Recent mitochondrial lineage extinction in the critically endangered Javan rhinoceros

Margaryan, Ashot; Sinding, Mikkel Holger S.; Shanlin, L. I.U.; Vieira, Filipe Garrett; Chan, Yvonne L.; Nathan, Senthivel K.S.S.; Moodley, Yoshan; Bruford, Michael W.; P. Gilbert, M. Thomas

Published in:
Zoological Journal of the Linnean Society

DOI:
10.1093/zoolinnean/zlaa004

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC

Citation for published version (APA):
Recent mitochondrial lineage extinction in the critically endangered Javan rhinoceros

ASHOT MARGARYAN1,2*, MIKKEL-HOLGER S. SINDING1,3, SHANLIN LIU1,4, FILIPE GARRETT VIEIRA1, YVONNE L. CHAN5, SENTHILVEL K. S. S. NATHAN6, YOSHAN MOODLEY7, MICHAEL W. BRUFORD8 and M. THOMAS P. GILBERT1,9

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


INTRODUCTION

The Javan rhinoceros (Rhinoceros sondaicus Desmarest, 1822) is the rarest of the five extant rhinoceros (rino) 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...
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
which new mtDNA-based assays might be designed into its former diversity and for future studies, in species will be valuable both for providing insights mtDNA sequences from this rare and understudied to a relatively small sample size, we hope that these reveal the loss of mtDNA diversity in Javan rhinos 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.

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; Tougaard 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 (Tougaard 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 genomescale 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 Alenoot 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, Dicerorhinus 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).

BIOMINFORMATICS

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.
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 rhinoceroses 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; MacPadden, 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_f) 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
the mapping results are shown in Table 3. For five out of eight historical samples, the depth of coverage (DoC) of mtDNA was > 24X, 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 (1.64–1.96X), 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' end 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%
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)
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...
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 (Tougaard 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) 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/721576587656786767/with/21555267631/. The artistic drawings of Elasmotherium and Stephanorhinus were produced by and used with permission from Dmitry Bogdanov: https://digbd.deviantart.com/art/Elasmotherium-sibiricum-119777176 and https://www.deviantart.com/digbd/art/Stephanorhinus-hemitoechus-200077685. The image of the woolly rhino was used with permission from Britannica: https://www.britannica.com/animal/woolly-rhinoceros.

REFERENCES


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.