the protocol of Dabney *et al.* (2013), in combination with modified aDNA binding buffer described by Allentoft *et al.* (2015). Next-generation sequencing libraries were prepared following the single-tube protocol (Carøe *et al.*, 2018) with the modifications described by Mak *et al.* (2017), using Illumina-specific adapters (Meyer & Kircher, 2010) for single-read sequencing on the Illumina HiSeq 2500 platform.

We complemented this dataset with the rhinoceros mitogenome sequences available in GenBank (Supporting Information, Table S1) and through reconstruction of an additional four full mitochondrial genomes from currently unpublished data from the rhinoceros genome sequencing consortium (Dalen L, Gilbert MTP, unpublished data). This unpublished dataset includes three modern rhinos [one each for the black rhino, *Diceros bicornis* (Linnaeus, 1758), the Indian rhino, *Rhinoceros unicornis*, and the Sumatran rhino, *Dicerorhinus sumatrensis* (Fischer, 1814)] and the extinct woolly rhino, *Coelodonta antiquitatis* (Bronn, 1831).

Bioinformatics

We used the BAM Pipeline in PALEOMIX to trim and map the sequencing reads from the eight historical Javan rhinoceros specimens (Schubert *et al.*, 2014). In brief, Illumina adaptor sequences and stretches of Ns at both ends of the reads were trimmed from all of the aDNA sequences using AdapterRemoval v.2.2 (Schubert *et al.*, 2016), keeping only sequences with a minimum length of 30 bp. The trimmed sequences were subsequently mapped against the published *R. sondaicus* mitochondrial genome (GenBank ID: FJ905815; Willerslev *et al.*, 2009) using BWA

Table 1. List of historical samples of *Rhinoceros sondaicus* sequenced for this study

Sample	Collection ID	Location	Latitude	Longitude	Specimen	Sample	Collection date	Registration date	GenBank ID
JR26	NHM, DK CN26	Java	7.49°S	110°E	Skeleton	Nail, cartilage	1800s	N/A	MK909146
JR27	NHM, DK CN27	Calcutta?	N/A	N/A	Skull	Bone	N/A	28/05/1839	MK909142
JR28	NHM, DK CN28	Java	7.49°S	110°E	Skull	Bone	N/A	15/12/1846	MK909144
JR29	NHM, DK CN29	Java	7.49°S	110°E	Skull	Bone	N/A	21/06/1854	MK909147
JR524	NHM, DK CN524	Bhutan	26.72°N	88.99°E	Skull	Bone	24 February 1881	10/06/1887	MK909150
JR734	NHM, UiO 734	Java	7.49°S	110°E	Skull	Bone, tissue	30 November 1838	10/09/1879	MK909149
JR3320	NHM, DK CN3320	N/A	N/A	N/A	Skull+skin	Skin, tissue	N/A	15/01/1961	MK909148
JR3851	NHM, UiO 3851	Sumatra	0°N	102°E	Skull	Bone	30 June 1896	N/A	MK909151

Abbreviations: DK, Denmark; N/A, not available; NHM, Natural History Museum; UiO, University of Oslo.

v.0.6.2 with the 'aln' algorithm (Li & Durbin, 2009), with the seed disabled to higher sensitivity (Schubert *et al.*, 2012). The aligned sequences were filtered for mapping quality 30 and sorted using Picard (http://broadinstitute.github.io/picard/) and SAMtools (Li *et al.*, 2009). Duplicate sequences at the library level were removed by Picard MarkDuplicates (http://broadinstitute.github.io/picard/).

The mtDNA consensus sequences for each historical rhinoceros sample were constructed using GENEIOUS v.9 (http://www.geneious.com/), based on the final bam files, and considering only bases with a sequencing depth of ≥ 2X. At every position, a nucleotide was called only if it was observed in ≥ 75% of the reads covering that site. Typical aDNA damage parameters, such as the C→T transition rates (Pääbo, 1989), were estimated using mapDamage2.0 with default settings (Jónsson *et al.*, 2013).

The mtDNA sequences of four additional species (black, Indian, Sumatran and woolly rhinoceroses) were constructed by mapping the adapter-free DNA reads to the relevant reference genomes (Table 2) with 'bwa mem', using sequences with mapping quality ≥ 30, and calling the consensus mtDNA sequences using GENEIOUS v.9. Raw sequencing reads for these four samples were taken from unpublished data. The fasta files of three modern (black, Indian and Sumatran), woolly and eight historical Javan rhino mitochondrial genomes (five full and three partial) have been deposited in GenBank under accession numbers MK909142–MK909153.

Phylogenetic analysis

The software MAFFT v.7 (https://mafft.cbrc.jp/alignment/server/) was used with E-INS-i algorithm to obtain the mtDNA sequence alignment, which was checked manually for possible misalignments around the indels. The maximum likelihood (ML) analysis was conducted with RAxML (Stamatakis, 2014) to assess the phylogenetic relationships of the sequenced species with a GTR+GAMMA model of nucleotide substitution. We performed 1000 bootstrap replicates to obtain node support. A phylogenetic network analysis based on a 413 bp region (tRNA-Pro gene and partial D-loop) of published modern (Fernando *et al.*, 2006) and historical Javan rhinos, in addition to the Indian rhinos, was conducted with POPART (http://popart.otago.ac.nz) using the 'Integer Neighbor-Joining' algorithm. The dataset included five high-coverage historical Javan rhino sequences from the present study (JR26, JR27, JR734, JR3320 and JR3851), three published historical sequences from Java (FJ905815, KY117574 and AY739628), two modern sequences (AY739626 and AY739627) representing 56 Javan rhino samples from the Ujung Kulon National

Park, and recently extinct *R. sondaicus annamiticus* from Vietnam (AY739625) in addition to three Indian rhino sequences as a comparative dataset (Supporting Information, Table S2).

Interspecies analysis

We used this newly assembled rhinoceros mitogenome dataset to reconstruct the phylogenetic relationships between all rhinos, both in order to confirm the species origin for each of the historical samples and as a basis for subsequent molecular dating analyses in order to enhance our insight into the divergence times between the Javan rhinoceros samples. We used BEAST v.1.8.4 (Drummond *et al.*, 2012) for creating Bayesian phylogenies, using tip calibration for ancient samples. For this analysis, a coalescent skyline tree prior was implemented with an uncorrelated, relaxed, lognormal clock model, which was shown to be a reliable prior for interspecies-level phylogenetic trees (Ritchie *et al.*, 2017). We used complete mtDNA sequences of all available rhino samples and *Equus asinus* Linnaeus, 1758 and *Equus caballus* Linnaeus, 1758 sequences in order to calibrate the age of the Bayesian tree based on fossil records of the age of the split between the rhinos and equids (Supporting Information, Table S1). The GTR+I+G nucleotide substitution model was used, based on results of jModelTest v.2.1.10 (Darriba *et al.*, 2012), with the Akaike information criterion. Mitochondrial DNA sequences from three historical Javan rhinos sequenced in the present study were not included in this analysis owing to their low mtDNA coverage. The remaining five high-coverage Javan rhinos, in addition to the four additional newly sequenced rhinoceros whole mitogenomes, were used alongside 29 published rhinoceros and *Equus* mitogenomes for the Bayesian phylogenetic analysis (Supporting Information, Table S1). This included two previously published mitogenomes from historical Javan rhinos that originated in Java: FJ905815, a ~100-year-old sample from Oxford University Museum of Natural History, UK (Willerslev *et al.*, 2009), and KY117574, a sample of unknown collection date from the Natural History Museum of Denmark (Mohd Salleh *et al.*, 2017). We used the following dates as priors with normal distribution, based on fossil evidence for the origin of the following groups: modern equids (4 ± 0.5 Mya; MacFadden, 2005), also based on molecular data (Orlando *et al.*, 2013); Javan and Indian rhinos (3 ± 0.5 Mya; Carroll, 1988; Lacombat, 2005); and rhinos and equids (55 ± 3 Mya; Prothero & Schoch, 1989).

We ran the Markov chain Monte Carlo simulations for 50×10^8 states, sampling every 50×10^4 states, and designating the first 10% of the states as burn-in. The BEAST analyses were performed using the CIPRES open-access server for phylogenetic studies (Miller *et al.*, 2010). We checked the output data for convergence and sufficient effective sample size (ESS) estimates using TRACER v.1.7 (Rambaut *et al.*, 2018). TreeAnnotator v.1.8.4 from the BEAST package and FigTree (http://tree.bio.ed.ac.uk/software/figtree/) were used to visualize the results of the BEAST analysis.

Intraspecies analysis

Next, we used BEAST v.1.8.4 to estimate the divergence times within the Javan rhino mtDNA lineage, using the Indian rhino (*R. unicornis*; GenBank ID: X97336) as an outgroup. In contrast to the interspecies analysis, in this case a strict clock model was used when reconstructing the Bayesian phylogeny for the genus *Rhinoceros* (Indian and Javan rhinos) owing to the much shorter evolutionary time scale. Here, we used the divergence time of ~3.2 Mya between the Indian and Javan rhinos obtained from the interspecies analysis. To estimate the divergence times of the three low-coverage samples (JR28, JR29 and JR524) that were not included in the interspecies analysis, we constructed three different Bayesian phylogenetic trees. Each tree included one of the low-coverage samples alongside the two previously published and five newly sequenced complete mitogenomes. All ambiguous sites (Ns) were removed from the alignments of low-coverage genomes, restricting the analysis to ~40–50% of the complete mtDNA length (7606 bp in the case of the analysis presented in Fig. 3).

BEAST v.1.8.4 was also used to track changes in effective female population size (N_e) of Javan rhinos through time by applying the Bayesian skyline plot method (Heled & Drummond, 2008), with the average generation time of 16.5 years (Haryono *et al.*, 2016), using the seven available complete mtDNA sequences of Javan rhinos.

RESULTS

In total, 362 066 719 sequence reads were produced from the eight historical Javan rhinoceroses. After removal of adapters and trimming for stretches of Ns and low-quality bases, 315 609 438 sequences remained, with an average read length ranging from 48.4 to 63.2 bp.

The number of reads mapping to the mtDNA genome of *R. sondaicus (*GeneBank ID: FJ905815) varied greatly among the samples and ranged from as few as 494 (JR524) to 558 067 reads (JR734), which probably reflected different preservation states of these historical samples. The summary statistics of

Table 2. Summary statistics of mapping results for newly assembled mitochondrial DNA sequences of comparative species

Rhinoceros (species)	Mapped	Unique	Duplicates	DoC	Reference	Collection; ID	GenBank ID
Black (*Diceros bicornis*)	391 712	34 729	0.91	307	FJ905814	Zululand Rhino Reserve of South Africa; 46373	MK909143
Indian (*Rhinoceros unicornis*)	4 466 326	39 753	0.99	317	X97336	San Diego Zoo, USA; KB14498	MK909145
Sumatran (*Dicerorhinus sumatrensis*)	40 638	22 751	0.44	172	FJ905816	Sabah Wildlife Department, Malaysia; Tam	MK909153
Woolly (*Coelodonta antiquitatis*)	326 384	32 398	0.9	166	FJ905813	Muus-Khaya site, Yana River, Russia; MK019 (YC1)	MK909152

For the three modern rhino samples in this list, we obtained the required CITES permits and CITES exemption form (in case of the Indian rhino).
Abbreviations: DoC, estimated depth of coverage (*X*) of the mitochondrial DNA; Duplicates, fraction of PCR duplicates; Mapped, number of reads mapped to the relevant reference mitochondrial genome; Reference, GenBank IDs of the mitochondrial DNA sequences used as reference for mapping; Unique, number of uniquely mapped reads.

Table 3. Summary statistics of mapping results

Sample	Total	Retained	Mapped	Unique	DoC	L_unique	Damage	Coverage
JR26	65 915 186	62 101 360	67 563	30 291	113.2	61.35	4.16	0.999
JR27	51 244 821	44 799 046	49 435	25 649	104.91	67.15	2.82	0.999
JR28	21 911 221	20 838 531	520	488	1.96	66.04	4.65	0.572
JR29	53 065 860	48 306 145	563	514	1.59	50.72	1.44	0.411
JR524	21 544 225	20 269 794	494	468	1.64	57.61	3.31	0.463
JR734	69 464 120	64 821 307	558 067	75 917	333.77	72.18	3.37	0.998
JR3320	35 764 897	28 569 566	169 555	47 105	186.94	65.15	3.78	0.999
JR3851	43 156 389	25 903 689	7934	6901	24.76	58.91	5.68	0.993

The samples were mapped to the *Rhinoceros sondaicus* mitochondrial genome (GenBank ID: FJ905815).
Abbreviations: Coverage, fraction of mitochondrial DNA with called genotypes; Damage, C→T transition rate (as a percentage) at the first nucleotide position of the 5′ end of the reads; DoC, estimated depth of coverage (*X*) of the mitochondrial DNA; L_unique, average length (in base pairs) of the uniquely mapped reads; Mapped, number of mapped reads to the *R. sondaicus* reference mitochondrial genome; Retained, number of reads retained after removing adapters and stretches of Ns; Total, total number of sequenced reads; Unique, number of reads uniquely mapped to the *R. sondaicus* reference mitochondrial genome.

the mapping results are shown in Table 3. For five out of eight historical samples, the depth of coverage (DoC) of mtDNA was > 24*X*, which allowed us to reconstruct the whole mtDNA sequences of these samples reliably, with close to 100% mtDNA coverage. For the remaining three samples with low DoC (1.64*X*–1.96*X*), only partial mtDNA sequences were generated, spanning ~50% of the genome (Table 3). The shotgun sequencing reads from all historical samples showed increased C→T deamination rates at the 5′ ends of the sequences when compared with the *R. sondaicus* reference mitochondrial sequence. Together, these elevated C→T damage profiles and the short average length of mapped reads are consistent with the notion that the DNA molecules were of ancient origin (Table 3).

The phylogenetic relationship between the Javan and other rhinoceros species (*N* = 36) based on whole mtDNA genomes is presented in Figure 2. The results showed that *Elasmotherium* was a sister group to all modern rhinos (the Rhinocerotinae group), with 89% posterior probability with the split time of ~31 Mya. The most recent common ancestor of the group including all three extant rhinoceros species was estimated to have lived ~18 Mya. According to the Bayesian phylogeny, the Javan rhino and the Indian rhino were closest to the African species, although the node support was low, with ~70% posterior probability. We also noted that the maximum likelihood tree using RaxML (Supporting Information, Fig. S1) based on all rhinoceros whole mtDNA sequences (*N* = 36) showed identical topology to Figure 2, with a lack of high resolution in the Rhinocerotinae group.

Our expanded dataset of Javan sequences derived principally from Indonesian samples did enable us to show, for the first time, that the most recent common ancestor of this group lived ≥ ~400 000 years ago (95%

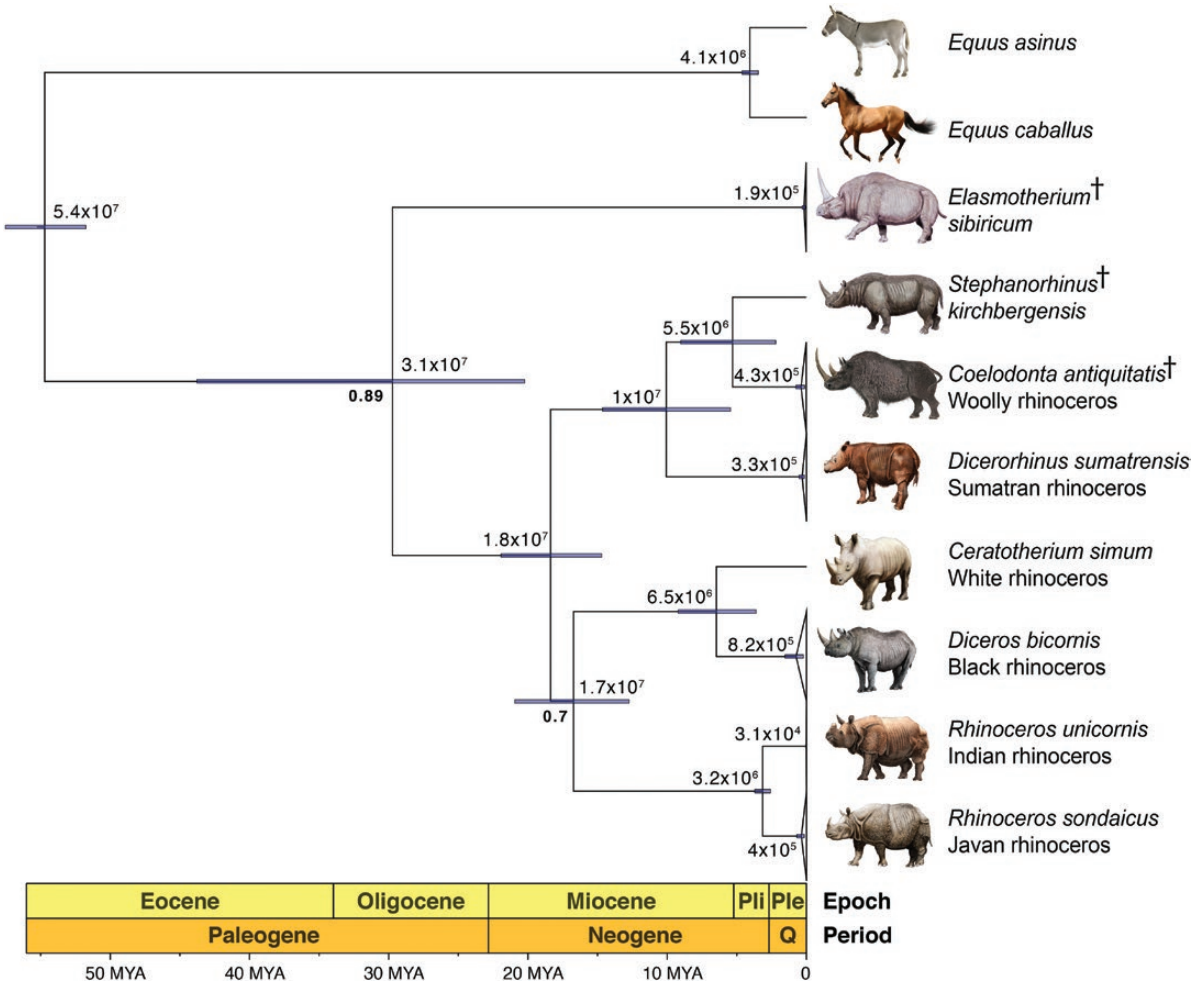

Figure 2. Bayesian phylogeny of all available (N = 36) rhino whole mitochondrial DNA sequences (published + new sequences from this study). Two *Equus* mitochondrial DNA sequences were included as the outgroup. The blue bars represent the 95% highest posterior density (HPD) intervals of the divergence times. All branches within a species level have been collapsed to improve readability. Node labels in bold show the Bayesian posterior probability (PP) values for two nodes; the rest of the nodes had PP values of one. Geological time scale abbreviations: Q, Quaternary; Ple, Pleistocene; Pli, Pliocene. Extinct rhino lineages are shown with daggers after species names.

HPD, 165 000–697 000 years ago; Fig. 2), with the oldest branches represented by two historical samples, JR3851 and JR734. A Bayesian skyline plot based on the seven Javan rhinos with complete mtDNA sequences showed a largely constant female population size of the Javan rhinos over the past 300 000 years until ~150 years ago (when most of the samples were collected), when the effective female population size was likely to have been ~9000 individuals (95% HPD, 1000–23 000). The average substitution rate of the whole mtDNA for the *R. sondaicus* lineages was estimated to be ~8×10^{-9} per site per year (95% HPD, 6.5×10^{-9} to 9.7×10^{-9}).

Our subsequent analyses, which included the lower coverage samples, provided further insights. First, although two of the low-coverage samples (JR28 and JR29) were closely related to most other samples (data not shown), sample JR524 fell basal (Fig. 3) to all other specimens, diverging in the Pleistocene at ~540 000 years BP (95% HPD, 450 000–640 000 years BP). Identical tree topologies with similar time splits were observed when using more stringent criteria (using at least three reads for calling a base) for constructing consensus mtDNA sequences of the three low-coverage samples, although this reduced the number of informative sites further. This indicated that: (1) the method for constructing the consensus mtDNA sequences was relatively robust; and (2) relatively few informative sites were enough to assess the phylogenetic relationship within the *R. sondaicus* lineage.

The Neighbor-Joining network analysis of modern and historical Javan rhino mtDNA sequences (N = 11)

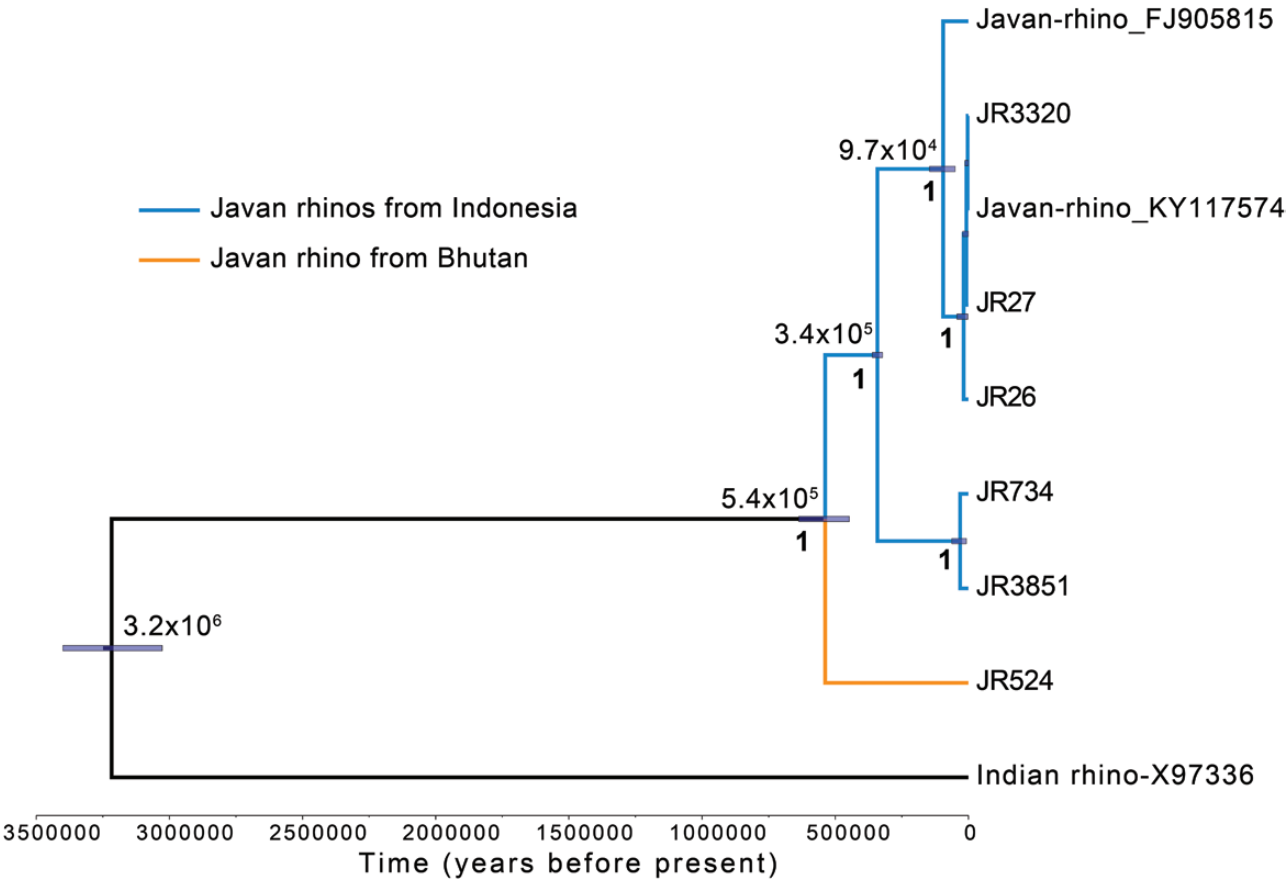

Figure 3. Bayesian phylogeny of the historical Javan rhinoceros mitochondrial DNA lineages, using the Indian rhino (*Rhinoceros unicornis*; GenBank ID: X97336) as the outgroup and including the partial (7606-bp-long) mitogenome of the Javan rhino sample collected in Bhutan (JR524). The blue bars represent the 95% highest posterior density (HPD) intervals of the divergence times. Node labels in bold show the Bayesian posterior probability (PP) values for the major clades. All specimens in the blue clade originated from Indonesia (apart from JR27, which has largely unknown 'Calcutta?' label) and are clearly divergent from the Bhutanese sample.

is presented in Figure 4. Given that no whole mtDNA genomes have been published for modern Javan rhinos, here we restricted our analysis to the few tRNA-Pro gene and partial D-loop sequences available for modern samples. Despite the small number of modern sequences, the analysis showed that most of the genetic diversity within the species was represented by some of the newly analysed historical sequences (e.g. JR734, JR3851 and JR26), in addition to the now extinct (since 2010) Vietnamese subspecies of the Javan rhino, *R. sondaicus annamiticus* (Fernando *et al.*, 2006).

DISCUSSION

We reconstructed five complete and three partial mtDNA sequences of critically endangered Javan rhinos by sequencing eight historical museum specimens, which more than doubled the number of available complete mtDNA lineages from this species.

The phylogenetic placement of the newly sequenced Javan rhino samples confirmed the species identity, and the overall tree topology based on all available mtDNA sequences was in accordance with recent studies (Kirillova *et al.*, 2017; Kosintsev *et al.*, 2019). The phylogenetic relationship in the Rhinocerotinae group (all rhinos excluding *Elasmotherium*) was similar to those published by Willerslev *et al.* (2009) and Orlando *et al.* (2003), namely showing a lack of resolution among the three main lineages of rhinoceros species (Indian + Javan, white + black and Sumatran + woolly + *Stephanorhinus*). However, the highest Bayesian probability tree reflected a greater genetic similarity between the two African and two Asian *Rhinoceros* species (albeit with low-probability node support of 0.7), with the clade of the Sumatran rhino as a sister group (Fig. 2). These results differed

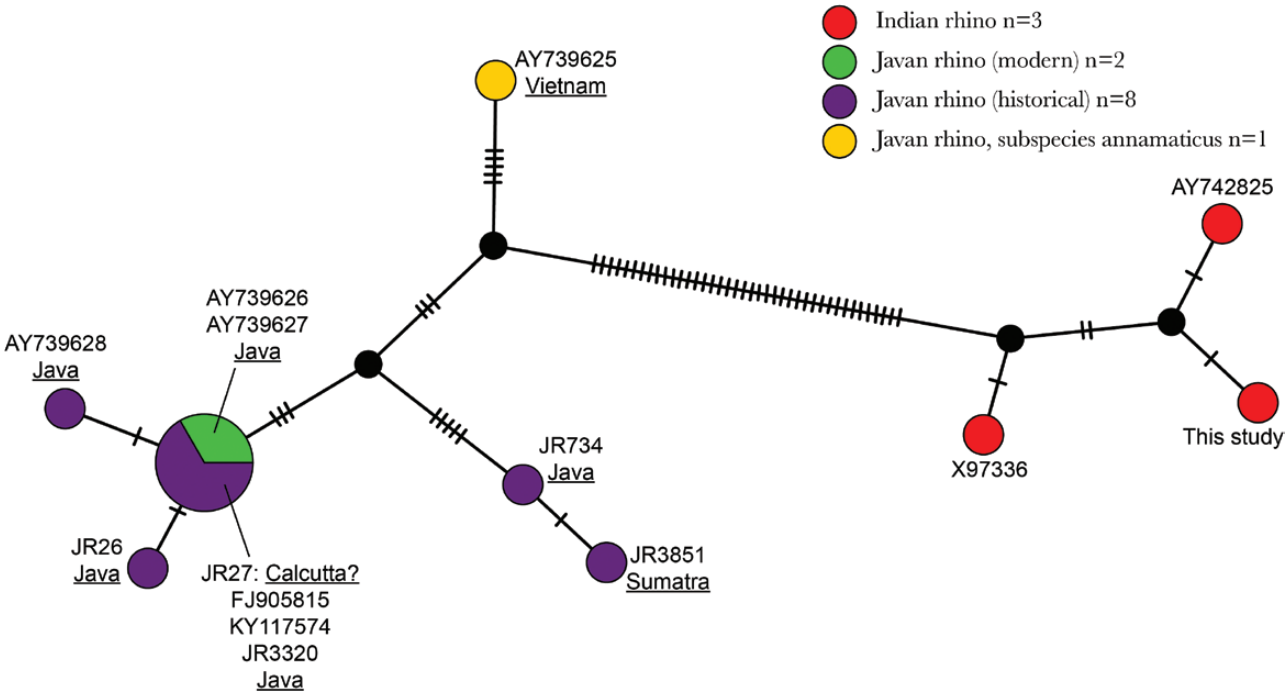

Figure 4. Neighbor-Joining network. The analysis of 14 modern and historical Javan and Indian rhinoceros mitochondrial DNA sequences was conducted using the 'Integer Neighbor-Joining' algorithm implemented in POPART. We used the partial D-loop region of 413 bp to maximize the number of available Javan rhino samples. 'Javan_annamiticus' represents the recently extinct (since 2010) Vietnamese subspecies *Rhinoceros sondaicus annamiticus*. Each circle represents a certain haplotype; smaller black circles indicate median vectors. Small black lines connecting branches between the haplotypes denote the number of mutation steps separating the haplotypes. Sample identities and regions of origin for the Javan rhinoceros samples are indicated next to each circle. Numbers in the legend indicate the number of sequences in each group.

from those based on phylogenetic analyses of protein (collagen alpha) sequences, which showed higher resolution within Rhinocerotinae and a different tree topology, with the African rhinos forming a sister group to the Asian species (Welker *et al.*, 2017). However, we caution that this might reflect the different genetic histories and/or power of resolution between nuclear (coding collagen alpha in this case) and mtDNA genomes (Steiner & Ryder, 2011). Ultimately, full nuclear genome-based analyses will be needed to resolve this question satisfactorily.

Although the overall topology of the mtDNA phylogenetic tree from the present study is similar to those from previous studies, the molecular dating estimates differ significantly. This is attributable to the fact that we used well-documented fossil data for node calibration of the phylogenetic tree rather than the molecular estimates by Tougard *et al.* (2001), which are likely to be overestimated by a large margin, as has been shown for the *Equus* genus based on nuclear data (Tougard *et al.*, 2001; Orlando *et al.*, 2013). However, it is worth mentioning that genome-wide data will be required from various rhino species to assess whether these observed differences do, in fact, reflect different nuclear and mtDNA evolutionary histories.

The relatively constant effective female population size of the Javan rhinos for the past ~300 000 years (until 150 years ago) indicated that the dramatic decline of their numbers in the past two centuries is attributable solely to anthropogenic factors.

As one might expect, the oldest lineage of the Javan rhinos in our dataset (represented by the sample JR524) was from the specimen sampled in Bhutan, in continental Asia (Table 1). We therefore hypothesize that it might represent an individual of the now-extinct subspecies *R. sondaicus inermis*, although additional samples and the incorporation of nuclear DNA will be needed to test this further.

The dramatic decline of the population size of the Javan rhino in recent years is reflected in the network analysis of the mtDNA sequences, in which we show that the most diverse lineages are represented by individuals that no longer exist, i.e. the historic samples from our dataset and recently extinct *R. sondaicus annamiticus*.

This difference in genetic diversity is likely due to the temporal as opposed to geographic differences

since most (apart from JR27 which has largely unknown "Calcutta?" label) of the *R. sondaicus* lineages originated from Indonesia (Java and Sumatra). This result is similar to a recent study comparing museum specimens of now extinct populations of black rhinos with modern samples, suggesting a general reduction in genetic diversity in modern rhinoceros populations, a consequence of anthropogenic population collapse (Moodley *et al*., 2017). Unfortunately, we were unable to recover the tRNA-Pro gene and partial D loop region from our three least sequenced samples including JR524, which represented the oldest branch in the Javan rhinoceros mtDNA lineage.

In summary, although our dataset is relatively small, reflecting the challenge of obtaining genetic data from *R. sondaicus* owing to the rarity of modern specimens and the poor preservation conditions of the historical material, our results clearly show that the genetic diversity of its mitochondrial lineage has contracted significantly during the past two centuries. With two subspecies already extinct, the importance of survival of the last one (*R. sondaicus sondaicus)* cannot be overstressed. Currently, there are no modern complete mtDNA sequences available from this species. Therefore, we hope that our newly assembled sequences from historical samples might provide a valuable starting dataset, upon which future studies and conservation efforts might be able to build, in order that more insights can be gained into the evolutionary history of this critically endangered species.

ACKNOWLEDGEMENTS

We thank Kieren Mitchell and Adrian Lister for early access to the *Elasmotherium* shotgun data, and members of the *Rhinoceros* genome sequencing consortium for early access to shotgun data from the other species. We also thank Love Dalén for fruitful discussions and facilitating data access. Furthermore, we thank the staff at the Danish National High-Throughput Sequencing Centre for technical assistance in generating the data. We acknowledge ERC Consolidator Grant 681396 'Extinction Genomics' for funding this research. Images of five modern rhinos were produced by and used with permission from Roger Hall: https://www.flickr.com/photos/roger_inkart/albums/72157658765678676/with/21555267631/. The artistic drawings of *Elasmotherium* and *Stephanorhinus* were produced by and used with permission from Dmitry Bogdanov: https://dibgd.deviantart.com/art/Elasmotherium-sibiricum-119777176 and https://www.deviantart.com/dibgd/art/Stephanorhinus-hemitoechus-200077685. The image of the woolly rhino was used with permission from Britannica: https://www.britannica.com/animal/woolly-rhinoceros.

REFERENCES

Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, Malaspinas AS, Margaryan A, Higham T, Chivall D, Lynnerup N, Harvig L, Baron J, Della Casa P, Dąbrowski P, Duffy PR, Ebel AV, Epimakhov A, Frei K, Furmanek M, Gralak T, Gromov A, Gronkiewicz S, Grupe G, Hajdu T, Jarysz R, Khartanovich V, Khokhlov A, Kiss V, Kolář J, Kriiska A, Lasak I, Longhi C, McGlynn G, Merkevicius A, Merkyte I, Metspalu M, Mkrtchyan R, Moiseyev V, Paja L, Pálfi G, Pokutta D, Pospieszny Ł, Price TD, Saag L, Sablin M, Shishlina N, Smrčka V, Soenov VI, Szeverényi V, Tóth G, Trifanova SV, Varul L, Vicze M, Yepiskoposyan L, Zhitenev V, Orlando L, Sicheritz-Pontén T, Brunak S, Nielsen R, Kristiansen K, Willerslev E. 2015. Population genomics of Bronze Age Eurasia. *Nature* **522:** 167–172.

Brook SM, Dudley N, Mahood SP, Polet G, Williams AC, Duckworth JW, Van Ngoc T, Long B. 2014. Lessons learned from the loss of a flagship: the extinction of the Javan rhinoceros *Rhinoceros sondaicus annamiticus* from Vietnam. *Biological Conservation* **174:** 21–29.

Cappellini E, Jensen LJ, Szklarczyk D, Ginolhac A, da Fonseca RAR, Stafford TW, Holen SR, Collins MJ, Orlando L, Willerslev E, Gilbert MTP, Olsen JV. 2012. Proteomic analysis of a Pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *Journal of Proteome Research* **11:** 917–926.

Carøe C, Gopalakrishnan S, Vinner L, Mak SST, Sinding MHS, Samaniego JA, Wales N, Sicheritz-Pontén T, Gilbert MTP. 2018. Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution* **9:** 410–419.

Carroll RL. 1988. Ungulates, edentates, and whales. In: *Vertebrate paleontology and evolution*. New York: WH Freeman and Company, 502–568.

Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga JL, Meyer M. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 15758–15763.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9:** 772.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29:** 1969–1973.

Ersmark E, Orlando L, Sandoval-Castellanos E, Barnes I, Barnett R, Stuart A, Lister A, Dalén L. 2015. Population demography and genetic diversity in the Pleistocene cave lion. *Open Quaternary* **1:** 1–14.

Fernando P, **Polet G**, **Foead N**, **Ng LS**, **Pastorini J**, **Melnick DJ. 2006.** Genetic diversity, phylogeny and conservation of the Javan rhinoceros (*Rhinoceros sondaicus*). *Conservation Genetics* **7:** 439–448.

Gilbert MTP, **Tomsho LP**, **Rendulic S**, **Packard M**, **Drautz DI**, **Sher A**, **Tikhonov A**, **Dalén L**, **Kuznetsova T**, **Kosintsev P**, **Campos PF**, **Higham T**, **Collins MJ**, **Wilson AS**, **Shidlovskiy F**, **Buigues B**, **Ericson PGP**, **Germonpré M**, **Götherström A**, **Iacumin P**, **Nikolaev V**, **Nowak-Kemp M**, **Willerslev E**, **Knight JR**, **Irzyk GP**, **Perbost CS**, **Fredrikson KM**, **Harkins TT**, **Sheridan S**, **Miller W**, **Schuster SC. 2007.** Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* **317:** 1927–1930.

Groves CP, **Leslie DM. 2011.** *Rhinoceros sondaicus* (Perissodactyla: Rhinocerotidae). *Mammalian Species* **43:** 190–208.

Haryono M, **Miller PS**, **Lees C**, **Ramono W**, **Purnomo A**, **Long B**, **Sectionov WI**, **Aji BD**, **Talukdar B**, **Ellis S. 2016.** *Population and habitat viability assessment for the Javan rhino*. Apple Valley: IUCN/SSC Conservation Breeding Specialist Group.

Heled J, **Drummond AJ. 2008.** Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology* **8:** 289.

Jónsson H, **Ginolhac A**, **Schubert M**, **Johnson PLF**, **Orlando L. 2013.** mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29:** 1682–1684.

Khairani KO, **Nydam D**, **Felippe MJ**, **McDonough P**, **Barry J**, **Mahmud R**, **Haryono M**, **Radcliffe RW. 2018.** Surveillance for hemorrhagic septicemia in buffalo (*Bubalus bubalis*) as an aid to range expansion of the Javan rhinoceros (*Rhinoceros sondaicus*) in Ujung Kulon National Park, Indonesia. *Journal of Wildlife Diseases* **54:** 14–25.

Kirillova IV, **Chernova OF**, **Van der Made J**, **Kukarskih VV**, **Shapiro B**, **Van der Plicht J**, **Shidlovskiy FK**, **Heintzman PD**, **Van Kolfschoten T**, **Zanina OG. 2017.** Discovery of the skull of *Stephanorhinus kirchbergensis* (Jäger, 1839) above the Arctic Circle. *Quaternary Research* **88:** 537–550.

Kosintsev P, **Mitchell KJ**, **Devièse T**, **Van der Plicht J**, **Kuitems M**, **Petrova E**, **Tikhonov A**, **Higham T**, **Comeskey D**, **Turney C**, **Cooper A**, **Van Kolfschoten T**, **Stuart AJ**, **Lister AM. 2019.** Evolution and extinction of the giant rhinoceros *Elasmotherium sibiricum* sheds light on late Quaternary megafaunal extinctions. *Nature Ecology & Evolution* **3:** 31–38.

Lacombat F. 2005. The evolution of the rhinoceros. *Save the rhinos: EAZA Rhino Campaign 2005/6 info pack*. London: Save the Rhinos, 46–49.

Li H, **Durbin R. 2009.** Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25:** 1754–1760.

Li H, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R**, **1000 Genome Project Data Processing Subgroup**. **2009.** The sequence alignment/map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

MacFadden BJ. 2005. Fossil horses – evidence for evolution. *Science* **307:** 1728–1730.

Mak SST, **Gopalakrishnan S**, **Carøe C**, **Geng C**, **Liu S**, **Sinding MHS**, **Kuderna LFK**, **Zhang W**, **Fu S**, **Vieira FG**, **Germonpré M**, **Bocherens H**, **Fedorov S**, **Petersen B**, **Sicheritz-Pontén T**, **Marques-Bonet T**, **Zhang G**, **Jiang H**, **Gilbert MTP. 2017.** Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. *GigaScience* **6:** 1–13.

Meyer M, **Kircher M. 2010.** Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010:** db.prot5448.

Miller MA, **Pfeiffer W**, **Schwartz T. 2010**. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)*: 1–8. Available at: http://www.phylo.org/sub_sections/portal/cite.php

Mohd Salleh F, **Ramos-Madrigal J**, **Peñaloza F**, **Liu S**, **Mikkel-Holger SS**, **Riddhi PP**, **Martins R**, **Lenz D**, **Fickel J**, **Roos C**, **Shamsir MS**, **Azman MS**, **Burton KL**, **Stephen JR**, **Wilting A**, **Gilbert MTP. 2017.** An expanded mammal mitogenome dataset from Southeast Asia. *GigaScience* **6:** 1–8.

Moodley Y, **Russo IRM**, **Robovský J**, **Dalton DL**, **Kotzé A**, **Smith S**, **Stejskal J**, **Ryder OA**, **Hermes R**, **Walzer C**, **Bruford MW. 2018.** Contrasting evolutionary history, anthropogenic declines and genetic contact in the northern and southern white rhinoceros (*Ceratotherium simum*). *Proceedings of the Royal Society B: Biological Sciences* **285:** 20181567.

Orlando L, **Ginolhac A**, **Raghavan M**, **Vilstrup J**, **Rasmussen M**, **Magnussen K**, **Steinmann KE**, **Kapranov P**, **Thompson JF**, **Zazula G**, **Froese D**, **Moltke I**, **Shapiro B**, **Hofreiter M**, **Al-Rasheid KAS**, **Gilbert MTP**, **Willerslev E. 2011.** True single-molecule DNA sequencing of a Pleistocene horse bone. *Genome Research* **21:** 1705–1719.

Orlando L, **Ginolhac A**, **Zhang G**, **Froese D**, **Albrechtsen A**, **Stiller M**, **Schubert M**, **Cappellini E**, **Petersen B**, **Moltke I**, **Johnson PLF**, **Fumagalli M**, **Vilstrup JT**, **Raghavan M**, **Korneliussen T**, **Malaspinas AS**, **Vogt J**, **Szklarczyk D**, **Kelstrup CD**, **Vinther J**, **Dolocan A**, **Stenderup J**, **Velazquez AMV**, **Cahill J**, **Rasmussen M**, **Wang X**, **Min J**, **Zazula GD**, **Seguin-Orlando A**, **Mortensen C**, **Magnussen K**, **Thompson JF**, **Weinstock J**, **Gregersen K**, **Røed KH**, **Eisenmann V**, **Rubin CJ**, **Miller DC**, **Antczak DF**, **Bertelsen MF**, **Brunak S**, **Al-Rasheid KAS**, **Ryder O**, **Andersson L**, **Mundy J**, **Krogh A**, **Gilbert MTP**, **Kjær K**, **Sicheritz-Ponten T**, **Jensen LJ**, **Olsen JV**, **Hofreiter M**, **Nielsen R**, **Shapiro B**, **Wang J**, **Willerslev E. 2013.** Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499:** 74–78.

Orlando L, **Leonard JA**, **Thenot A**, **Laudet V**, **Guerin C**, **Hänni C. 2003.** Ancient DNA analysis reveals woolly rhino

evolutionary relationships. *Molecular Phylogenetics and Evolution* **28:** 485–499.

Pääbo S. 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America* **86:** 1939–1943.

Prothero DR, **Schoch RM. 1989.** *The evolution of perissodactyls*. Oxford: Oxford University Press.

Rambaut A, **Drummond AJ**, **Xie D**, **Baele G**, **Suchard MA. 2018.** Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67:** 901–904.

Ritchie AM, **Lo N**, **Ho SYW. 2017.** The impact of the tree prior on molecular dating of data sets containing a mixture of inter- and intraspecies sampling. *Systematic Biology* **66:** 413–425.

Rookmaaker LC. 1997. Records of the Sundarbans rhinoceros (*Rhinoceros sondaicus inermis*) in India and Bangladesh. *Pachyderm* **24:** 37–45.

Schubert M, **Ermini L**, **Der Sarkissian C**, **Jónsson H**, **Ginolhac A**, **Schaefer R**, **Martin MD**, **Fernández R**, **Kircher M**, **McCue M**, **Willerslev E**, **Orlando L. 2014.** Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols* **9:** 1056–1082.

Schubert M, **Ginolhac A**, **Lindgreen S**, **Thompson JF**, **Al-Rasheid KAS**, **Willerslev E**, **Krogh A**, **Orlando L. 2012.** Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13:** 178.

Schubert M, **Lindgreen S**, **Orlando L. 2016.** AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes* **9:** 88.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312–1313.

Steiner CC, **Ryder OA. 2011.** Molecular phylogeny and evolution of the Perissodactyla. *Zoological Journal of the Linnean Society* **163:** 1289–1303.

Tougard C, **Delefosse T**, **Hänni C**, **Montgelard C. 2001.** Phylogenetic relationships of the five extant rhinoceros species (Rhinocerotidae, Perissodactyla) based on mitochondrial cytochrome *b* and 12S rRNA genes. *Molecular Phylogenetics and Evolution* **19:** 34–44.

Welker F, **Smith GM**, **Hutson JM**, **Kindler L**, **Garcia-Moreno A**, **Villaluenga A**, **Turner E**, **Gaudzinski-Windheuser S. 2017.** Middle Pleistocene protein sequences from the rhinoceros genus *Stephanorhinus* and the phylogeny of extant and extinct Middle/Late Pleistocene Rhinocerotidae. *PeerJ* **5:** e3033.

Willerslev E, **Gilbert MTP**, **Binladen J**, **Ho SYW**, **Campos PF**, **Ratan A**, **Tomsho LP**, **da Fonseca RR**, **Sher A**, **Kuznetsova TV**, **Nowak-Kemp M**, **Roth TL**, **Miller W**, **Schuster SC. 2009.** Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. *BMC Evolutionary Biology* **9:** 95.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Figure S1. Maximum parsimony tree. All available ($N = 36$) rhino whole mitochondrial DNA sequences (published + new sequences from this study) were used for the analysis. Two *Equus* mitochondrial DNA sequences were included as the outgroup. All branches within a species level have been collapsed to improve readability. The node labels indicate bootstrap support values.

Table S1. List of published rhinoceros whole mitochondrial DNA sequences used as the comparative dataset for the phylogenetic analysis. Two *Equus* samples were used as the outgroup.

Table S2. List of rhinoceros partial D-loop mitochondrial DNA sequences used for the Neighbor-Joining network analysis.