Sanchez-Barreiro, F., 2020. The missing horns of Africa: A temporal, population genomics study of the white and black rhinoceroses. Ph.D. Thesis presented to Globe Institute, University of Copenhagen, Denmark, pp. 1-170

This thesis consists of the following parts.

PhD thesis | Fátima Sánchez Barreiro

Supervisor: Professor M. Thomas P. Gilbert

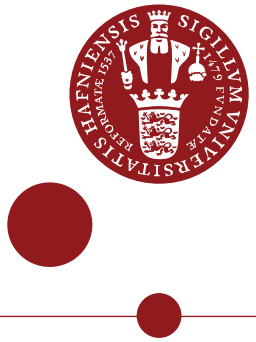
Co-supervisor: Professor Yoshan Moodley

Submitted on: May 2020

UNIVERSITY OF COPENHAGEN

The Missing horns of Africa

A temporal, population genomics study of the white and the black rhinoceroses



PhD thesis

The missing horns of Africa

A temporal, population genomics study of the white and the black rhinoceroses

Fátima Sánchez Barreiro

May 2020

Host institution: University of Copenhagen
Department: The GLOBE Institute
Section: Section for Evolutionary Genomics

Author: Fátima Sánchez Barreiro
Title: The missing horns of Africa
A temporal, population genomics study of the white and the black rhinoceroses
Submitted on: May 2020

Supervisor: Professor M. Thomas P. Gilbert
Co-supervisor: Professor Yoshan Moodley
Funding: ERC Consolidator Grant 681396 ‘Extinction Genomics’ to M. Thomas P. Gilbert and
EMBO Short-Term Fellowship 7578 to Fátima Sánchez Barreiro

Cover and back cover illustrations:
Elvira Sánchez Barreiro, alias Treboada

To all the rhinos left in this world.

May they continue being horny.

Preface

Rhinos appeared in my life rather unexpectedly. Three and a half years ago the most accurate biological statement I could make about them was that they belong to the Animal Kingdom. After some time entertaining the thought of directing my unborn research career toward plant genomics and evolution, being suggested to work on rhinos felt too exotic to be legit. My parents certainly did not take this turn in my academic path with a smile… But it sounded all too exciting and challenging and odd, so I jumped right in.

My supervisor, Tom, knowing of my inclination toward the veggie world, recommended that I thought of these beasts as broccoli, ‘just big, grey, angry broccoli’. I quite like eating broccoli though, so for most of my PhD I refrained from seeking comfort in this comparison. In fact, it turned unnecessary, because rhinos became the most motivating and cherished part of this endeavour.

This rhino project came with a certain aura, that of population decline, near-extinction, and the potential threats associated with that. To help me dive into this theoretical background, I was lucky to be involved in the writing of an opinion paper about these matters at the start of my PhD. This work is included in this thesis as Chapter 1, and connects population genomics to conservation by proposing an approach to measure genomic erosion with temporally sampled genomic data.

I used much of this framework when I moved on to focus on the empirical, population genomics study of the African rhinos, the main objective in my PhD plan. This constituted the bulk of my work, and entailed a phase of generation of genomic data from historical and modern samples, and the subsequent data analysis and interpretation of results. The presentation of the work is split into two more research chapters, each devoted to one of the African rhino species, Chapter 2 for the white rhino, and Chapter 3 for the black rhino, owing to the different research questions each of their stories spurred.

My PhD education, however, involved other activities as well, such as coursework, teaching and supervising, preparing and giving talks, and discussing with peers. And the workflow in most research and related ventures entailed trying, failing, re-thinking, realising, reflecting and, above all, learning. It might sound redundant and obvious that the thing I did the most during my PhD education was learning, but the intensity, breadth and speed of the learning process have been

amazing, way beyond what I could have expected (and I expected quite a lot). I hope I have been able to express all that along this dissertation.

Doing a PhD has been a strange journey, full of discoveries of all kinds, some joyful and rewarding, some contradictory and disappointing. All along, however, the story of the African rhinos has been a fertile substrate for the growth of skills, insight, reflections and connections to other parts of the world. As nourishing as broccoli, if not more.

Fátima Sánchez Barreiro
May 2020, Copenhagen (Denmark)

Acknowledgements

If someone were to ask me what I have learnt the most about during my PhD education, I would say that about myself. And I am quite certain many other PhD candidates and graduates would reply similarly. As a *millennial* who has never 'left school' in trying to postpone growing up, this degree has been a journey toward and around herself. Therefore, with an absolute lack of modesty, the first person I would like to acknowledge is myself: for arriving at this destination maybe not triumphantly, but certainly more resilient, cynical and wiser.

Throughout this PhD journey I have crossed various landscapes, the Cape of Beginner's Hope, the Hills of Time Management, Conference-Abroad Beach, Submission swamp... All these landscapes were populated, and luckily for me, by people who have made this trip enriching and exciting. These pages contain an account of my gratitude to all of them.

Roads, bridges and infrastructure in general are essential for having smooth travels, so I would like to thank the funding bodies that financed these three and a half years of work, the European Research Council (ERC), and the European Molecular Biology Organization (EMBO). I am also thankful to the institutions that hosted me, mostly the University of Copenhagen (Denmark), but also the Institut de Biologia Evolutiva (IBE, Catalonia), the Swedish Museum of Natural History (Sweden) and the University of Venda (South Africa). And not to be forgotten, I thank all the museums that contributed samples to this project, and the Danish National Supercomputer for Life Sciences (Computerome), thanks to which I could conduct the data analysis.

I would not have embarked on this expedition had it not been for an expert travel agent and his magnetic ideas. I am very glad Tom Gilbert trusted me to carry out this project, and I sincerely appreciate all the wonderful opportunities he has offered me, all the advice and, of course, the displays of wit. However, neither Tom nor I knew enough about the territory we were venturing into, so I was fortunate to have Yoshan Moodley as a travel guide. He introduced me to what actually matters in life: wildlife, conservation genetics, and safariying. I can only be grateful for his infinite enthusiasm (and the supermarket bag full of rhino samples!).

Of course this endeavour was far from being the product of the good will and work of just three people. Lots have participated, contributed intellectually, helped generously and given advice. As the wandering traveller who is given guidance by kind locals, I am thankful to all of them for making this project thrive and for fueling my learning curve: Shyam, Jazmín, Mick, Marc, Ashot,

Marta, Claudia, Filipe… Also, two people in particular helped me in the making of this travel journal that you are reading: George, whose feedback I have very much appreciated, and Camilla who put the Scandinavian touch by translating the Summary into Danish.

A few hosts generously sheltered me for a while. The hospitality of Love Dalén, Yoshan and Tomàs Marquès-Bonet has been key to my PhD because their homes were full of inhabitants from whom I learnt enormously. I am very thankful to all the members of Love's group for the kind help, the fun, and for the unforgettable weekend on top of a rock in the middle of a fjord writing papers and listening to ABBA. From my time in South Africa, I particularly cherished the determination and the eagerness of the students at Moodley Lab to learn. I arrived at Port Enthusiasm when I needed it the most.

When it comes to the people at IBE, my second academic home, I have to confess the profound influence that they have had on me, especially those in the group of Tomàs. Discussions, journal clubs, seminars, and drinks at Bambú have been spontaneous celebrations during my travels. Having met and learnt from these bright guys has been one of the most rewarding bits of my PhD.

Besides the treks to other lands, my academic development benefited immensely from joining the *popgen* community in Copenhagen. There have been epic moments, where I have been fortunate to drink from these brilliant minds while they drank from their beer cans.

Hills and strong winds were part of this journey as well. More than once my hiking boots started to hurt, and I wondered why on Earth I began this trip. In those moments, it has been most comforting to feel part of a large community of travelers, who faced similar troubles and were happy to share stories and advice. In that regard, I am thankful to my writers' group, a refuge in my last months of PhD, and to the inspiring people I met at UCAPS (the University of Copenhagen Association of PhD Students), for the fun times, and for enticing me to reflect about what a PhD education is, and what we would like it to be.

In every course, symposium, conference I went to, I always encountered friendly, smart, motivating people, and I am grateful for every discussion and enlightening perspective. I would like to mention in particular the gang *Las Meninas de Cesky*, the child of a bizarre mating between Czech nightlife and population genomics theory.

My base camp during these years of PhD has been the vibrant Section for EvoGenomics. Despite a couple of affiliation changes and a relocation, its spirit has remained the same: open, lively, diverse,

fascinating. Some of the comrades here, I have been travelling with for quite some time, others were short-term buddies. My gratitude to all these wonderful people is vast. It would be hard to cite them all, and if I missed someone I would not forgive myself. So I trust that they know who they are; I have tried to let them see during the past few years how important they are to me.

But there is more to life than academia and its environments (right...?). When it came to taking a rest, I detoured to cosy places where I got by with a little help from my friends. I am endlessly thankful to them (Ro, Agata, Alex, Sabiha, Stephanie, Sabe, Martes, Zaira...) for helping me recalibrate my compass every now and then.

Also, a thoughtful, fun and supportive travel companion can make the way much more enriching. In fact, some of the background reflection poured into this work evolved thanks to it being shared with one such companion. What can I say? Thanks Jonathan.

During this journey, I regularly received postcards from some very special senders whom, from far away, trusted that I knew what I was doing. They never ask too much, they always care, and they cheer the loudest. If there is any glory in this endeavour, some of it definitely belongs to my parents.

In my PhD bubble, I would have forgotten what a good sense of humour and political *incorrectness* are had it not been for my sister. I could not be more glad about how our big-sister-young-sister relationship evolved in the past few years. Plus she has given me the most amazing PhD thesis cover ever seen.

I feel privileged I had the chance to travel as a PhD backpacker, but I am also glad that the trip is over, and that I can now recount what I have learnt. So to everyone and everything that somehow helped me entering and getting through this adventure, *graciñas*!

English summary

Human activities are fueling a global defaunation phenomenon to which large vertebrates are particularly sensitive. They have suffered rapid and dramatic population declines (i.e. bottlenecks), and have been cornered into tiny remnant fragments of their historical ranges of distribution. Such demographic histories are hypothesised to erode genomic diversity, and alter population structure and gene flow in these species. Evaluating these potential consequences is critical for future species management and conservation, and to achieve this we ideally need genomic information from not only current, but also from past populations. In this thesis, I applied such an approach to the case of the African rhinoceroses. Both the white rhinoceros (*Ceratotherium simum* Burchell) and the black rhinoceros (*Diceros bicornis* L.) underwent population extirpations and bottlenecks in a brutal human-wildlife conflict through the past two centuries, and are now dependent on active conservation efforts. I generated whole-genome data from both historical and modern samples, and complemented this with extra available genomic data from various modern samples. This yielded a final dataset of 143 re-sequenced genomes, which I then analysed in a population genomics framework. The power of coupling museum samples with palaeogenomic DNA sequencing methods allowed me to study these two rhinoceros species from a temporal, intra-species perspective, and in doing so pinpoint changes in their genomic makeup due to their recent demographic history.

Following an **Introduction** to the relevant background and the aims of the thesis, I present three research chapters that describe the bulk of the work I conducted during my PhD education.
In **Chapter 1**, my collaborators and I formalised a conceptual and methodological framework in which to assess genomic erosion in threatened species, relying on comparing genomic data from pre- and post-decline samples. In **Chapter 2**, I present an assessment of genomic erosion in the white rhinoceros, following closely the approach proposed in Chapter 1. By comparing pre- and post-bottleneck genomic data of 52 northern and the southern white rhinoceroses (NWR and SWR respectively), I discovered previously unknown patterns of population structure driven by geography in the pre-bottleneck NWR and SWR, as well as significant symptoms of genomic erosion among the post-bottleneck individuals. In **Chapter 3**, the focus shifts to the historical biogeography of the black rhinoceros. With whole-genome data from 64 historical samples, I found that six populations of black rhinoceros existed before the species' precipitous decline, following a latitudinal cline with two major discontinuities. With the resulting genomic-geographic historical map, I identified the origin of 27 re-sequenced genomes from extant populations.
Lastly, a section of **Conclusions and perspectives** closes this dissertation with an attempt to place this work in a broader context, particularly that of applied conservation.

Dansk resumé

Menneskelig aktivitet påvirker et allerede eksisterende globalt fænomen: defaunation, hvor det især er de store vertebrater som er specielt sensitive. Disse har lidt hurtige og voldsomme fald i populationsstørrelse (i.e. flaskehalse), og er blevet klemt inde i resterende små fragmenter af deres ellers tidligere historiske udbredelsesområder. Hypoteser lyder, at et sådan mønster kan erodere genomisk diversitet, samt ændre populationsstrukturen og genflowet i disse arter. Det er yderst vigtigt at evaluere disse potentielle konsekvenser for at fremme fremtidig artsbevarelse, og for at opnå dette har vi ideelt set behov for genomisk information fra de nuværende populationer såvel som også de historiske populationer. I denne afhandling har jeg benyttet en sådan fremgangsmåde til at undersøge det afrikanske næsehorn. Igennem de sidste to århundreder har en brutal konflikt imellem mennesker og dyr, resulteret i udryddelsen af populationer og flaskehalse for både det hvide næsehorn (*Ceratotherium simum* Burchell) og sorte næsehorn (*Diceros bicornis* L), som nu begge er afhængige af en aktiv bevaringsindsats. Jeg genererede data for hele genomer af både historiske og moderne prøver, og komplementerede dette med data fra yderligere tilgængelige genomer fra forskellige moderne prøver. Dette gav et endeligt datasæt på 143 re-sekventerede genomer, som jeg derefter analyserede fra et populationsgenetisk perspektiv. Det at kunne koble museums prøver sammen med palæogenomisk DNA sekventering, gjorde det muligt for mig at undersøge disse to arter af næsehorn fra et tidsmæssigt, artspecifikt perspektiv, og dermed lokalisere ændringer i deres genomiske sammensætning som skyldes deres nylige demografiske historie.

Efter en **Introduktion** bestående af den relevante baggrund og formål med afhandlingen, vil jeg præsentere tre forskningskapitler som beskriver hovedparten af det arbejde jeg udførte under min Ph.d.-uddannelse.
I **Kapitel 1**, har mine samarbejdspartnere og jeg formaliseret en konceptuel og metodisk ramme til at vurdere genomisk erodering hos truede arter, på baggrund af at sammenligne genom data fra prøver præ- og post-tilbagegangen i populationsstørrelse. I **Kapitel 2**, præsenterer jeg en vurdering af genomisk erodering i det hvide næsehorn, som følger fremgangsmåden foreslået i kapitel 1. Ved at sammenligne genom data fra præ- og post- flaskehals af 52 nordlige og sydlige hvide næsehorn (henholdsvis NWR og SWR), fandt jeg hidtil ukendte mønstre af populationsstruktur som er drevet af geografi i præ-flaskehals NWR og SWR, samt signifikante symptomer af genomisk erodering blandt individer post-flaskehals. I **Kapitel 3**, ændrer fokusset sig til den historiske biogeografi af det sorte næsehorn. Med genom data fra 64 historiske prøver, fandt jeg, at der eksisterede seks populationer af det sorte næsehorn som fulgte en breddegrads gradient med to større diskontinuiteter, før artens bratte fald. Baseret på denne information, kunne jeg lave et genom-geografisk kort som jeg brugte til at identificere den geografiske oprindelse af 27 re-sekventerede moderne individer.
Til sidst, for at afslutte denne afhandling, har jeg en sektion bestående af **Konklusion og perspektivering** med et forsøg på at placere dette arbejde i en bredere kontekst, især det om anvendt artsbevarelse.

Translated by Camilla Hjorth Scharff-Olsen

Table of contents

Introduction & Aims, contributions and structure

Introduction

This work aims to be a contribution to the study of threatened species in light of their recent demographic histories. In particular, the present dissertation looks through the lens of population genomics at what African rhinoceroses are and used to be. Historical sampling is a cornerstone of this project, in conjunction with population genetics theory and modern population genomics analytical tools. The sections below introduce the relevant concepts and the background stories that frame this project.

1. Big, cute and yet so fragile

1.1 The decline of the megafauna: old stories and new threats

There is no consensus on a precise definition of megafauna, but as the name suggests it refers to wild, vertebrate species of large body size. Some authors include all species ≥44 kg in this definition [1,2], others take into account the relative body size difference within their taxonomic class [3]. More encompassing, trophic level-dependent definitions have also been proposed such as in where megafauna includes “large herbivores (45–999 kg), megaherbivores (≥1000 kg), large carnivores (21.5–99 kg), and megacarnivores (≥100 kg)” [4].

Under all definitions, however, an identifiable decline of this large-bodied fauna happened right before and during the Pleistocence-Holocene transition, between ~50 kya and ~8 kya [1,4]. Several waves of extinction struck different continental regions in a strongly size-selective manner [5], with the subsequent downscaling of vertebrate size altering entire ecosystems [2,4].

The timing, intensity and specific mechanisms of this phenomenon varied across continents. Australia was the first area to experience a burst of extinctions between ~50-40 kya, coinciding with the arrival of *Homo sapiens* to the region, and no unusual climatic shifts occurring [1]. Later, two waves of extinction happened in Eurasia, between 48-23 kya and, more intensely, between 14-10 kya, in overlap with the arrival and then the increase in abundance of *H. sapiens*, but also with the two major climate cooling events, the Last Glacial Maximum (LGM) and the Young Dryas (YD) [1]. The Americas suffered a megafaunal collapse between 13-11 kya in North America and between 12-8 kya in South America, where again the arrival and expansion of *H. sapiens* coincided with an oscillating climate [1]. In Africa, where *H. sapiens* and other hominins had long existed, and where climatic oscillations were milder, megafaunal extinction was not so severe [1].

This late-Pleistocene downturn of the megafauna happened within the frame of a rapidly shifting climate, which has been largely considered a plausible culprit. However, these events also happened during the expansion of modern humans across the Globe, and their increase in population size. Megafaunal species are particularly likely to enter into conflict with humans, owing to harvesting of these species or competing use of habitat [3]. The characteristic size-selectivity of this extinction, the co-occurrence of *H. sapiens* settlement and expansion, archaeological findings indicating the role of modern humans as hunters of megafauna, among other lines of evidence, suggest a prominent role of humans in the late-Pleistocene megafaunal decline [5].

Although this megafaunal extinction can be seen as an isolated event in geological time-scales, the fact is that the current trend of biodiversity seems to be a continuation of this phenomenon [1]. Biodiversity loss during recent times has been quantified, monitored and ascribed to particular causes, and the magnitude of what we are witnessing is so severe that it has been coined *the Sixth Mass Extinction* [6,7].

Current extinction rates are up to 100-fold higher than conservative estimates of background extinction rates [6], and even if extinction rate is a suitable proxy for quantification of biodiversity loss, *defaunation* is an equally alarming phenomenon [8]. Loss of wildlife in general has become a pervasive phenomenon that is depleting ecosystems and enlarging catalogues of threatened species *via* range contractions, decreased species abundance and population extirpations [8,9]. As a compelling example, population sizes of vertebrates have overall dropped by 60% between 1970 and 2014 according to the Living Planet Index Report of 2018 [10] (Figure 1).

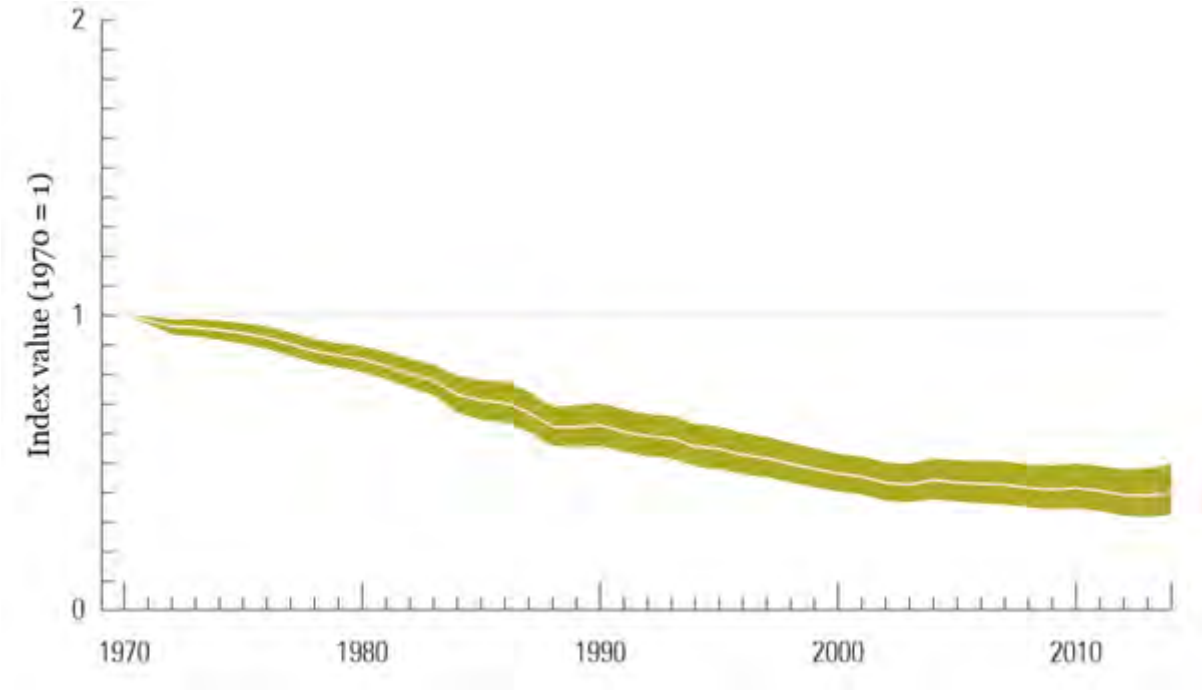


Figure 1. *Evolution of the global Living Planet Index (LPI) from 1970 until 2014.* This index encapsulates the size trends of over 20,000 populations belonging to more than 4,200 vertebrate species. It was developed by the Zoological Society of London (ZSL) and the World Wildlife Fund (WWF), and is often included in guidelines for policy-making regarding the preservation of biodiversity. Graph from the Living Planet Report 2018 [10].

If cosmic events and geological phenomena were the usual culprits of mass extinctions, with the added 'human touch' to the late-Pleistocene megafaunal decline, this sixth event comes in a very different flavour. Human disturbance of natural environments and their biological communities has accelerated remarkably since the beginning of the industrial revolution. As human population size, resource extraction and technological advancement grew, wild biodiversity biomass followed a negative trend [1]. Denoting the relevance of human impact on Earth, the current geological period has started to be referred to as the Anthropocene [11].

Although a generalised downturn is observed across numerous taxa and all trophic levels [9], once again, species of larger body sizes are more susceptible to entering into conflict with humans and suffer more intense consequences than smaller animal species [3] (Figure 2).

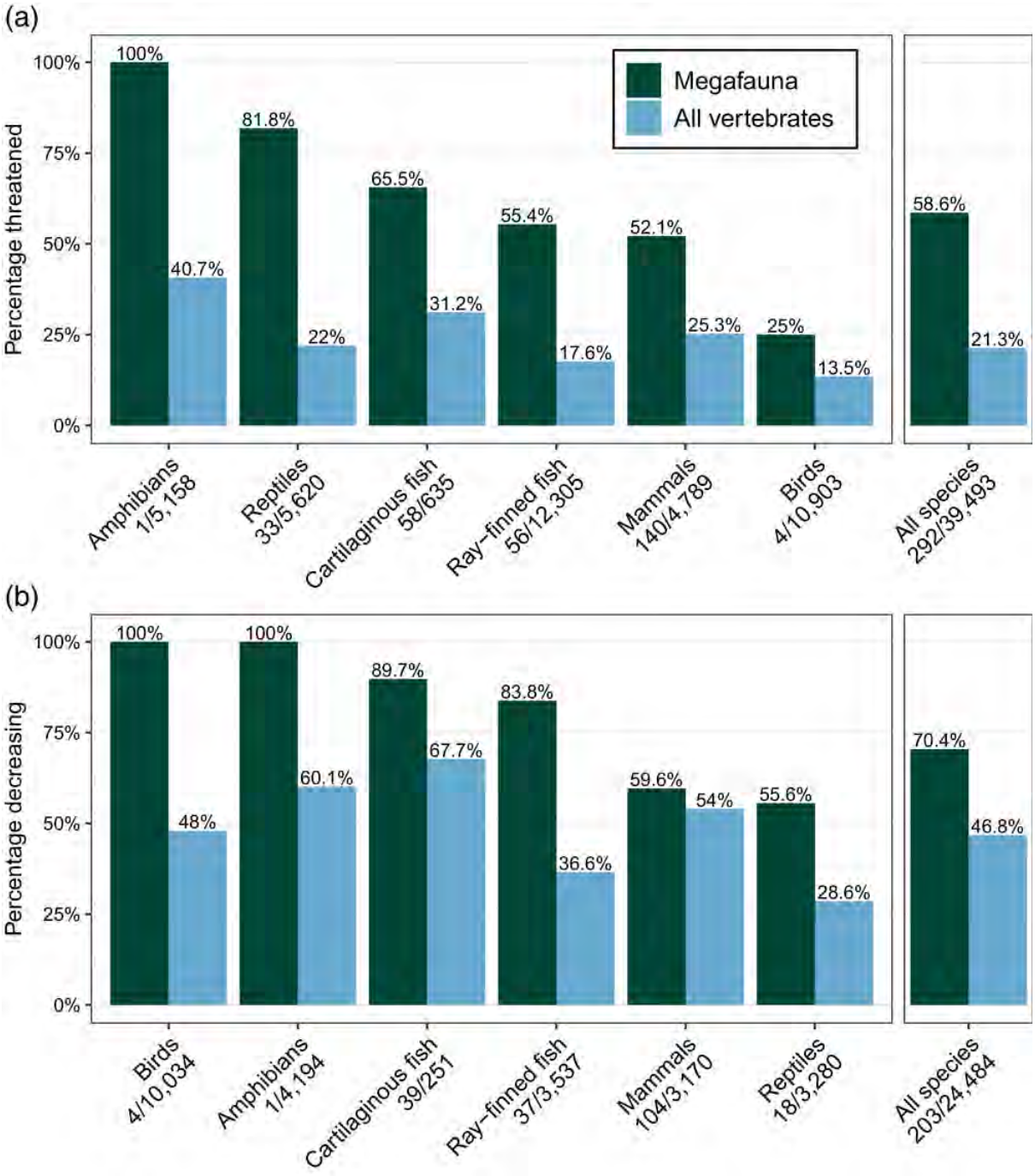


Figure 2. *Levels of threat across vertebrate groups at large, and the megafauna in particular.* The proportion of vertebrate species that are threatened (a), and those that show a decreasing population trend (b) according to the IUCN Red List, for the entire vertebrate group (blue) and for megafaunal species within each group only (green). Megafauna here were mammals, ray-finned fish, and cartilaginous fish ≥100 kg of body mass, and amphibians, birds, and reptiles ≥40 kg. Image extracted from [3].

1.2 Ecological consequences of the loss of the megafauna

The downturn of the megafauna is a major onslaught to global biodiversity by itself, but this depletion has also deep impacts on entire ecosystems. Size matters when it comes to ecological roles, and large-bodied animals are known to be major players on ecosystem functioning and structure [4].

Megacarnivores exert control over the abundance and species composition of herbivores and mesopredators, and the cascading effects on the trophic ladder regulate the stability of the ecosystem [4]. Megaherbivores, on the other hand, are key regulators of vegetation structure and composition, being able to drive shifts from, e.g. woody vegetation to grassland or vice versa [4]. Beyond that, the regulatory role of the megafauna controls large-scale events such as fire regimes and nutrient cycling [2]. In fact, the megafauna has been demonstrated to be crucial in pumping nutrients from ocean bottoms to the surface, and from there back to land; and on land, from high-concentration areas toward nutrient-poor regions [2].

Given their position as key ecological actors, it is not surprising that the loss of megafauna in the late Pleistocene caused substantial changes to whole ecosystems. Co-extinction of species that were establishing parasitic, commensal or mutualistic interactions with megafaunal species is a well recorded phenomenon [12]. Likewise, anachronism is another outcome of megafaunal extinction [12], e.g. large, fleshy fruits of many plant species lack today their means of dispersion without large-bodied vertebrates roaming around [12]. Regarding large-scale effects, it has been shown that global nutrient distribution capacity has decreased to 6% of its value before the late-Pleistocene extinctions [2], and average animal body size has been downgraded [1].

The Earth has suffered a major shift from allocating biomass and energy on numerous species with different ecological niches and roles, to allocating most energy and biomass on just a few: humans and its domesticated companions [1]. Amid this major disturbance, the loss of cornerstone megafaunal species leaves ecosystems devoid of their engineer species, thus entering a phase of major readjustment.

2. A megafaunal saga: rise and fall of the family Rhinocerotidae

There was a time in which rhinoceroses (referred to as *rhinos* throughout the remainder of this text) rocked the terrestrial world. What we see today of the Rhinocerotidae is just some obscurely connected remnants of a once rich, diverse and widely distributed mammal family within the order Perissodactyla.

The family originated in Asia sometime in the Middle Eocene (48-40 mya) (see Figure 3 for a guide on geological epochs), from where it diversified and expanded. Early forms reached North America already in the Middle Eocene; in the Early Oligocene (34-28 mya), coinciding with a pulse of diversification driven by climate cooling, they expanded toward Europe. Lastly, rhinos reached Africa in the Early Miocene (23-16 mya) [13].

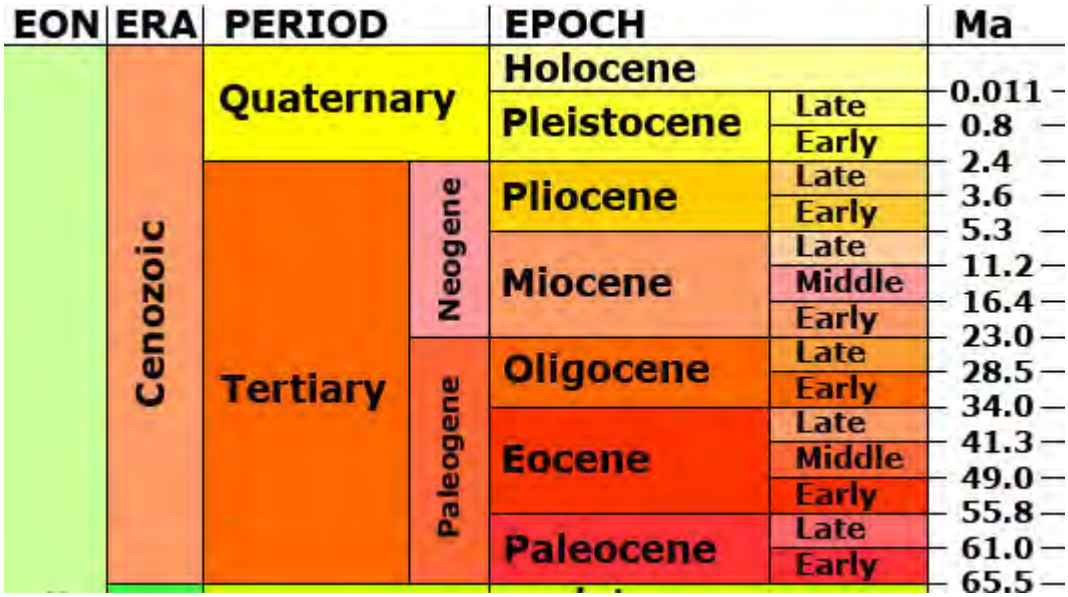


Figure 3. *Cenozoic geological chronology*. The family Rhinocerotidae emerged in the Middle Eocene and their splendour occurred during the Miocene. Image modified from [14].

Several waves of diversification, migration, replacement and extinction happened throughout their evolutionary history, mostly as a consequence of climatic changes [13], but the Miocene (23-5.3 mya) was the golden age of rhinos, as they reached their peaks of species richness and ecological diversity [15]. The end of the Late Miocene (11.6-5.3 mya), however, signified the beginning of the end: the expansion of C4 plant-dominated vegetation, and the drier conditions triggered a faunal turnover during which rhino diversity started to decline. In the Early Pliocene (5.3-3.6 mya), rhinos disappeared from North America [13], and as the climate became colder and drier during the

Pliocene (5.3-2.6 mya) and Pleistocene (2.6-0.01 mya), the rhino family continued losing members, although the iconic, cold-adapted woolly rhino (*Coelodonta antiquitatis*) and some other species (e.g. *Stephanorhinus*) were abundant across Eurasia [13].

The distribution of species richness and average body size showed an inter-continental pattern, although there was a prevalent tendency toward larger body sizes over the course of their evolutionary history. Eurasia, hosted the highest diversity of rhinos, and a broad range of body sizes. The rhino faunas of North America were slightly less diverse, and small-sized species were predominant, despite the tendency to increase over time. In Africa, as a contrast, low species richness and large body sizes have been the rule [13].

The differences in body size were linked to differential ecological strategies, that sort rhinos into three main types: a) small, gregarious browsers, b) medium-large browsers and occasional grazers that colonised many different habitats, c) large grazers with amphibious lifestyles and potentially gregarious [13]. The development of hypsodonty, i.e. teeth adapted to highly abrasive forage, has been another key evolutionary trend, given its strong association with grazing. Many rhino species, especially at the peak of diversity in the Miocene, are speculated to have been ecologically flexible browsers, but their level of hypsodonty would reflect the extent of their tendency to grazing [13]. In today's rhino fauna, browsing is prevalent, but an exclusive, hypsodont grazer exists, the white rhino [16].

The success of the rhino family, in terms of diversification and persistence, is probably linked to their ability to thrive on even very nutrient-poor forage, as well as being flexible in the spectrum between browsing and grazing. Moreover, the evolutionary tendency toward larger body sizes has granted them a reduced predation risk, as well as facilitating living among rough vegetation [17].

The survivors of this past glory include five extant species, three in Asia and two in Africa. The Asian species are namely the Sumatran rhino (*Dicerorhinus sumatrensis*), the Indian rhino (*Rhinoceros unicornis*) and the Javan rhino (*Rhinoceros sondaicus*); the African species are the white rhino (*Ceratotherium simum*) and the black rhino (*Diceros bicornis*) [17].

The evolutionary relationships between the extant species of rhinos have been a conundrum for decades. Both morphological and molecular approaches have yielded contradictory results, and even the inclusion of mitochondrial DNA sequences of extinct rhino species did not resolve the phylogeny [18]. The Indian and Javan rhinos systematically appear as sister taxa, as do the African white and black rhinos, but the position of the Sumatran rhino remains equivocal [18,19]. Phylogenomics analyses might be able to shed light on this long-standing question.

All five rhino species were abundant in their historical ranges of distribution up until the end of the 19th century. From then onward, the relentless increase of anthropogenic pressure over their habitats and populations have pushed them to terrible demographic scenarios [20,21]. The critically endangered Javan rhino is as of today the rarest mammal in the planet, with some 46-66 individuals left in two disjunct locations, one in Indonesia, and one in Vietnam [22]. The Indian rhino, of which ~2000 individuals remain, shows a positive population trend, but inhabits just fragments of its former range and is ranked as vulnerable by the IUCN Red List [23]. The Sumatran rhino is also a vulnerable species whose total number of individuals, scattered across several small populations, ranged between 220 and 275 in 2008 [24]. The recent population histories of the African rhinos are described in detail in Section 3.2, but overall they follow schemes similar to those of their Asian counterparts.

Besides the damaging range fragmentation and habitat loss, rhinos continue to be poached for their horn, which is sold in the black market at exorbitant prices [17]. Therefore, all rhino populations today depend on human management for their survival. Ironically, it is also human intervention that propels them toward extinction.

3. The rhinos of Africa

3.1 Evolution, ecology and biogeography of the white and the black rhinos

Despite the strong connection we commonly draw today between rhinos and the African continent, proportionally few lineages represented the *Rhinocerotidae* family there. Rhinos arrived in Africa from Eurasia first in the Early Miocene (23-16 Mya), but the predecessors of the extant African rhinos entered the continent early in the Pliocene (5.3-2.6 Mya) and replaced all prior rhino diversity [13]. The current rhino fauna in Africa includes two species: the white rhino (*Ceratotherium simum* Burchell, 1817), and the black rhino (*Diceros bicornis* Linnaeus, 1758) [13] (Figure 4).

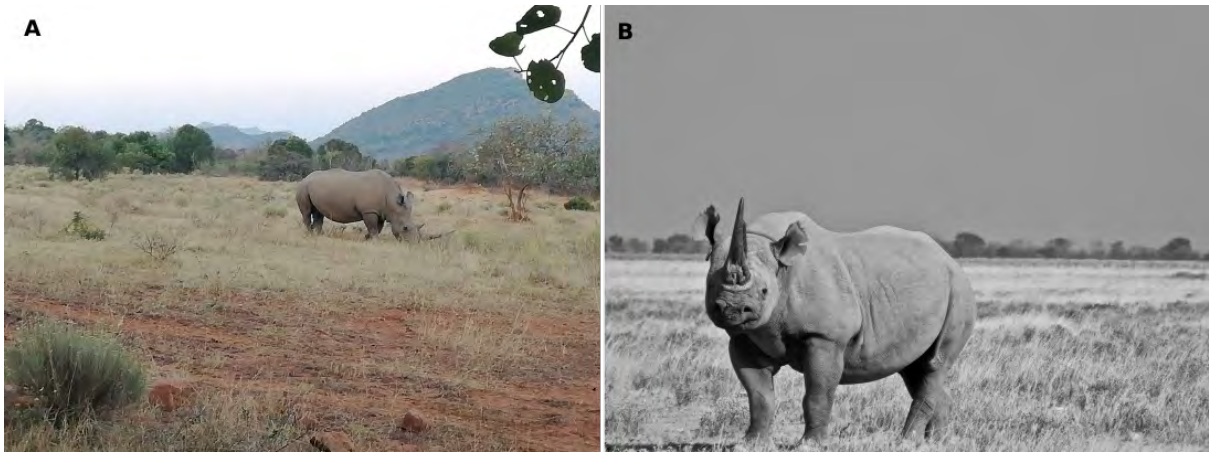


Figure 4. *The two extant African rhino species, the white and the black rhino.* A) White rhino in a private game reserve in South Africa. Photograph by Fátima Sánchez Barreiro, 2019. B) Black rhino in Etosha National Park, Namibia. Photograph by Marta Ciucani, 2016.

The key ecological and evolutionary difference between these two lineages lies in their dietary habits, the white rhino being an obligate grazer, while the black rhino is a browser. The common ancestor of the current African rhinos was probably a mixed feeder, and the specialised strategies likely emerged upon the split of the ancestral lineage shortly after the Miocene-Pliocene boundary (5.3 mya) [25]. Recent research based on whole-genome data suggests that the two emerging rhino lineages experienced gene flow until 3.3-4.1 mya; eventually reproductive isolation was fully established, along with ecological specialisation [26]. Their diverging ecologies underlie key morphological adaptations, and explain the preference of the white rhino for habitats of grassland and savanna, and that of the black rhino for bushland and woodland [25].

The white rhino, as an obligate grazer that feeds on nitrogen-rich short grasses, plays a key role in maintaining the grazing lawns of the grassland plains [27]. The black rhino, on the other hand, can feed on a broad range of food items, namely leaves, twigs, flowers and fruits, so they are relevant seed dispersers in some parts of their range. Also, they avoid areas of dense forest, but otherwise thrive in a remarkable variety of habitats, from moist shrubland to desertic areas [17].

Mouth shape is an evident difference driven by dietary habits between these two species, and has granted each a second vernacular name, the square-lipped (white) rhino and the hook-lipped (black) rhino. This alternative nomenclature might be more appropriate given that both species are in fact grey; the naming of *C. simum* as *white* rhino derives from a mistranslation to English of *wijdt*, meaning ‘wide’ in early Cape Dutch [20].

In accordance with their body size, they require large home ranges, whose extension depends on forage availability and quality [17]. Age at first reproduction lies around 6-7 years, and females remain fertile until the end of their lives. The most durable social bond is that of a mother and her calf, which lasts around 2-3 years, otherwise they are mostly solitary or aggregate only temporarily in small groups [17].

Regarding biogeography, historically the white rhino encompassed two allopatric populations that had not come into secondary contact for millennia [28]: the northern white rhino (*C. s. cottoni*, hereafter referred to as NWR), and the southern white rhino (*C. s. simum*, hereafter referred to as SWR). The NWR inhabited the plains of present-day South Sudan, Northeastern Democratic Republic of the Congo (DRC), Central African Republic and Uganda. The SWR roamed across regions within present-day South Africa, Botswana, Namibia, Zimbabwe and Mozambique (Figure 5A). During the Pleistocene, the white rhino had a larger range of distribution that reached the Mediterranean [16], and was possibly continuous, as pointed by fossil evidence and cave paintings [20]. The current geographical separation of ~2000 km between the ranges of the NWR and the SWR would have appeared due to vegetation changes linked to climatic oscillations of the Pleistocene [28].

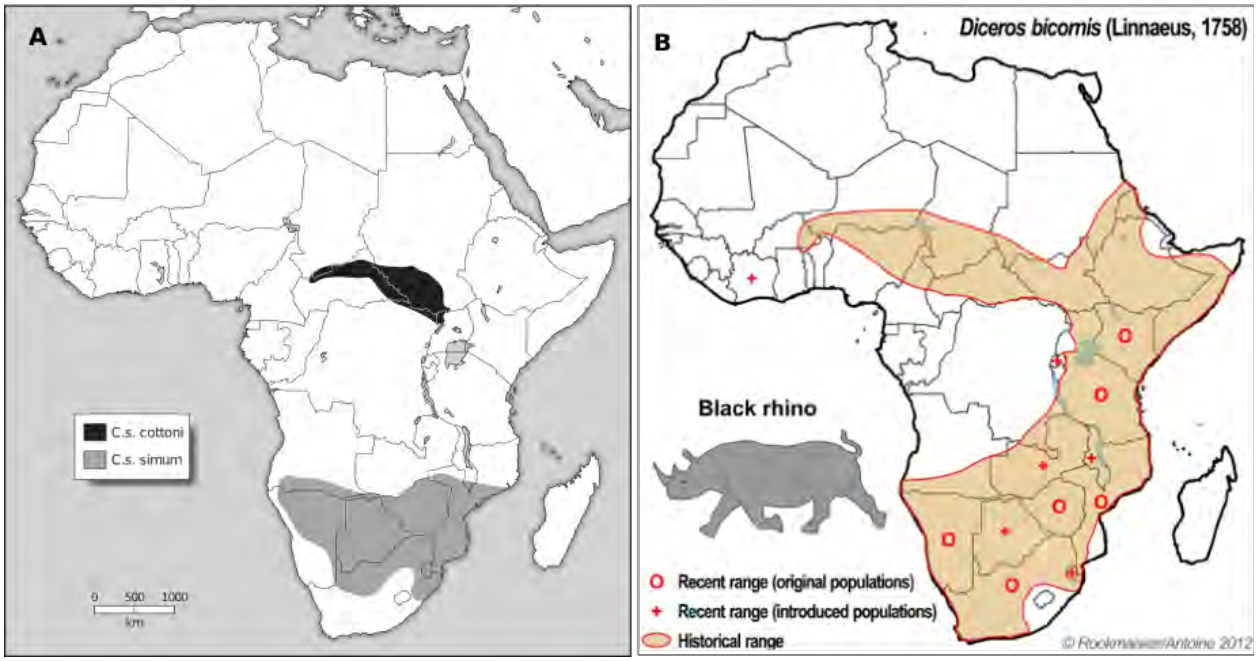


Figure 5. *Historical ranges of distribution of the African rhinos*. A) Historical distribution of the white rhino (*Ceratotherium simum*), for both subspecies, the northern white rhino (*C. s. cottoni*) and the southern white rhino (*C. s. simum*). Map adapted from [20]. B) Historical distribution of the black rhino (*Diceros bicornis*). Locations of current populations, either original or introduced, are highlighted in red. Map adapted from [29].

From a taxonomic perspective, controversy exists as to whether the NWR and the SWR should be considered two separate species. A morphological and genetic assessment prompted some authors to propose that the NWR be promoted to the category of species as the Nile rhino (*Ceratotherium cottoni*) [30]. However, more recent studies support the genetic closeness of these two populations, and that they should be maintained as subspecies [28,31].

The black rhino never ventured north of the present-day Sahara, but inhabited a vast, continuous range of distribution across Subsaharan Africa, avoiding only areas of dense forest [16]. Until the 20th century, the black rhino was abundantly found as far west as Cameroon and Nigeria, to the Horn of Africa on the East, and all the way south to South Africa, Namibia and Angola, with the exception of the tropical rainforest area of the Congo river basin [29] (Figure 5B).

The subspecies level taxonomy of the black rhino has been both a conundrum and a matter of debate for decades [32]. Although several taxonomic classifications have been proposed since the mid-20th century, a system of four conservation units or ecotypes has been broadly used in a taxonomic sense as well: *D.b. bicornis* (south-western black rhino), *D.b. minor* (south-eastern black rhino), *D.b. michaeli* (eastern black rhino) and *D.b. longipes* (western black rhino, already extinct) [32]. Recent genetic work [33], taking into consideration the historical diversity of the species, promoted the update of current management guidelines by highlighting some of the flaws of the four-ecotype classification. Nonetheless the intra-species diversity of the black rhino is far from being fully explored, as detailed in Chapter 3 of this dissertation.

3.2 No matter if you are black or white: recent demography of the African rhinos

The focus of this section lies in the recent history of the African rhinos, since the mid-19th century until today. The (only) positive aspect of the intense conflict between rhino populations and humans in the past two centuries is the remarkably good record we have of population trends in these wild species.

Human-rhino conflict started escalating upon settlement and expansion of European colonial presence in Africa throughout the 19th century. During the first half of the century, European settlers and explorers ventured into interior regions of Africa and reported the extraordinary abundance of game, including rhinos. The detailed accounts of these explorations show that these were also

hunting trips [34] . Beyond recreational hunting, European settlement imposed profound changes on the landscape that affected wildlife populations severely. A paradigmatic example are the anti-tsetse fly campaigns in Zululand (today's Eastern South Africa) in the early 1900s, whereby wild game was exterminated for the sake of clearing land to create pastures [35]. Probably a substantial proportion of rhinos was killed during these times, and already in the late 19th century, concerns were raised regarding their decreasing populations sizes and potential disappearance [34].

Habitat disturbances and hunting put a toll on the populations of both white and black rhinos, but the second half of the 20th century witnessed an even more dramatic turn of events. As civil unrest and armed conflict ravaged many parts of Africa, rhino poaching to supply the growing trade of rhino horn became a devastating threat to the remaining rhino populations, especially from the 1970s onwards [20,36]. The main centers of rhino horn demand have been in Yemen, for the production of traditional daggers called *jambiyas*, and East Asian countries such as China, Taiwan, Japan, Thailand and Vietnam, where uses of rhino horn are linked to traditional medicine, ornaments and the display of wealth and status [36,37].

Since the first decades of the 20th century conservation measures have attempted to provide space and protection to the remaining African rhinos, but their degree of success has not been even through time and space [20]. The victories of preservation efforts have often been countered by periodic upsurges of poaching, and despite the increases in census size of some rhino populations, the overall trend was negative all the way until the mid-1990s [38]. A short recovery occurred afterwards owing to decisive law enforcement and strong protection from some of the countries still hosting rhino populations [38], but since 2008 another wave of poaching has taken the lives of thousands of rhinos [39].

What those historical events have signified for the populations of black and white rhino during the past 200 years is summarised below. The SWR was met by European settlers early in the 19th century, and the resulting human-rhino conflict forced them through a bottleneck from several hundred thousands, to an estimated low of 200 individuals in one single surviving population in Kwa-Zulu (South Africa) at the turn of the 20th century [35]. Subsequent conservation efforts, starting as early the 1920s, boosted a remarkable recovery throughout the 20th century [35]. As a result, there are currently ~18,000 wild individuals, mostly in South Africa [40] (Figure 6B).

The NWR did not experience the consequences of colonial settlement until the late 19th century. The onset of their decline promoted conservation measures such as the creation of national parks in Sudan and DRC already in the 1930s. Until 1960, the NWR were more abundant, ~2,250 animals, than the recovering SWR. Unfortunately, from then onward subsequent poaching bursts drove their

populations to extinction [20,29,38] (Figure 6B). By 1984, 350 NWR remained in the wild, and by 1998, the figure had shrunk to 25 individuals in Garamba National Park (DRC) [20]. They were declared extinct in the wild in 2011 by the IUCN Red List [40]. Despite the disparity between the NWR and SWR, the white rhino species is currently ranked as Near Threatened in the IUCN Red List [40].

The black rhino was the most abundant of all five extant rhino species in the early 19th century, it is speculated that hundreds of thousands inhabited their vast and continuous range of distribution. Population extirpation became common already in the first decades of the 20th century due to hunting and habitat loss, but some 100,000 black rhinos still survived by 1960 [20]. The number dropped to 65,000 by 1970, and the years between 1970 and 1992 witnessed an astounding 96% decrease of black rhino population size; in 1995 a low of 2,354 individuals was recorded [41] (Figure 6A). Ever since, black rhino numbers have increased slightly in the five countries that host the biggest populations (Kenya, Tanzania, Zimbabwe, Namibia and South Africa), owing to large conservation efforts [38]. As of 2017, there were ~5,500 black rhinos left in Africa, but the species is still critically endangered [41].

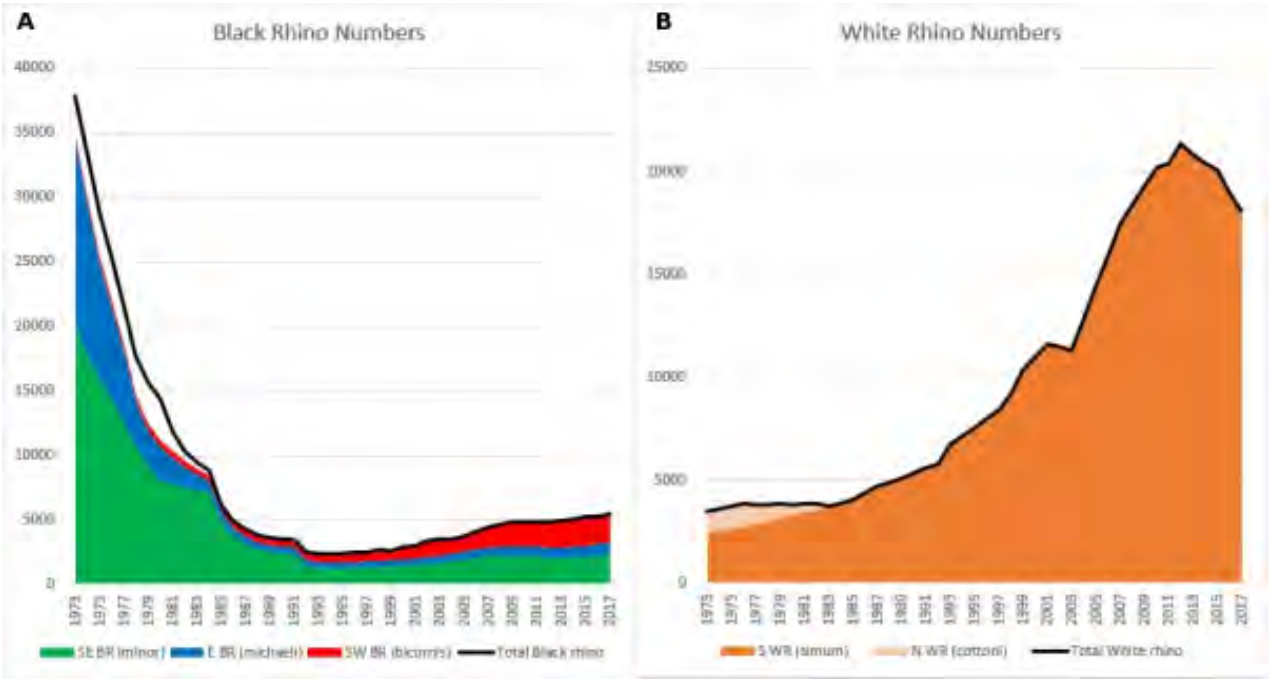


Figure 6. *Population size trends of the black and the white rhinos in Africa from 1973 to 2017*. A) Census size of the black rhino; colour shades refer to the three surviving ecotypes, the black line indicates total black rhino number. B) census size of the white rhino. From the supplementary material of the IUCN African Rhino Specialist Group 2020 report [41].

Poaching is a major threat to rhinos nowadays. Rhino horn is an illegal commodity that can reach prices higher than those of gold in the black market [37], and its trade is often orchestrated by transnational organised-crime syndicates [42]. In this scenario, political factors play key roles because they shape the contrasting socioeconomic circumstances at the ends of the trade: the increased wealth and demand of horn in the consumer societies, and the impoverishment and marginalisation of the communities living closest to the rhinos [42,43].

Ultimately, solving the human-rhino conflict will require multifaceted, decisive strategies, and conservation measures so far have proven vital, but not sufficient. In the prediction of further population reductions and necessary management of these species, it will be crucial to gather as much insight about white and black rhinos as possible to ensure their effective monitoring and preservation.

4. The dynamics of extinction from a population genetics perspective

Population declines and local extirpations are concerning because they lay the ground for species extinction. These phenomena are known to pose genetic consequences for populations, thus drawing a link between genetics and extinction. The background presented below revolves around this connection within the framework of population genetics theory. As it is a key concept in this discipline, when mentioning *population size* throughout this section, it will refer to the *effective population size* (N_e), as opposed to the *census size*. The effective size is defined as the size of an idealised Wright-Fisher population that experiences the same amount of genetic drift as the real population [44]. We will make the simplifying assumption that N_e correlates well with the census size, and thus imply that ‘large’ and ‘small’ N_e reflect a ‘large’ and ‘small’ census size respectively.

4.1 Size matters: the effects of ‘smallness’

When populations turn small, randomness takes control. As [45] phrases it “The dynamics of a small population are governed by the specific fortunes of each of its few individuals. In contrast, the dynamics of a large population are governed by the law of averages”. In the paragraphs below an

account of the consequences of ‘smallness’ on population genetic diversity is presented.

If we take a population of diploid individuals following a Wright-Fisher model (i.e. random mating, constant and finite population size and non-overlapping generations), and we do the exercise of populating the subsequent generations one after the other, we will be drawing at random as many gametes from the current generation as needed to match the set population size. Each individual is then the combination of two of these gametes drawn at random one after the other. If we then measured the heterozygosity in each generation as the proportion of individuals harbouring two non-identical alleles at a given locus, we would see that heterozygosity becomes lower and lower over time. In a Wright-Fisher population, heterozygosity decays at a rate of 1/N per generation, where N is the total number of sampled gametes; this diversity decay due to random sampling in a finite population is the genetic drift (see Chapter 2 of [46]). Following this, if the population size is large, the rate of heterozygosity decay will be low, while in a small population heterozygosity decays much faster from one generation to the next.

If we followed the trajectory of the frequency of an allele that is solely governed by genetic drift, we would watch it rise and drop randomly over time upon eventual fixation or disappearance. If the population size is large, these swings will be slight, and time to fixation or loss is likely to be long; in a small population, the allele frequency will bounce wildly, and become fixed or disappear much faster (see Chapter 2 of [47]). This process can be grasped intuitively: the loss of two individuals in a population of 15 rhinos will have a much bigger impact on the allele frequencies than the same loss in a population of 5,000. Ultimately, whether the allele we are observing is fixed or vanishes, the unavoidable outcome of this process is a diversity loss.

Indubitably, other forces play a part in determining genetic diversity in natural populations, e.g. mutation introduces variation in every generation, and differential selective advantages might promote the increase in frequency over time of the selected allele (see Chapter 4 of [46]. However, from modelling genetic drift alone two important points emerge: there is a stochastic component among the forces driving the fate of genetic diversity in a population, and the smaller the population, the stronger this component becomes.

Besides drift, another phenomenon is likely to occur in small populations: matings between related individuals, i.e. inbreeding [45]. Like genetic drift, inbreeding reduces heterozygosity in the subsequent generations, because individuals will show higher probabilities of identity by descent across loci, i.e. a higher inbreeding coefficient [45]. Homozygosity, the chance that two alleles at a given locus are identical, therefore increases, and with it the risk for inbreeding depression.

Inbreeding depression is a long studied question in evolutionary and population genetics, and refers to the reduction in fitness (at individual or population level) that emerges as a consequence of inbreeding [48]. The negative effects on fitness can be driven by a reduction of heterozygosity at sites where the heterozygous genotype is the most advantageous, but also, and more importantly, by the increased homozygosity exposing recessive deleterious variants [49].

Along with inbreeding, genetic drift contributes to the increase in frequency, and even fixation, of deleterious variants, thus inflating the genetic load [50]. Strong negative selection or *purging* might be efficient at removing strongly deleterious alleles in very small, inbred populations [49]. However, owing to the overriding effects of drift and inbreeding, small populations do not excel at eliminating by negative selection the mildly deleterious to deleterious variants whose accumulating detrimental effects might readily diminish the fitness [51].

Ultimately, ‘smallness’ is connected to an increased extinction risk, where the conflation of environmental and demographic stochasticity with genetic threats might trigger an extinction vortex [52] (Figure 7). A population that grows small starts suffering from the symptoms of ‘smallness’, which might lead to a reduction in fitness, which prompts the population to grow even smaller, thus entering a loop headed toward annihilation. Moreover, if variability is the substrate to adaptation, a homogeneous and homozygous population has bleak chances over evolutionary time-scales. Depauperate levels of diversity are likely to diminish the evolutionary potential of the population, thus rendering it less resilient to environmental change [53,54].

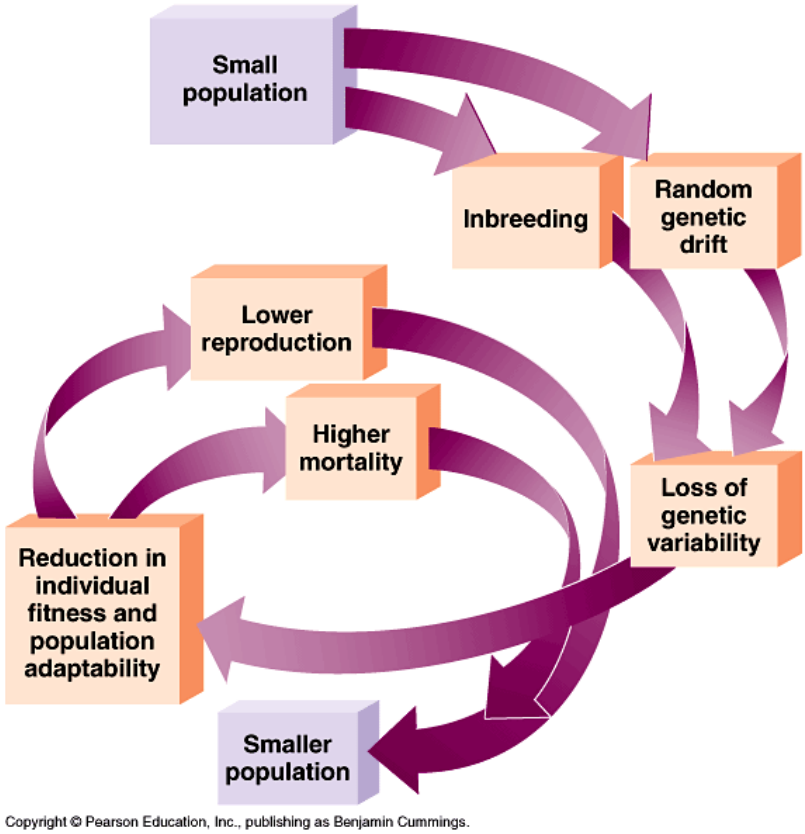


Figure 7. *Sequence of events in an extinction vortex*. Diagram from [55].

Empirical evidence abounds supporting the notion that populations do show genetic warning signs before becoming extinct [52,54,56], and this is concerning in a time where population declines are a pervasive phenomenon (see Section 1.1). Small population sizes have become the rule, a particularly obvious pattern among threatened species such as the African rhinos. As described in the next section, *how* populations arrive at 'smallness' seems to matter as well, but in any case it does not sound implausible that many of these remnant populations might be experiencing the genetic consequences of 'smallness'.

4.2 Time matters: the outcomes of population history

The amount and maintenance of genetic diversity across species and taxa is a mysterious aspect of biology that still lacks full explanation [57]. Seemingly, an assortment of factors ranging from life-history traits and ecological strategies, to natural selection and mutation might play a part in this conundrum [57], but it has been long established that there is a broad, positive correlation between genetic diversity and population size [53,57]. In a conservation framework, the recent demographic history is often of utmost relevance, because populations that are small today arrived at this stage recently. And unsurprisingly, *how* 'smallness' comes to be shapes its genetic consequences.

A population bottleneck is a stochastic *coup de force*, a demographic event where population size is reduced to a small fraction in a relatively short period of time. Upon a bottleneck, a general cutback to standing genetic diversity occurs, but low-frequency variants in the population will be particularly susceptible to disappearing [54]. Ultimately, the intensity and the duration of the bottleneck will determine the specific resulting distribution of genetic diversity [54].

As a derivation of this, the timing of the bottleneck, taking the present as the reference point, will also be responsible for the patterns of diversity we detect. Immediately after a bottleneck, the surviving population will feature a patchy version of the full landscape of diversity it displayed before the event; if a longer time period has elapsed since the bottleneck, the surviving population will additionally show symptoms of 'smallness'. An instance of this contrast is discussed in Chapter 2 of this thesis.

Populations harbour some proportion of deleterious genetic variation, and when population size dwindles, a fraction of that deleterious variation is likely to survive through the bottleneck and become frequent or fixed owing to drift and inbreeding, hence increasing the genetic load [58]. Conversely, recent research suggests that a population that has been invariably small for a long time, despite the high levels of inbreeding and low diversity, will display lower levels of genetic load than a large population, owing to efficient purging [59]. This highlights the relevance of population history in understanding its current genetic status.

The biodiversity trends of the Anthropocene fit the worrisome scenario where populations have been driven away from their equilibrium size via strong population bottlenecks. Probably the only positive aspect linked to these recent demographic events is that they are well recorded for many threatened species, such as the African rhinos. This information can be a key complement to any genetic/genomic study of the current status of threatened species.

4.3 Space matters: habitat loss and fragmentation

Habitat loss, and the often co-occurring habitat fragmentation, exert harsh effects on biodiversity and ecosystem functioning [60,61]. Fragmentation of a continuous habitat into smaller, discrete patches produces a landscape that resembles an island biogeography model, where more or less isolated 'islands' of suitable habitat are surrounded by an uninhabitable matrix [61]. Local extinction and immigration from the 'mainland' would be the forces controlling diversity, and smaller and more isolated islands would harbour less diversity [62]. A parallel with this model can

be drawn for intra-species population biology: in species with fragmented ranges of distribution, genetic diversity is negatively correlated with habitat patch size and isolation [63,64].

In a population genetics framework, a species distributed over a continuous range is likely to follow an isolation-by-distance pattern, i.e. the main driver of genetic differentiation will be the geographical distance between populations or subpopulations [65]. When landscape transformations convert continuous ranges into islands in an ‘ocean’ of disturbed territory [61], the new ecological scenario approaches a metapopulation model, where the dynamics between local extinction and migration rates among patches become key determinants of viability [66]. The emerging interplay between gene flow among patches and genetic drift within each patch will determine the new patterns of population differentiation. If local drift is stronger than gene flow, inter-population differentiation is expected to increase [67].

Current scenarios often involve strong human-driven landscape transformations that provoke habitat loss and fragmentation, and establish insurmountable barriers among the remaining populations, such as roads and railways. Besides the direct extirpation of individuals and local populations, with the consequent loss of genetic diversity, the long-term consequences of these disturbances have been proven to reshape patterns of genetic differentiation and to dwindle genetic diversity [68–70].

***

The phenomena described in 4.1, 4.2 and 4.3 often go hand in hand, so much so that the following story of an imaginary species could well fit the actual history of many threatened organisms today. Let us consider a wild, widespread species whose range of distribution starts being encroached and fragmented by human pressure on the landscape. During this habitat disturbance, some of the species’ populations are lost and others undergo bottlenecks. The immediate consequence of this is a massive loss of standing genetic variation at the species level.

The surviving populations are small and isolated, and they start suffering from strong drift and inbreeding, which erode heterozygosity and increase the genetic load. The negative consequences to the fitness of the population might take some time to arise, depending on the biology of the species, but if they do they might render the net population growth rate null or even negative, in which case extinction is to be expected.

It might be that conservation measures do manage to maintain population sizes afloat, but the impoverished adaptive potential of the populations might become apparent in the form of little resilience to environmental changes. Extinction risk, even after the immediate success of conservation measures, would still loom over the species, hence the importance of temporal

approaches to the study and monitoring of species under decline.

5. Studying biodiversity in an eroding world: looking back to move forward

In light of the background stories and the conceptual framework stated above, how do we study species or populations that are today little more than a phantom of what they used to be? One of the possible approaches, employed in this project, hinges on sampling historical specimens, and generating whole-genome sequence data from them.

5.1 Ancient DNA, genomics and their offspring

Ancient DNA (aDNA) samples are those that originate from specimens that were not preserved for the purposes of molecular biology work. As its name implies, a time component is one of the definitory traits: aDNA samples are often retrieved from long-dead biological material, such as archaeological or museum specimens. More than age, however, other features of aDNA determine how these samples must be handled to successfully retrieve genetic or genomic information from them: fragmentation, chemical damage and contamination [71].

Upon cell death, membranes are disrupted, cell compartmentalization breaks apart, and enzymatic activity fragments the DNA [72]. Moreover, hydrolytic breakage and oxidative reactions contribute to the degradation as repair mechanisms are not active anymore [71,73]. Later, the enzymatic action of bacteria and fungi that might sit on the sample enhance the degradation process [71]. The chemical and physical conditions surrounding the dead material are key determinants of the extent and specificities of the long-term degradation process. Crucially temperature, but also water and oxygen content have been identified as major players [73].

Chemical changes to the bases of the DNA molecules alter the original sequence [74]. Some of these miscoding lesions are particularly frequent, namely the deamination of adenine to hypoxanthine, and especially the deamination of cytosine to uracil [75], which produces a very characteristic signature in aDNA sequence data. Cytosines (C) pair with guanines (G), but the uracils (U) resulting from deamination pair with thymine (T), therefore, during the data generation workflow, CG base pairs become TA [71]. This pattern is reflected as an excess of C-to-T

transitions at 5’ ends of the DNA fragments (and the complementary G-to-A on the 3’ ends of the opposite strand) [76].

Besides all the post-mortem damage, an additional problem of aDNA samples is posed by the often high levels of contamination. The fragments of DNA that belong to the original organism, the endogenous content, are commonly blended in a complex mix of exogenous DNA from different sources, typically microbial [77]. No less important is to minimise the risk for contamination with modern DNA during sample processing [71].

In short, aDNA samples feature a collection of fragments of DNA of sizes typically between 100 and 500 bp [71], of which only some actually belong to the organism of interest, and these typically harbour characteristic miscoding lesions. These seemingly inconvenient features turned out to be unequivocal signatures of the authenticity of aDNA data. But obviously they pose challenges when working with aDNA samples, and they impose the need to follow strict guidelines that ensure the reliability and reproducibility of aDNA studies [78].

Since the first successful extraction and sequencing of aDNA, from a museum specimen of quagga (a South African equid), in 1984 [79] the field of aDNA has grown remarkably. Along the way, the associated challenges prompted detailed investigation of the chemical characteristics of aDNA, and boosted the development of a wealth of new methods to generate and analyse aDNA data [80].

Two technological revolutions have been pivotal in the advancement of aDNA studies: the polymerase chain reaction (PCR) in the late 1980’s, and the advent of high-throughput sequencing (HTS) in the early 2000’s [80]. With PCR, the small and scarce fragments of endogenous DNA could be replicated, initially only for targeted regions with specific primers, then for potentially all endogenous fragments. The HTS technologies that have dominated the market for the past decade are based on sequencing-by-synthesis of small fragments of DNA, an approach that proved very suitable for aDNA samples. Crucially, the decreasing costs of HTS have been a paramount factor in the expansion of aDNA-based research [80].

Discrete molecular markers have been extensively exploited in aDNA studies, notably mitochondrial markers, owing to the multicopy nature of this locus. However, the aforementioned decrease in price of HTS gave birth to the era of genomics; now molecular information from throughout entire genomes can be retrieved in large amounts. Ancient DNA studies quickly joined in, and tailored methodological solutions were developed to facilitate the switch to this ambitious path [81].

Today, the burgeoning field of *palaeogenomics* is generating genomic data from historical and archaeological material of diverse kinds [81], and has become a key contributor to numerous other disciplines, e.g. the study of human history, evolution and genetics [82,83], or the archaeological research of the domestication of crops [84] and animals [80]. Naturally, the study of wildlife is also riding this wave.

In the realm of empirical population genetics, the transition to genome-wide loci sampling has boosted the investigation of patterns of population structure, selection, gene flow, introgression, inbreeding, etc. at unprecedented levels of resolution and in a myriad of non-model species (e.g. [85–88]). The disciplines of population genomics and wildlife genomics have readily incorporated palaeogenomics in their scope too. Whole-genome data now exists for historical and ancient specimens of extinct [89,90] and extant species [91,92].

Given that museum collections constitute a particularly rich source of organisms spanning wide geographical and temporal ranges, it is not surprising that temporal, genomic studies relying on museum specimens are flourishing [91,93–95]. Like the ones presented in this dissertation, some lie within the scope of conservation genomics, under the premise that knowing the past will help us understand the present, to better prepare for the future.

5.2 Genomics through time for the study of the African rhinos

African rhinos are charismatic species of conservation concern. As detailed in Section 3, they are indeed severely threatened and considerable efforts are being invested to protect them. Given their recent demographic history, it is reasonable to fear that the processes described in Section 4 might be affecting the remaining rhinos. With this backdrop, the research presented in this thesis revolves around two specific themes: genomic erosion and historical biogeography in the African rhinos.

This project attempted to explore the past by means of collecting genomic data from historical rhino specimens housed in museums, to compare with or to complete the information provided by present-day genomic data. I have attempted to shed light onto whether white rhinos have suffered genomic erosion, understood as the process of loss of genomic diversity due to demographic history, and what was the population structure, connectivity and diversity of the black rhino in historical times.

By adopting a palaeogenomic approach, I aimed at increasing the resolution and scope of previous

genetic studies delving into the population structure and diversity of African rhinos with mitochondrial and microsatellite markers [31,33]. The ultimate goals would be to build upon the body of knowledge on the process of extinction, and to pour the insight gained into the realm of conservation.

Because rhinos have been hunted for recreational and scientific purposes for over two centuries, they are well represented in museums, and fairly good records of their recent demography exist. Additionally, reference genomes and prior genetic studies are available for these species, thus facilitating the cumulative continuation of this line of work via genome-wide, temporal, population-level sampling.

At the start of this project, a number of tissue samples from museum specimens of white and black rhino were handed to me. They came from museums all over the world, and they luckily represented most of the historical geographical range of the species. However I covered some gaps by ‘hunting’ a few additional museum specimens, and was lucky to be able to conduct some ‘field work’ myself at a museum.

The molecular work was performed separately for the historical and modern material in facilities dedicated to aDNA and modern DNA respectively, partly at the Swedish Museum for Natural History and mostly at the University of Copenhagen (Denmark). From both historical and modern samples, DNA was extracted, and from the extracts sequencing libraries were built and amplified using optimised protocols. Libraries of good quality were then outsourced for shotgun sequencing on a high-throughput platform that produced whole-genome data from each sample.

The voluminous raw data was processed bioinformatically, including its alignment against a reference genome. Once the data was aligned, then the fun began, as it was used for various statistical analyses within the framework of population genomics to characterise genomic diversity, population structure, gene flow, etc. This dataset has opened the doors to answering a myriad of questions. The results presented in the following chapters of this dissertation stem from the work carried out during the time frame of my PhD education.

References

1. Barnosky, A.D. (2008). Colloquium paper: Megafauna biomass tradeoff as a driver of Quaternary and future extinctions. Proc. Natl. Acad. Sci. U. S. A. *105 Suppl 1*, 11543–11548.

2. Doughty, C.E., Roman, J., Faurby, S., Wolf, A., Haque, A., Bakker, E.S., Malhi, Y., Dunning, J.B., Jr, and Svenning, J.-C. (2016). Global nutrient transport in a world of giants. Proc. Natl. Acad. Sci. U. S. A. *113*, 868–873.

3. Ripple, W.J., Wolf, C., Newsome, T.M., Betts, M.G., Ceballos, G., Courchamp, F., Hayward, M.W., Valkenburgh, B., Wallach, A.D., and Worm, B. (2019). Are we eating the world's megafauna to extinction? Conservation Letters *358*, e12627.

4. Malhi, Y., Doughty, C.E., Galetti, M., Smith, F.A., Svenning, J.-C., and Terborgh, J.W. (2016). Megafauna and ecosystem function from the Pleistocene to the Anthropocene. Proc. Natl. Acad. Sci. U. S. A. *113*, 838–846.

5. Smith, F.A., Elliott Smith, R.E., Lyons, S.K., and Payne, J.L. (2018). Body size downgrading of mammals over the late Quaternary. Science *360*, 310–313.

6. Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M., and Palmer, T.M. (2015). Accelerated modern human-induced species losses: Entering the sixth mass extinction. Sci Adv *1*, e1400253.

7. Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., *et al.* (2011). Has the Earth's sixth mass extinction already arrived? Nature *471*, 51–57.

8. Ceballos, G., Ehrlich, P.R., and Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proc. Natl. Acad. Sci. U. S. A. *114*, E6089–E6096.

9. Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N.J.B., and Collen, B. (2014). Defaunation in the Anthropocene. Science *345*, 401–406.

10. WWF (2018). Living Planet Report - 2018: Aiming Higher Grooten, M. and Almond, R.E.A., ed. (WWF) Available at: https://www.wwf.org.uk/sites/default/files/2018-10/LPR2018_Full%20Report.pdf.

11. Waters, C.N., Zalasiewicz, J., Summerhayes, C., Barnosky, A.D., Poirier, C., Gałuszka, A., Cearreta, A., Edgeworth, M., Ellis, E.C., Ellis, M., *et al.* (2016). The Anthropocene is functionally and stratigraphically distinct from the Holocene. Science *351*, aad2622.

12. Galetti, M., Moleón, M., Jordano, P., Pires, M.M., Guimarães, P.R., Jr., Pape, T., Nichols, E., Hansen, D., Olesen, J.M., Munk, M., *et al.* (2018). Ecological and evolutionary legacy of megafauna extinctions:

Anachronisms and megafauna interactions. Biol Rev *93*, 845–862.

13. Cerdeño, E. (1998). Diversity and evolutionary trends of the Family Rhinocerotidae (Perissodactyla). Palaeogeogr. Palaeoclimatol. Palaeoecol. *141*, 13–34.

14. The United States Geological Survey (2009). Geologic time scale. Available at: https://commons.wikimedia.org/wiki/File:Geologic_time_scale.jpg.

15. Fortelius, M., Heissig, K., Sarac, G., and Sen, S. (2003). Rhinocerotidae (Perissodactyla). In Geology and Paleontology of the Miocene Sinap Formation, Turkey, Mikael Fortelius, John Kappelman, Sevket Sen, and Raymond L. Bernor, ed. (Columbia University Press), pp. 282–307.

16. Geraads, D. (2010). Rhinocerotidae. In Cenozoic Mammals of Africa (University of California Press).

17. Dinerstein, E. (2011). Family text: Rhinocerotidae (Rhinoceroses). In Handbook of the mammals of the world, vol. 2: Hoofed animals., Wilson, Don E. & Mittermeier, Russell A., ed. (Lynx Edicions).

18. 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.

19. Steiner, C.C., and Ryder, O.A. (2011). Molecular phylogeny and evolution of the Perissodactyla. Zool. J. Linn. Soc. *163*, 1289–1303.

20. Emslie, R., and Brooks, M. (1999). African rhino: status survey and conservation action plan IUCN/SSC African Rhino Specialist Group, ed. (Gland, Switzerland and Cambridge, UK: IUCN).

21. Mohd Khan Bin, van Strien, N.J., and IUCN/SSC Asian Rhino Specialist Group (1997). Asian Rhinos: Status Survey and Conservation Action Plan (IUCN).

22. van Strien, N.J., Steinmetz, R., Manullang, B., Sectionov, Han, K.H., Isnan, W., Rookmaaker, K., Sumardja, E., Khan, M.K.M. & Ellis, S. (2008). IUCN Red List of Threatened Species: Javan Rhinoceros. IUCN Red List of Threatened Species. Available at: https://www.iucnredlist.org/species/19495/8925965

23. Talukdar, B., and Susie Ellis (AsRSG, International Rhino Foundation) (2018). IUCN Red List of Threatened Species: Greater One-horned Rhino. IUCN Red List of Threatened Species. Available at: https://www.iucnredlist.org/species/19496/18494149

24. van Strien, N.J., Manullang, B., Sectionov, Isnan, W., Khan, M.K.M, Sumardja, E., Ellis, S., Han, K.H., Boeadi, Payne, J. & Bradley Martin, E. (2008). IUCN Red List of Threatened Species: Sumatran Rhinoceros. IUCN Red List of Threatened Species. Available at: https://www.iucnredlist.org/species/6553/12787457

25. Geraads, D. (2020). Perissodactyla (Rhinocerotidae and Equidae) from Kanapoi. J. Hum. Evol. *140*, 102373.

26. Moodley, Y., Westbury, M.V., Russo, I.M., Gopalakrishnan, S., Rakotoarivelo, A., Olsen, R., Prost, S., Tunstall, T., Ryder, O.A., Dalén, L., Bruford, M.W. Interspecific gene flow and the evolution of specialisation in black and white rhinoceros. Mol. Biol. Evol.

27. Waldram, M.S., Bond, W.J., and Stock, W.D. (2008). Ecological Engineering by a Mega-Grazer: White Rhino Impacts on a South African Savanna. Ecosystems *11*, 101–112.

28. 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.

29. Rookmaaker, K., and Antoine, P.-O. (2012). New maps representing the historical and recent distribution of the African species of rhinoceros: Diceros bicornis, Ceratotherium simum and Ceratotherium cottoni. Pachyderm *52*, 91–96.

30. Groves, C.P., Fernando, P., and Robovský, J. (2010). The sixth rhino: a taxonomic re-assessment of the critically endangered northern white rhinoceros. PLoS One *5*, e9703.

31. Moodley, Y., Russo, I.-R.M., Robovský, J., Dalton, D.L., Kotzé, A., Smith, S., Stejskal, J., Ryder, O.A., Hermes, R., Walzer, C., *et al.* (2018). Contrasting evolutionary history, anthropogenic declines and genetic contact in the northern and southern white rhinoceros (Ceratotherium simum). Proc. Biol. Sci. *285*, 20181567.

32. Rookmaaker, K. (2011). A review of black rhino systematics proposed in Ungulate Taxonomy by Groves and Grubb (2011) and its implications for rhino conservation. Pachyderm *50*, 72–76.

33. Moodley, Y., Russo, I.-R.M., Dalton, D.L., Kotzé, A., Muya, S., Haubensak, P., Bálint, B., Munimanda, G.K., Deimel, C., Setzer, A., *et al.* (2017). Extinctions, genetic erosion and conservation options for the black rhinoceros (Diceros bicornis). Sci. Rep. *7*, 41417.

34. Rookmaaker, L.C. (2005). Review of the European perception of the African rhinoceros. J. Zool. *265*, 365–376.

35. Rookmaaker, K. (2000). The alleged population reduction of the southern white rhinoceros (Ceratotherium simum simum) and the successful recovery. Säugetierkundliche Mitteilungen *45*, 55–70.

36. Amin, R., Thomas, K., Emslie, R.H., Foose, T.J., and Van Strien, N. (2006). An overview of the conservation status of and threats to rhinoceros species in the wild. Int. Zoo Yb. *40*, 96–117.

37. Ayling, J. (2013). What Sustains Wildlife Crime? Rhino Horn Trading and the Resilience of Criminal Networks. J. Int. Wildl. Law Policy *16*, 57–80.

38. Emslie, R. (2008). Rhino population sizes and trends. Pachyderm *44*, 88–95.

39. Knight, M. (2017). African Rhino Specialist Group report / Rapport du Groupe de Spécialistes du Rhinocéros d'Afrique. Pachyderm *58*, 17–35.

40. Richard Emslie (IUCN SSC African Rhino Specialist Group) (2020). IUCN Red List of Threatened Species: White Rhino (IUCN) Available at: https://www.iucnredlist.org/species/4185/45813880.

41. Richard Emslie (IUCN SSC African Rhino Specialist Group) (2020). IUCN Red List of Threatened Species: Black Rhino (IUCN) Available at: https://www.iucnredlist.org/en.

42. Haas, T.C., and Ferreira, S.M. (2016). Combating Rhino Horn Trafficking: The Need to Disrupt Criminal Networks. PLoS One *11*, e0167040.

43. Hübschle, A.M. (2017). The social economy of rhino poaching: Of economic freedom fighters, professional hunters and marginalized local people. Curr. Sociol. *65*, 427–447.

44. Charlesworth, B. (2009). Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. *10*, 195–205.

45. Caughley, G. (1994). Directions in Conservation Biology. J. Anim. Ecol. *63*, 215–244.

46. Wakeley, J. (2009). Coalescent theory: an introduction R. Camp, ed. (Greenwood Village, Colorado: Roberts and Company Publishers).

47. Slatkin, M., and Nielsen, R. (2013). An Introduction to Population Genetics: Theory and Applications (Sinauer Associates, Oxford University Press).

48. Frankham, R. (2005). Genetics and extinction. Biol. Conserv. *126*, 131–140.

49. Charlesworth, B., and Charlesworth, D. (1999). The genetic basis of inbreeding depression. Genet. Res. *74*, 329–340.

50. Lohr, J.N., and Haag, C.R. (2015). Genetic load, inbreeding depression, and hybrid vigor covary with population size: An empirical evaluation of theoretical predictions. Evolution *69*, 3109–3122.

51. Reed, D.H., Lowe, E.H., Briscoe, D.A., and Frankham, R. (2003). Inbreeding and extinction: Effects of rate of inbreeding. Conserv. Genet. *4*, 405–410.

52. Blomqvist, D., Pauliny, A., Larsson, M., and Flodin, L.-A. (2010). Trapped in the extinction vortex? Strong genetic effects in a declining vertebrate population. BMC Evol. Biol. *10*, 33.

53. Frankham, R. (1996). Relationship of Genetic Variation to Population Size in Wildlife. Conserv. Biol. *10*, 1500–1508.

54. Phillip R. England, Graham H.R. Osler, Lynn M. Woodworth, Margaret E. Montgomery, David A. Briscoe, Richard Frankham (2003). Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. Conservation, 595–604.

55. Campbell, N.A., Reece, J.B., Urry, L.A., Cain, M.L., Wasserman, S.A., Minorsky, P.V., and Jackson, R.B. (2008). Biology, 8th Edition (Pearson Benjamin Cummings).

56. Spielman, D., Brook, B.W., and Frankham, R. (2004). Most species are not driven to extinction before genetic factors impact them. Proc. Natl. Acad. Sci. U. S. A. *101*, 15261–15264.

57. Leffler, E.M., Bullaughey, K., Matute, D.R., Meyer, W.K., Ségurel, L., Venkat, A., Andolfatto, P., and Przeworski, M. (2012). Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? PLoS Biol. *10*, e1001388.

58. van Oosterhout, C. (2020). Mutation load is the spectre of species conservation. Nat Ecol Evol. Available at: http://dx.doi.org/10.1038/s41559-020-1204-8.

59. van der Valk, T., de Manuel, M., Marques-Bonet, T., and Guschanski, K. Estimates of genetic load in small populations suggest extensive purging of deleterious alleles. bioRxiv. Available at: http://dx.doi.org/10.1101/696831.

60. Fahrig, L. (2003). Effects of Habitat Fragmentation on Biodiversity. Annu. Rev. Ecol. Evol. Syst. *34*, 487–515.

61. Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D., *et al.* (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. Sci Adv *1*, e1500052.

62. MacArthur, R.H., and Wilson, E.O. (2001). The Theory of Island Biogeography (Princeton University Press).

63. Frankham, R. (1997). Do island populations have less genetic variation than mainland populations? Heredity *78 (Pt 3)*, 311–327.

64. Costanzi, J.-M., and Steifetten, Ø. (2019). Island biogeography theory explains the genetic diversity of a fragmented rock ptarmigan (Lagopus muta) population. Ecol. Evol. *9*, 3837–3849.

65. Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. Evolution *47*, 264–279.

66. Sih, A., Jonsson, B.G., and Luikart, G. (2000). Habitat loss: ecological, evolutionary and genetic consequences. Trends Ecol. Evol. *15*, 132–134.

67. DiLeo, M.F., and Wagner, H.H. (2016). A Landscape Ecologist's Agenda for Landscape Genetics. Current Landscape Ecology Reports *1*, 115–126.

68. Keller, I., and Largiadèr, C.R. (2003). Recent habitat fragmentation caused by major roads leads to reduction of gene flow and loss of genetic variability in ground beetles. Proc. Biol. Sci. *270*, 417–423.

69. Haag, T., Santos, A.S., Sana, D.A., Morato, R.G., Cullen, L., Jr, Crawshaw, P.G., Jr, De Angelo, C., Di Bitetti, M.S., Salzano, F.M., and Eizirik, E. (2010). The effect of habitat fragmentation on the genetic structure of a top predator: loss of diversity and high differentiation among remnant populations of Atlantic Forest jaguars (Panthera onca). Mol. Ecol. *19*, 4906–4921.

70. Dixon, J.D., Oli, M.K., Wooten, M.C., Eason, T.H., McCown, J.W., and Cunningham, M.W. (2007). Genetic consequences of habitat fragmentation and loss: the case of the Florida black bear (Ursus americanus floridanus). Conserv. Genet. *8*, 455–464.

71. Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., and Hofreiter, M. (2004). Genetic analyses from ancient DNA. Annu. Rev. Genet. *38*, 645–679.

72. Pääbo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc. Natl. Acad. Sci. U. S. A. *86*, 1939–1943.

73. Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709–715.

74. Axelsson, E., Willerslev, E., Gilbert, M.T.P., and Nielsen, R. (2008). The effect of ancient DNA damage on inferences of demographic histories. Mol. Biol. Evol. *25*, 2181–2187.

75. Gilbert, M.T.P., Hansen, A.J., Willerslev, E., Rudbeck, L., Barnes, I., Lynnerup, N., and Cooper, A. (2003). Characterization of genetic miscoding lesions caused by postmortem damage. Am. J. Hum. Genet. *72*, 48–61.

76. Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., *et al.* (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. U. S. A. *104*, 14616–14621.

77. Damgaard, P.B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., and Allentoft, M.E. (2015). Improving access to endogenous DNA in ancient bones and teeth. Sci. Rep. *5*, 11184.

78. Gilbert, M.T.P., and Willerslev, E. (2006). Authenticity in ancient DNA studies. Med. Secoli *18*, 701–723.

79. Higuchi, R., Bowman, B., Freiberger, M., Ryder, O.A., and Wilson, A.C. (1984). DNA sequences from the quagga, an extinct member of the horse family. Nature *312*, 282–284.

80. McHugo, G.P., Dover, M.J., and MacHugh, D.E. (2019). Unlocking the origins and biology of domestic animals using ancient DNA and paleogenomics. BMC Biol. *17*, 98.

81. Der Sarkissian, C., Allentoft, M.E., Ávila-Arcos, M.C., Barnett, R., Campos, P.F., Cappellini, E., Ermini, L., Fernández, R., da Fonseca, R., Ginolhac, A., *et al.* (2015). Ancient genomics. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20130387.

82. Slatkin, M., and Racimo, F. (2016). Ancient DNA and human history. Proc. Natl. Acad. Sci. U. S. A. *113*, 6380–6387.

83. Llamas, B., Willerslev, E., and Orlando, L. (2017). Human evolution: a tale from ancient genomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*. Available at: http://dx.doi.org/10.1098/rstb.2015.0484.

84. Kantar, M.B., Nashoba, A.R., Anderson, J.E., Blackman, B.K., and Rieseberg, L.H. (2017). The Genetics and Genomics of Plant Domestication. Bioscience *67*, 971–982.

85. Sinding, M.-H.S., Gopalakrishan, S., Vieira, F.G., Samaniego Castruita, J.A., Raundrup, K., Heide Jørgensen, M.P., Meldgaard, M., Petersen, B., Sicheritz-Ponten, T., Mikkelsen, J.B., *et al.* (2018). Population genomics of grey wolves and wolf-like canids in North America. PLoS Genet. *14*, e1007745.

86. de Manuel, M., Kuhlwilm, M., Frandsen, P., Sousa, V.C., Desai, T., Prado-Martinez, J., Hernandez-Rodriguez, J., Dupanloup, I., Lao, O., Hallast, P., *et al.* (2016). Chimpanzee genomic diversity reveals ancient admixture with bonobos. Science *354*, 477–481.

87. Westbury, M.V., Hartmann, S., Barlow, A., Wiesel, I., Leo, V., Welch, R., Parker, D.M., Sicks, F., Ludwig, A., Dalén, L., *et al.* (2018). Extended and Continuous Decline in Effective Population Size Results in Low Genomic Diversity in the World's Rarest Hyena Species, the Brown Hyena. Mol. Biol. Evol. *35*, 1225–1237.

88. Ribeiro, Â.M., Canchaya, C.A., Penaloza, F., Galindo, J., and da Fonseca, R.R. (2019). Population genomic footprints of environmental pollution pressure in natural populations of the Mediterranean mussel. Mar. Genomics *45*, 11–15.

89. 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.

90. Murray, G.G.R., Soares, A.E.R., Novak, B.J., Schaefer, N.K., Cahill, J.A., Baker, A.J., Demboski, J.R., Doll, A., Da Fonseca, R.R., Fulton, T.L., *et al.* (2017). Natural selection shaped the rise and fall of passenger pigeon genomic diversity. Science *358*, 951–954.

91. van der Valk, T., Díez-Del-Molino, D., Marques-Bonet, T., Guschanski, K., and Dalén, L. (2019). Historical Genomes Reveal the Genomic Consequences of Recent Population Decline in Eastern Gorillas. Curr. Biol. *29*, 165–170.e6.

92. de Manuel, M., Barnett, R., Sandoval-Velasco, M., Yamaguchi, N., Garrett Vieira, F., Zepeda Mendoza, M.L., Liu, S., Martin, M.D., Sinding, M.-H.S., Mak, S.S.T., *et al.* (2020). The evolutionary history of extinct and living lions. Proc. Natl. Acad. Sci. U. S. A. Available at: http://dx.doi.org/10.1073/pnas.1919423117.

93. Bi, K., Linderoth, T., Vanderpool, D., Good, J.M., Nielsen, R., and Moritz, C. (2013). Unlocking the vault: next-generation museum population genomics. Mol. Ecol. *22*, 6018–6032.

94. Feng, S., Fang, Q., Barnett, R., Li, C., Han, S., Kuhlwilm, M., Zhou, L., Pan, H., Deng, Y., Chen, G., *et al.* (2019). The Genomic Footprints of the Fall and Recovery of the Crested Ibis. Curr. Biol. *29*, 340–349.

95. Dussex, N., von Seth, J., Knapp, M., Kardailsky, O., Robertson, B.C., and Dalén, L. (2019). Complete genomes of two extinct New Zealand passerines show responses to climate fluctuations but no evidence for genomic erosion prior to extinction. Biol. Lett. *15*, 20190491.

Aims, contributions and structure

In brief, the aims of this thesis have been: a) establishing an approach for the research of threatened species by means of temporal sampling in a population genomics framework, b) generating whole-genome sequence data of historical and modern white and black rhinoceroses, and c) leveraging the power of this dataset to detect genomic erosion, and to investigate biogeographical patterns at an intra-species level in the African rhinos. With these in mind, the structure of the present thesis is the following:

Chapter 1. *Quantifying Temporal Genomic Erosion in Endangered Species*. Opinion paper, published in Trends in Ecology and Evolution in December 2017. This first chapter lays the conceptual basis for the study of genomic erosion based on using data from pre-decline museum specimens as a baseline. I contributed to this collaborative work with background literature review and writing from an early-draft stage. I also assisted with corrections and proofreading during the revision process.

Chapter 2. *Historical population declines prompted significant genomic erosion in the northern and southern white rhinoceros (*Ceratotherium simum*).* bioRxiv pre-print. This study's main focus is the quantification of genomic erosion in the white rhino.

Chapter 3. *Historical sampling portrays a vanishing beast: population structure, phylogeography and genomic diversity in the black rhinoceros (*Diceros bicornis*).* Manuscript in preparation. This chapter revolves around the range-wide biogeography of the black rhino in historical times.

The bulk of my research work is presented in Chapter 2 and Chapter 3. In both studies, I was the lead author, which involved handling the biological samples to generate most of the genomic data, processing bioinformatically the raw sequence data, and then analysing it computationally with guidance from mainly Shyam Gopalakrishnan, Jazmín Ramos-Madrigal, Michael V. Westbury, Marc de Manuel and Ashot Margaryan. I then made the figures and wrote the manuscripts.

Conclusions and perspectives. This section includes some reflections nurtured throughout the course of this project, particularly placing this work the context of conservation biology.

Appendix. I list here the research outputs of a series of projects in which I was involved to some extent during the years of my PhD education. These pieces of work are excluded from evaluation by the examining committee of this thesis.

Supplementary material of *Historical population declines prompted significant genomic erosion in the northern and southern white rhinoceros (*Ceratotherium simum*)*

Fátima Sánchez-Barreiro, Shyam Gopalakrishnan, Jazmín Ramos-Madrigal, Michael V. Westbury, Marc de Manuel, Ashot Margaryan, Marta M. Ciucani, Filipe G. Vieira, Yannis Patramanis, Daniela C. Kalthoff, Zena Timmons, Thomas Sicheritz-Pontén, Love Dalén, Oliver A. Ryder, Guojie Zhang, Tomás Marquès-Bonet, Yoshan Moodley, M. Thomas P. Gilbert

Mapping statistics and ancient DNA damage assessment

Out of a total of 55 white rhinoceros re-sequenced genomes, two were discarded from further analyses because they showed an average depth of coverage <1x. An additional sample was discarded since it was identical to another one sequenced at higher depth of coverage (see *Relatedness test* and Figure S3). Mapping statistics per sample, for the remaining 52, can be found in Table S1. Sample identifiers indicate the code of the country of origin, the date (of collection for pre-bottleneck samples, and of birth for post-bottleneck samples), as well as a counter to distinguish samples of same age and country of origin. Summary statistics of the sequencing data per group are reported in Table S2.

Table S1. *Metadata, mapping statistics and individual diversity measures per sample*. See spreadsheet *WR_TableS1_v3* (available at https://github.com/fasaba/WR_supplementary). The variable *retained_reads* refers to the number of reads that were not discarded by PALEOMIX based on size filtering (>25 bp), and could then be aligned to the reference assembly; the remaining mapping statistics were calculated by PALEOMIX based on this number. The column *depth_of_coverage* refers to the average depth of coverage across the entire reference assembly as calculated by PALEOMIX. The endogenous content of each sample is indicated by *mapped_reads_unique_fraction*, which is the fraction of raw reads mapped minus the fraction of those reads that were clonal.

		Retained reads		Mapped fraction raw		Clonality		Mapped fraction unique		Depth of coverage		5'CtoT damage	
	n	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
NWR1	16	7.73E+08	3.22E+07	0.7029	0.2190	0.3132	0.1347	0.4921	0.1979	12.10	5.42	0.0259	0.0118
NWR2	9	1.80E+08	1.34E+07	0.7427	0.0567	0.0039	0.0011	0.7398	0.0562	9.25	1.05	0.0008	0.0001
SWR1	9	8.04E+08	2.56E+07	0.6806	0.2084	0.2939	0.0896	0.4769	0.1629	10.88	4.66	0.0284	0.0146
SWR2	18	4.41E+08	1.87E+08	0.7965	0.0925	0.1430	0.0911	0.6847	0.1103	14.92	5.15	0.0022	0.0015

Table S2. *Summary statistics of sequence data per group*.

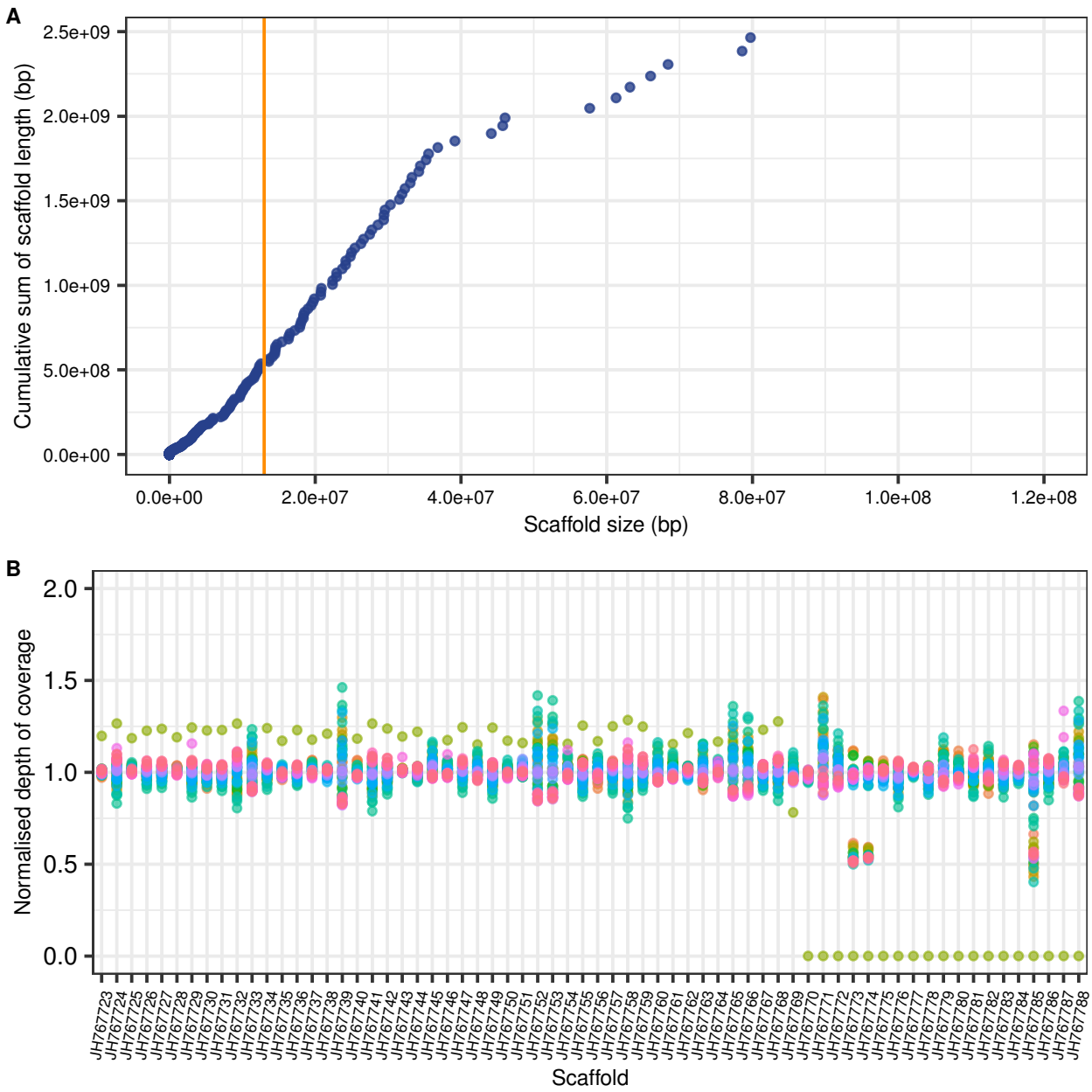


Figure S1. *Choice of scaffolds for variant site finding*. (A) Contribution of each scaffold in the reference assembly to the total length of it. The vertical orange line is placed at a scaffold size of 13 Mbp; only scaffolds above this size were considered for variant site finding. (B) Normalised depth of coverage for each scaffold >13 Mbp for 52 samples. Of the 66 scaffolds on the horizontal axis, three (JH767773, JH767774, JH767785) show a normalised depth of 0.5 for male samples, and were therefore excluded from further analyses.

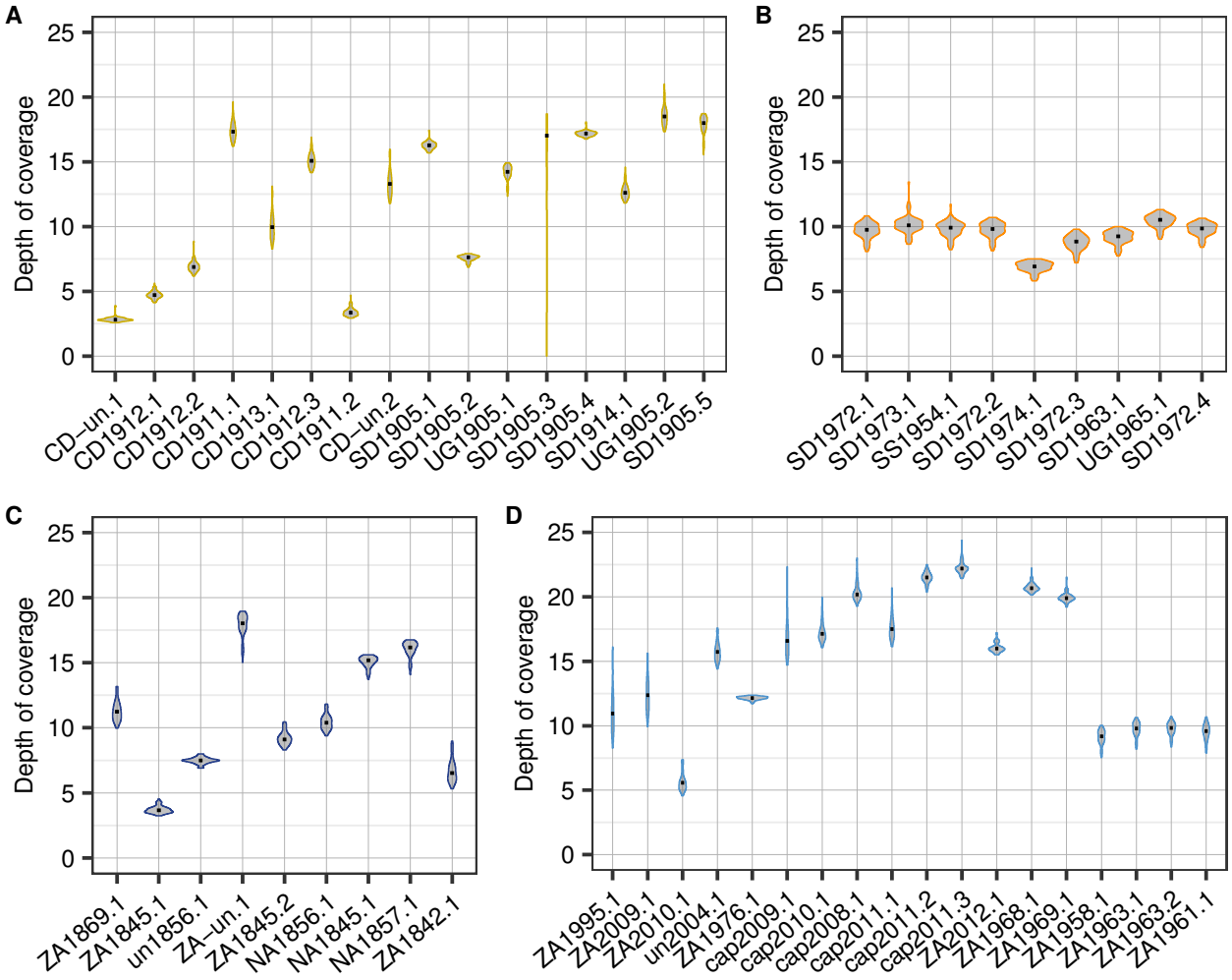


Figure S2. *Distribution of depth of coverage per sample.* For each sample, the distribution of depth of coverage across the 63 chosen scaffolds is shown; samples are sorted into: NWR1 (A), NWR2 (B), SWR1 (C) and SWR1 (D).

Cross-contamination tests

To verify that no major contamination issues among samples could bias further analyses, we extracted the fraction of reads that mapped against the mitochondrial scaffold (chrM) from each bam file, and generated a fasta file from those reads with Analysis of Next Generation Sequencing Data (ANGSD) v 0.921 [41] option -doFasta 1. The frequency of sites bearing 1, 2, 3, or 4 alleles was computed. Since the mitochondrial chromosome is haploid, for a given sample and site, a frequency of 1 is expected for a given allele; additional alleles occurring might be due to sequencing error, aDNA damage or contamination. Diagnostically, contamination is associated with frequencies above 20% of alternative alleles in these haploid sites; lower frequencies are likely due

to errors. We observed that, most samples contained sites with more than one allele in their mtDNA, but at frequencies below 2%.

Relatedness test

We ran an analysis of relatedness based on a panel of genotype likelihoods with ngsRelate v2 [54], following the approach described in [55]. Two samples appeared as identical (SD1905.5 and SD1905.6) (Figure S3A-B), therefore SD1905.6 was discarded from further analyses. In a separate analysis per subspecies, we found that two pairs of NWR and one pair of SWR samples showed a relatedness signal (Figure S3C-F), so for analyses of structure (i.e. PCA, UMAP and admixture), the sample of lowest depth of coverage from each pair was excluded (CD-un.1, un1856.1, ZA1842.1).

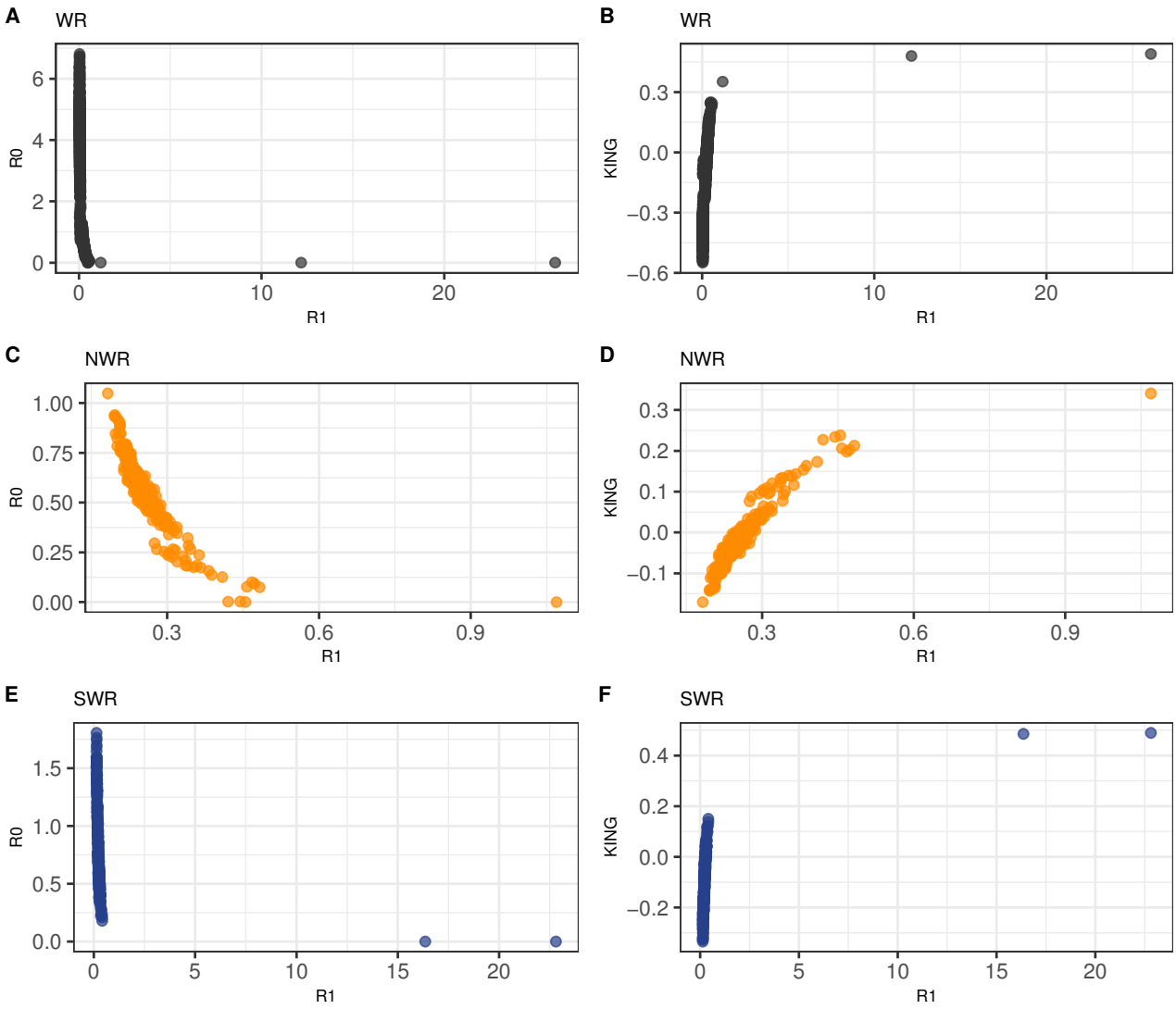


Figure S3. *Relatedness assessment of white rhinoceros samples with ngsRelate*. Analysis of 53 re-sequenced white rhinoceroses (A, B), for 25 NWR (C, D), and for 27 SWR (E, F). In all cases, analyses were based on a panel of genotype likelihoods for transversion sites. For each sample set, each point represents a pairwise

combination of samples, and its combination of coefficients is a proxy for the degree of relatedness [55]. In A) and B), the furthest outlier pair corresponds to SD1905.5 - SD1905.6, seemingly identical samples. In C) and D) the related pair corresponds to CD-un.1 - CD-un.2; in E) and F) related pairs are un1856.1 - ZA-un.1 and ZA1845.2 - ZA1842.1.

Principal component analysis (PCA)

Here are the visualizations of PCs one to three from a PCA analysis of 49 unrelated samples in our dataset. The first principal component (PC1) shows a clear separation between NWR and SWR (Figure S4); the PC2 uncovers population substructure within both subspecies. Post-bottleneck NWR fell within the Sudan-Uganda historical diversity only, but our pre-bottleneck NWR expanded beyond this gradient to represent a wide range of DRC diversity as well. In the SWR, PC2 revealed a split between pre- and post-bottleneck (Figure S4A), where pre-bottleneck samples were more widely scattered, while post-bottleneck SWR clustered tightly together. The PC3 resolved the distribution of SWR, as we observed substructure due to geography in the pre-bottleneck samples, and a comparatively small differentiation among post-bottleneck SWR (Figure S4B).

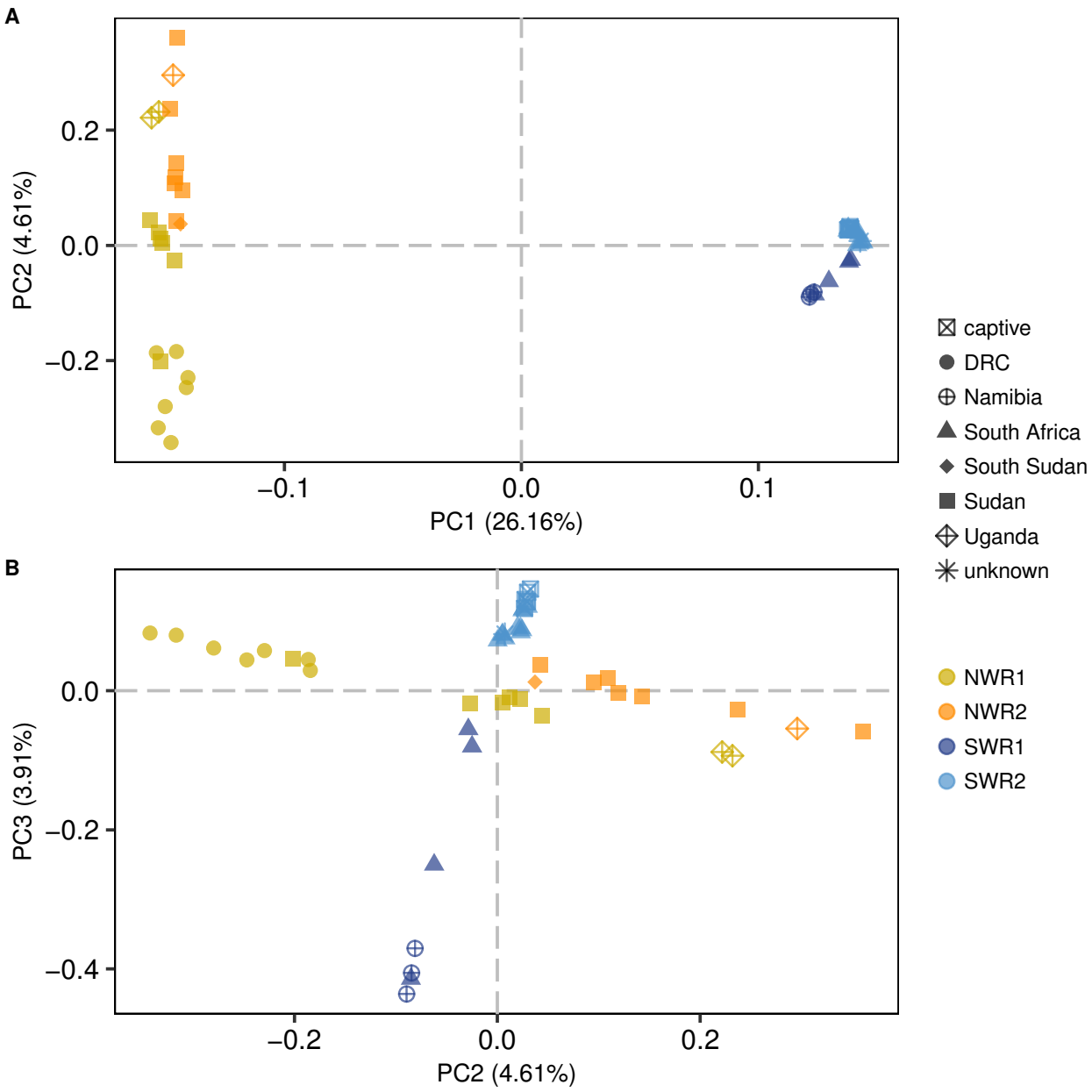


Figure S4. *PCA analysis of 49 unrelated white rhinoceroses*. Based on genotype likelihoods for transversion sites only. A) PC1 against PC2; B) PC2 against PC3.

Heterozygosity correction

To account for differences in depth of coverage across samples, the estimated genome-wide heterozygosity and exon heterozygosity were corrected based on how heterozygosity estimates decay with decreasing values of depth of coverage (see *Heterozygosity correction* in Methods for more details). The curves from which the correcting equations were drawn are depicted in Figure S5 for genome-wide heterozygosity estimates, and in Figure S6 for exon heterozygosity estimates.

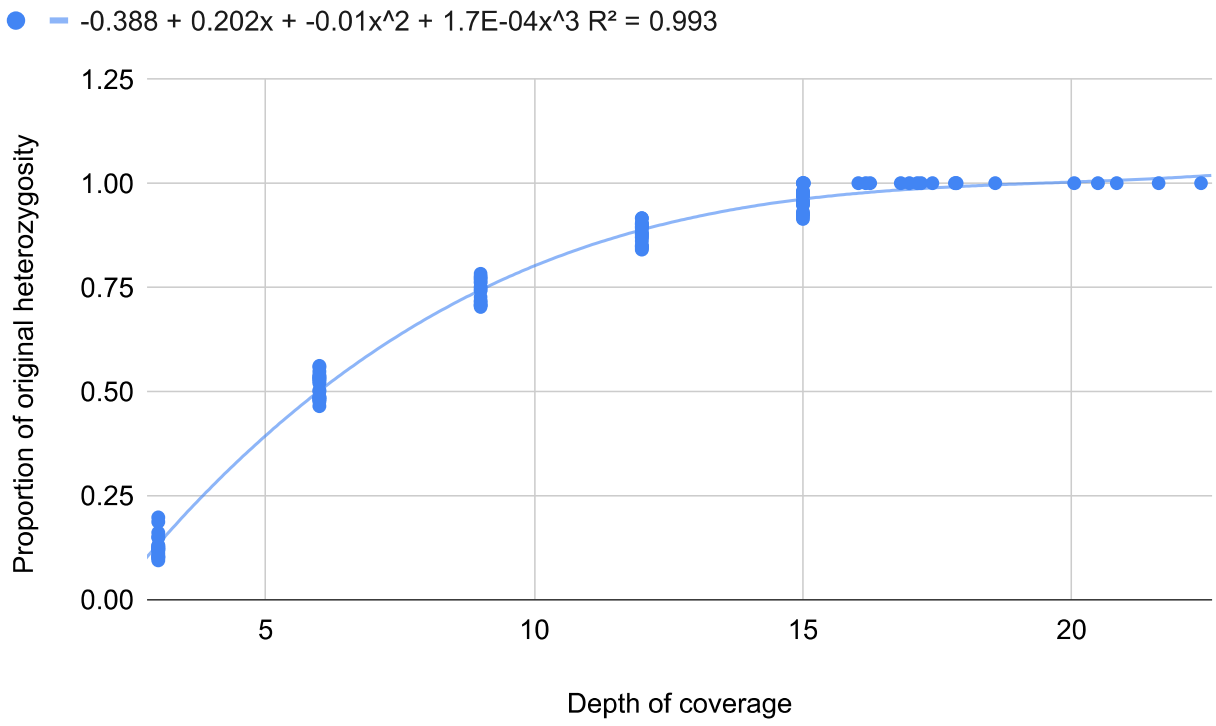


Figure S5. *Relation between the depth of coverage and the normalised genome-wide heterozygosity.* Estimates of genome-wide heterozygosity were calculated based on the individual SFS for decreasing values of depth for samples of mean depth of coverage >15x (n = 19). The trend line was generated in excel by choosing the polynomial equation of lowest degree and highest R^2 that best describes the distribution.

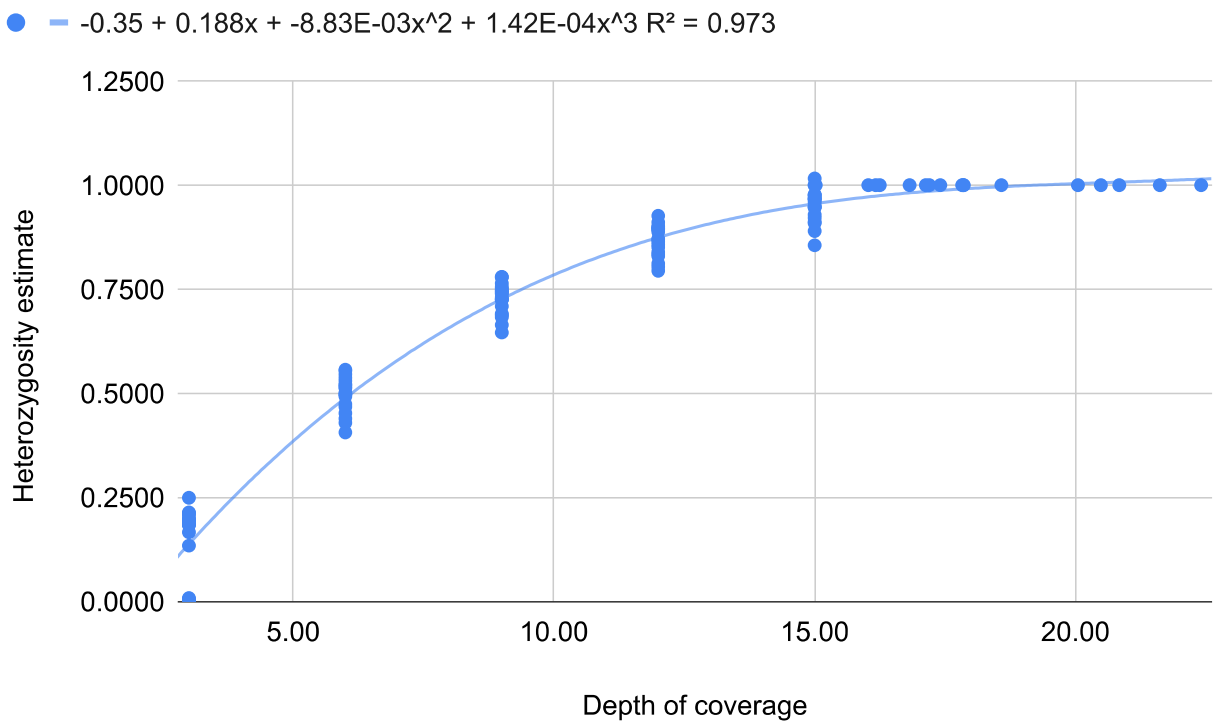


Figure S6. *Relation between the depth of coverage and the normalised exon heterozygosity.* Estimates of exon heterozygosity were calculated based on the individual SFS for decreasing values of depth for samples of mean depth of coverage >15x (n = 19). The trend line was generated in excel by choosing the polynomial equation of lowest degree and highest R^2 that best describes the distribution.

Local estimates of heterozygosity and identification of RoH

To calculate an inbreeding coefficient per sample, we visualised the distribution of windows falling along the range of local estimates of heterozygosity per sample (Figure S7). In NWR2 and SWR2, these distributions are skewed toward lower values of heterozygosity. There are three outliers (CD-un.1, CD1911.2 and ZA1845.1) that show a disproportionate number of windows with heterozygosity values very close to zero, probably because they have a substantial amount of missing data given that their mean depth of coverage is <4x. Pre-bottleneck groups, and particularly SWR1, show the longest lengths of RoH detected.

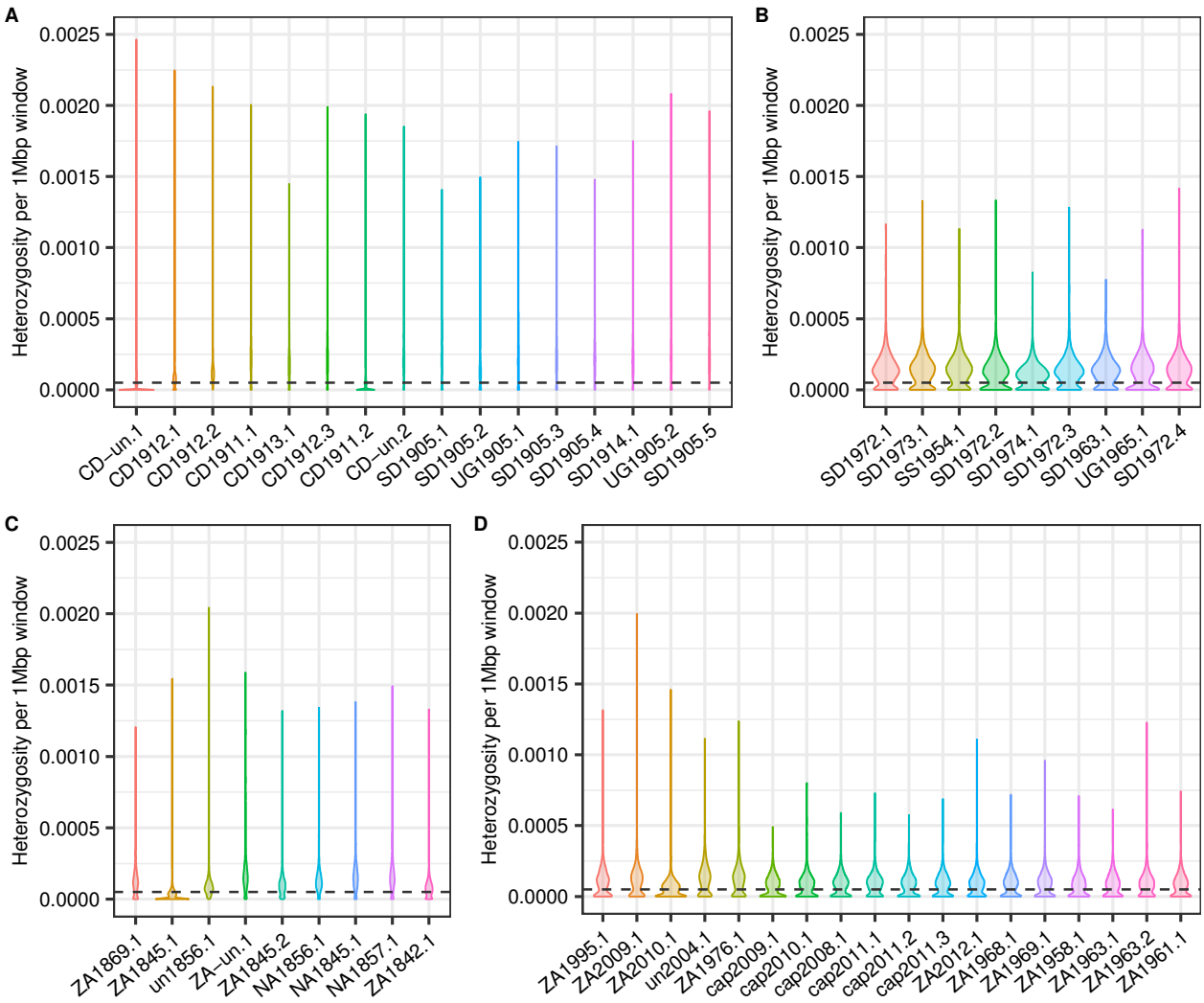


Figure S7. *Distribution of local estimates of heterozygosity per sample and threshold for RoH identification.* Density of windows (1 Mbp, 0.5 Mbp slide) along the range of heterozygosity values per sample for A) NWR1, B) NWR2, C) SWR1 and D) SWR2. Dashed lines indicate the chosen threshold of heterozygosity for assigning windows to RoH (5×10^{-5}). Samples CD-un.1 and CD1911.2 (NWR1), and ZA1845.1 (SWR1) were discarded from the F_{RoH} estimation due to an excess of missing data.

Genomic erosion in subsampled post-bottleneck groups

To exclude that sampling bias might be creating spurious patterns of genomic erosion, we calculated the individual metrics of genomic diversity for a subset of the samples. In NWR1, we kept only samples originating from Sudan-Uganda since all NWR2 were sourced from that area. From SWR2 only wild-born individuals were retained for comparison with all SWR1. This subsampled dataset consists of 38 individuals distributed across the four groups (see Table S3).

Delta estimators for genome-wide heterozygosity, exon heterozygosity and F_{RoH}, show that, after minimizing sampling bias, genomic erosion patterns remain (Figures S8 and S9). Post-bottleneck NWR and SWR show significantly lower levels of heterozygosity than pre-bottleneck counterparts (see Figure S8 and Table S3 for delta estimators), with the exception of F_{RoH} between NWR1 and NWR2, despite a slight increase among post-bottleneck NWR (Figure S9 and Table S3). Wild-born SWR2 remain significantly more inbred than SWR1 (Figure S9 and Table S3).

	n	Date of collection / birth	Median GW het	Unpaired Wilcoxon test p-value	Δ GW het	Median exon het	Unpaired Wilcoxon test p-value	Δ exon het	Median F_{RoH}	Unpaired Wilcoxon test p-value	Δ F_{RoH}
NWR1 SD-UG	8	1905-1914	0.000192	1.65E-04	-0.0827	0.000112	2.74E-02	-0.1431	0.1867	1.14E-01	0.1426
NWR2	9	1954-1974	0.000176			0.000096			0.2134		
SWR1	9	1905-1914	0.000158	2.04E-04	-0.2764	0.000087	3.39E-02	-0.1660	0.1831	5.49E-03	0.6761
SWR2 wild	12	1954-1974	0.000114			0.000073			0.3068		

Table S3. *Overview of the subsampled dataset and summary of delta estimators*. For NWR1 and SWR1, time span refers to collection dates; for NWR2 and SWR2, it refers to birth dates. For individual genome-wide heterozygosity, exon heterozygosity and F_{RoH}, comparisons of the medians were calculated with Wilcoxon tests. Delta estimators were calculated as median pre-bottleneck value minus median post-bottleneck value, divided by the median pre-bottleneck value. All metrics were based on transversion sites only.

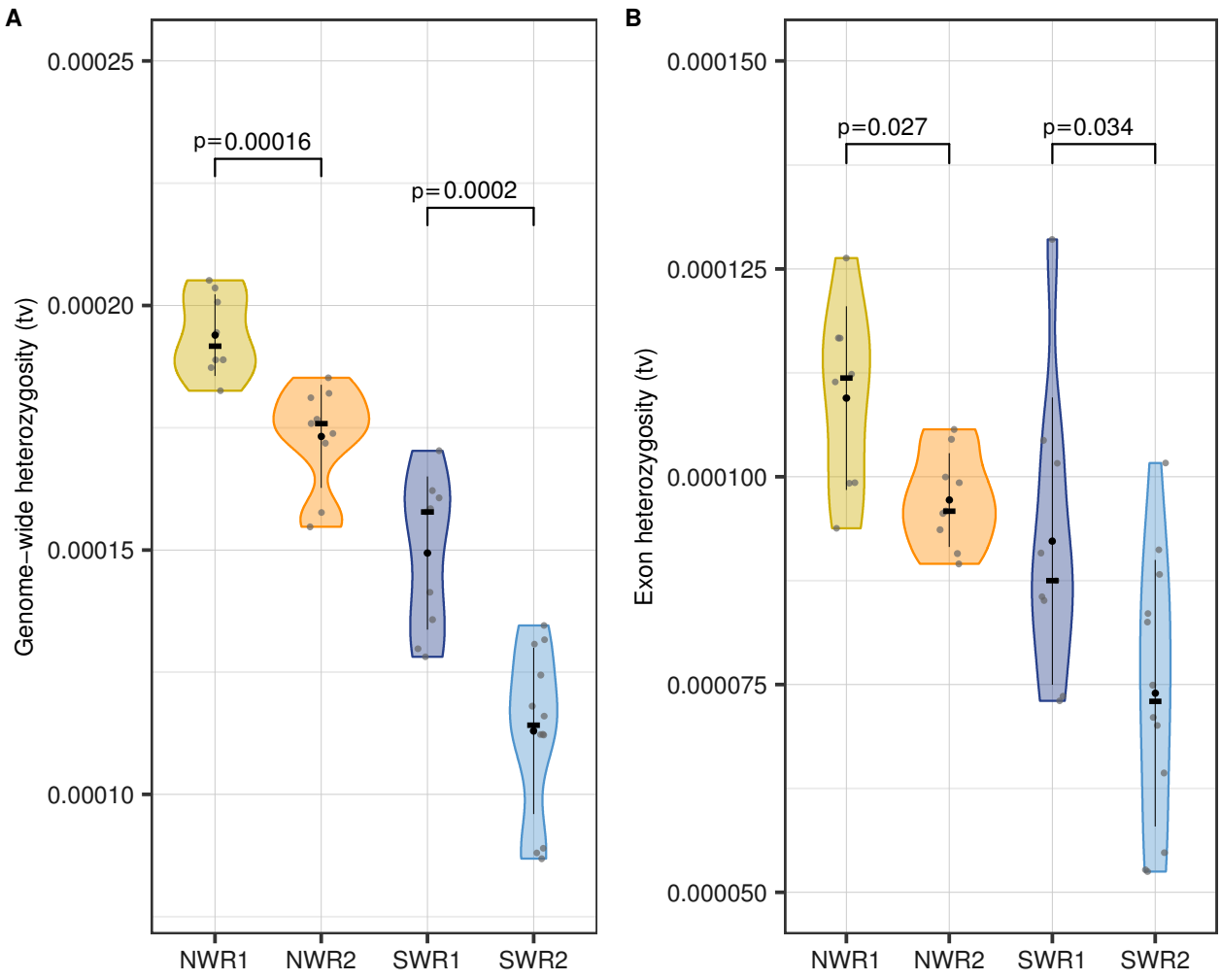


Figure S8. *Post-bottleneck white rhinoceroses show lower genomic diversity after subsampling to minimise sampling bias*. Estimates of genome-wide heterozygosity (A) and heterozygosity at regions annotated as exons (B) based on the per-sample SFS calculated for transversions and then corrected for depth of coverage for the subsampled dataset (n = 38 individuals). Black dots indicate the mean, black lines the standard deviation, and black cross-bars, the median per group. P-values correspond to median comparisons with unpaired Wilcoxon tests.

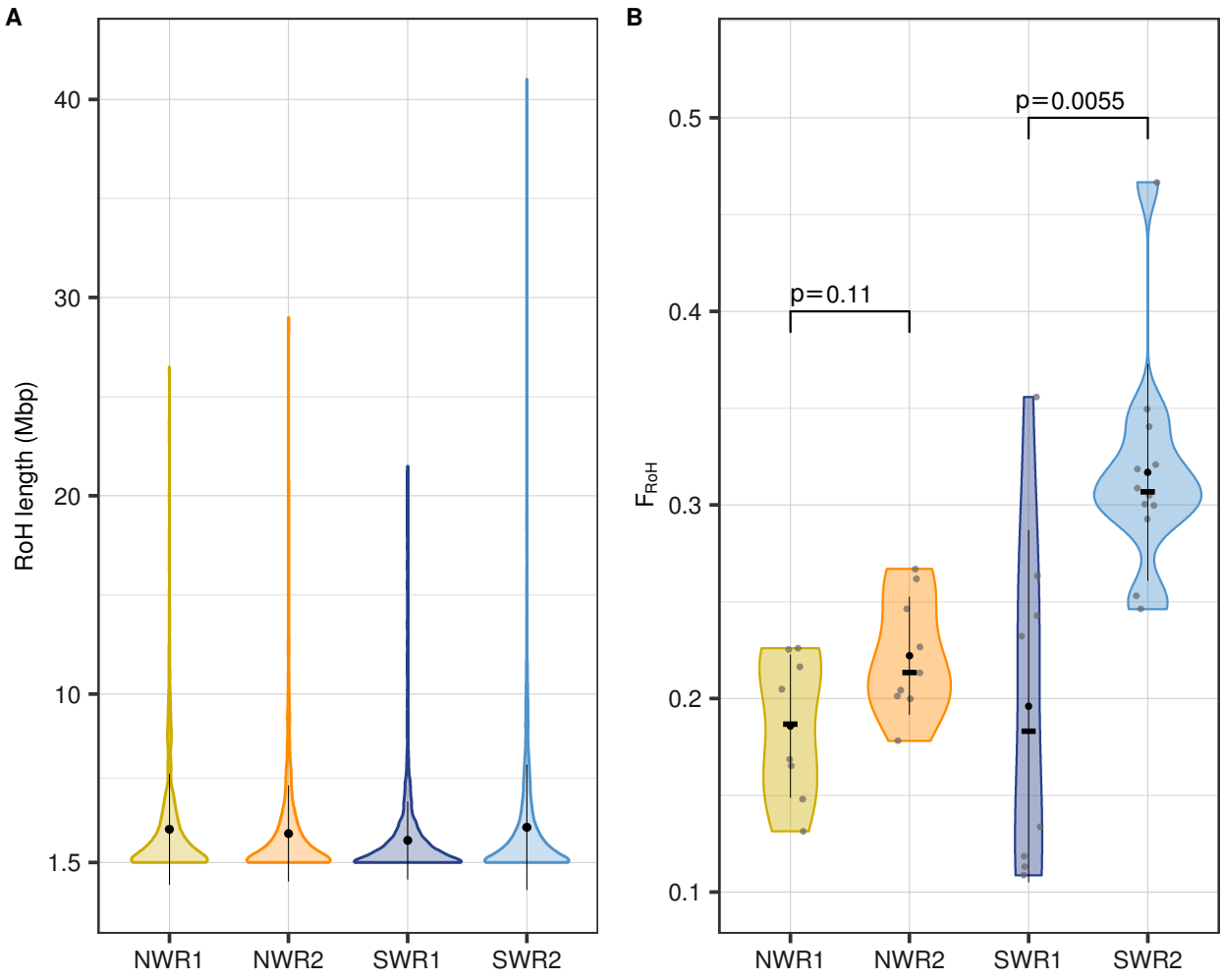


Figure S9. *Post-bottleneck white rhinoceroses show higher estimates of inbreeding after subsampling to minimise sampling bias*. A) Distribution of the length of RoHs for each of the four groups after subsampling and removing one outlier (n = 37 individuals); black dots indicate the mean RoH length per group and black lines the standard deviation. B) Estimates of individual F_{RoH} across groups after subsampling and removing one outlier (n = 37 individuals). Black dots indicate the mean, black lines the standard deviation, and black cross-bars the median per group. P-values refer to unpaired Wilcoxon tests to compare the medians.

Appendix to Chapter 2

Reduced version of Table S1 (see spreadsheet *WR_TableS1_v3* for full table; available at https://github.com/fasaba/WR_supplementary)

sample name	source	time window	country	locality	date	retained reads	mapped reads raw fraction	mapped reads clonal fraction	mapped reads unique fraction	depth of coverage	5'CtoT damage	corrected gw het	corrected exon het	F_{RoH}
ZA1869.1	Natural_History_Museum_Praha	SWR1	South_Africa	NA	1869	760368932	0.7681	0.2468	0.5785	11.30	0.0159	0.000130	0.000091	0.2555
CD-un.1	Royal_Museum_of_Central_Africa_Tervuren	NWR1	DRC	Faradje	NA	797010345	0.1999	0.4067	0.1186	2.84	0.0284	0.000455	0.000309	0.8814
CD1912.1	American_Museum_Natural_History	NWR1	DRC	Faradje	1912	790975295	0.4261	0.5166	0.2060	4.85	0.0261	0.000218	0.000099	0.3463
CD1912.2	American_Museum_Natural_History	NWR1	DRC	Vankerchovenville	1912	789271480	0.6338	0.5326	0.2962	7.08	0.0276	0.000202	0.000110	0.1222
CD1911.1	American_Museum_Natural_History	NWR1	DRC	Faradje	1911	734666100	0.8563	0.2316	0.6579	16.98	0.0200	0.000195	0.000122	0.2148
CD1913.1	American_Museum_Natural_History	NWR1	DRC	Faradje	1913	712628088	0.7267	0.2518	0.5437	9.92	0.0337	0.000201	0.000141	0.1670
CD1912.3	American_Museum_Natural_History	NWR1	DRC	Faradje	1912	812194932	0.7799	0.2174	0.6103	14.79	0.0210	0.000206	0.000126	0.1473
CD1911.2	American_Museum_Natural_History	NWR1	DRC	Faradje	1911	786298539	0.2310	0.2904	0.1640	3.39	0.0353	0.000242	0.000175	0.8218
CD-un.2	Royal_Museum_of_Central_Africa_Tervuren	NWR1	DRC	NA	NA	770376546	0.7856	0.2401	0.5970	13.09	0.0288	0.000202	0.000126	0.1399
SD1905.1	Powell_Cotton_Museum	NWR1	Sudan	By_Kero_Lado_Enclave	1905	796257179	0.7868	0.2339	0.6028	16.25	0.0244	0.000189	0.000112	0.2211
SD1905.2	Powell_Cotton_Museum	NWR1	Sudan	By_Kero_Lado_Enclave	1905	784968111	0.8500	0.6316	0.3131	7.42	0.0601	0.000189	0.000094	0.2460
UG1905.1	Powell_Cotton_Museum	NWR1	Uganda	Ouneri_Nr_Wadelai_Lado_Enclave	1905	731749543	0.8103	0.3335	0.5400	13.83	0.0317	0.000183	0.000099	0.2378
SD1905.3	Powell_Cotton_Museum	NWR1	Sudan	Fariala_Lado_Enclave	1905	772010465	0.8036	0.2412	0.6097	16.82	0.0230	0.000195	0.000099	0.1817
SD1905.4	Powell_Cotton_Museum	NWR1	Sudan	Fariala_Lado_Enclave	1905	764697017	0.8481	0.2342	0.6495	17.13	0.0171	0.000201	0.000117	0.1667
ZA1845.1	Swedish_Museum_of_Natural_History	SWR1	South_Africa	South_Africa	1845	845500662	0.1846	0.1666	0.1538	3.74	0.0210	0.000170	0.000129	0.8081
un1856.1	Swedish_Museum_of_Natural_History	SWR1	NA	NA	1856	783358580	0.7074	0.4615	0.3809	7.64	0.0364	0.000141	0.000074	0.2331
ZA-un.1	Swedish_Museum_of_Natural_History	SWR1	South_Africa	South_Africa	NA	799090671	0.8527	0.2616	0.6296	17.86	0.0153	0.000162	0.000085	0.1320
ZA1845.2	Swedish_Museum_of_Natural_History	SWR1	South_Africa	Kwazulu	1845	797602423	0.5859	0.3863	0.3596	9.38	0.0626	0.000128	0.000073	0.2457

NA1856.1	Swedish_Museum_of_Natural_History	SWR1	Namibia	Damaraland	1856	836210237	0.7245	0.2506	0.5429	10.25	0.0242	0.000158	0.000102	0.1219
NA1845.1	Swedish_Museum_of_Natural_History	SWR1	Namibia	Damaraland	1845	795489557	0.8286	0.2572	0.6155	15.02	0.0276	0.000161	0.000086	0.1244
NA1857.1	Swedish_Museum_of_Natural_History	SWR1	Namibia	Damaraland	1857	806548920	0.8468	0.2598	0.6269	16.03	0.0205	0.000158	0.000087	0.1394
ZA1842.1	Swedish_Museum_of_Natural_History	SWR1	South_Africa	Umfolozi	1842	808407179	0.6269	0.3549	0.4044	6.73	0.0321	0.000136	0.000104	0.3485
ZA1995.1	National_Museums_Scotland	SWR2	South_Africa	Umfolozi_Reserve	1995	709162239	0.7637	0.2919	0.5408	11.33	0.0059	0.000118	0.000084	0.3351
ZA2009.1	National_Museums_Scotland	SWR2	South_Africa	KwaZulu_Natal_Hluluwe_Umfolozi_National_Park	2009	644582780	0.8079	0.2416	0.6127	12.51	0.0041	0.000124	0.000091	0.3137
ZA2010.1	National_Museums_Scotland	SWR2	South_Africa	KwaZulu_Natal	2010	682559275	0.4502	0.2846	0.3221	5.57	0.0039	0.000132	0.000102	0.4908
un2004.1	National_Museums_Scotland	SWR2	NA	NA	2004	755695371	0.8280	0.2026	0.6602	15.42	0.0025	0.000131	0.000088	0.2779
SD1914.1	National_Museums_Scotland	NWR1	Sudan	Lado_Enclave	1914	717916347	0.8366	0.2369	0.6385	12.72	0.0171	0.000204	0.000126	0.1498
UG1905.2	Powell_Cotton_Museum	NWR1	Uganda	Ouneri_Nr_Wadelai_Lado_Enclave	1905	797945772	0.8531	0.2217	0.6640	18.58	0.0090	0.000187	0.000111	0.2348
SD1905.5	Powell_Cotton_Museum	NWR1	Sudan	By_Kero_Lado_Enclave	1905	810866100	0.8193	0.1916	0.6623	17.83	0.0117	0.000205	0.000117	0.1410
ZA1976.1	Natural_History_Museum_Vienna	SWR2	South_Africa	KwaZulu_Natal_Hluluwe_Umfolozi_National_Park	1976	595924021	0.8011	0.2268	0.6194	11.88	0.0037	0.000135	0.000083	0.2763
cap2009.1	San_Diego_Zoo_Institute_for_Conservation_Research	SWR2	captive	captive	2009	396393903	0.8152	0.1535	0.6901	17.20	0.0035	0.000080	0.000045	0.3886
cap2010.1	San_Diego_Zoo_Institute_for_Conservation_Research	SWR2	captive	captive	2010	391652474	0.8343	0.1562	0.7039	17.41	0.0016	0.000083	0.000050	0.3547
cap2008.1	San_Diego_Zoo_Institute_for_Conservation_Research	SWR2	captive	captive	2008	464991217	0.8454	0.1681	0.7032	20.49	0.0015	0.000086	0.000054	0.3444
cap2011.1	San_Diego_Zoo_Institute_for_Conservation_Research	SWR2	captive	captive	2011	392551261	0.8363	0.1344	0.7238	17.85	0.0016	0.000091	0.000054	0.2960
cap2011.2	San_Diego_Zoo_Institute_for_Conservation_Research	SWR2	captive	captive	2011	462564978	0.8542	0.1337	0.7400	21.62	0.0015	0.000083	0.000050	0.3539
cap2011.3	San_Diego_Zoo_Institute_for_Conservation_Research	SWR2	captive	captive	2011	493996657	0.8497	0.1530	0.7197	22.41	0.0017	0.000085	0.000053	0.3435
ZA2012.1	San_Diego_Zoo_Institute_for_Conservation_Re	SWR2	South_Africa	NA	2012	355398018	0.8556	0.1550	0.7230	16.17	0.0016	0.000089	0.000055	0.3299

	search													
ZA1968.1	San_Diego_Zoo_Institute_for_Conservation_Research	SWR2	South_Africa	NA	1968	450347121	0.8493	0.1317	0.7374	20.84	0.0017	0.000087	0.000052	0.3397
ZA1969.1	San_Diego_Zoo_Institute_for_Conservation_Research	SWR2	South_Africa	NA	1969	427489688	0.8499	0.1223	0.7460	20.05	0.0015	0.000088	0.000053	0.3386
ZA1958.1	San_Diego_Zoo_Institute_for_Conservation_Research	SWR2	South_Africa	NA	1958	166146617	0.7899	0.0037	0.7869	9.08	0.0005	0.000112	0.000064	0.3714
ZA1963.1	San_Diego_Zoo_Institute_for_Conservation_Research	SWR2	South_Africa	NA	1963	182943677	0.7640	0.0041	0.7608	9.70	0.0005	0.000112	0.000071	0.3490
ZA1963.2	San_Diego_Zoo_Institute_for_Conservation_Research	SWR2	South_Africa	NA	1963	183262608	0.7710	0.0048	0.7673	9.76	0.0007	0.000116	0.000075	0.3086
SD1972.1	San_Diego_Zoo_Institute_for_Conservation_Research	NWR2	Sudan	NA	1972	179636474	0.7785	0.0052	0.7744	9.68	0.0009	0.000177	0.000100	0.2154
SD1973.1	San_Diego_Zoo_Institute_for_Conservation_Research	NWR2	Sudan	NA	1973	187434291	0.7626	0.0031	0.7603	9.86	0.0009	0.000172	0.000094	0.2370
SS1954.1	San_Diego_Zoo_Institute_for_Conservation_Research	NWR2	South_Sudan	Equatoria	1954	183307507	0.7677	0.0032	0.7653	9.74	0.0007	0.000181	0.000099	0.2044
SD1972.2	San_Diego_Zoo_Institute_for_Conservation_Research	NWR2	Sudan	NA	1972	185450379	0.7549	0.0037	0.7521	9.74	0.0008	0.000158	0.000090	0.2725
SD1974.1	San_Diego_Zoo_Institute_for_Conservation_Research	NWR2	Sudan	NA	1974	166083178	0.5951	0.0026	0.5936	6.87	0.0008	0.000176	0.000096	0.2799
SD1972.3	San_Diego_Zoo_Institute_for_Conservation_Research	NWR2	Sudan	NA	1972	158755543	0.7706	0.0034	0.7680	8.50	0.0009	0.000185	0.000106	0.2145
SD1963.1	San_Diego_Zoo_Institute_for_Conservation_Research	NWR2	Sudan	NA	1963	171373726	0.7569	0.0047	0.7533	8.93	0.0009	0.000182	0.000104	0.2077
UG1965.1	San_Diego_Zoo_Institute_for_Conservation_Research	NWR2	Uganda	NA	1965	205073417	0.7345	0.0060	0.7302	10.42	0.0009	0.000155	0.000091	0.2982
SD1972.4	San_Diego_Zoo_Institute_for_Conservation_Research	NWR2	Sudan	NA	1972	179506453	0.7633	0.0035	0.7606	9.51	0.0005	0.000174	0.000096	0.2397
ZA1961.1	San_Diego_Zoo_Institute_for_Conservation_Research	SWR2	South_Africa	NA	1961	173971516	0.7718	0.0057	0.7675	9.27	0.0008	0.000112	0.000070	0.3608

Chapter 3

Historical sampling portrays a vanishing beast: population structure, phylogeography and genomic diversity in the black rhinoceros (*Diceros bicornis*)

Historical sampling portrays a vanishing beast: population structure, phylogeography and genomic diversity in the black rhinoceros (*Diceros bicornis*)

Fátima Sánchez-Barreiro[1,17,*], Ashot Margaryan[1], Michael V. Westbury[1], Claudia Fontsere[2], Michael W. Bruford[3,16], Isa-Rita M. Russo[3], Daniela C. Kalthoff[4], Thomas Sicheritz-Pontén[1,5], Love Dalén[6,7], Guojie Zhang[8-11], Tomás Marquès-Bonet[2,12,13], Yoshan Moodley[14,18], M. Thomas P. Gilbert[1,15,18,*]

[1]GLOBE Institute, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark

[2]Institut de Biologia Evolutiva (Consejo Superior de Investigaciones Científicas–Universitat Pompeu Fabra), Barcelona Biomedical Research Park, Doctor Aiguader 88, 08003 Barcelona, Catalonia, Spain.

[3]Cardiff School of Biosciences, Sir Martin Evans Building, Cardiff University, Museum Avenue, Cardiff, CF10 3AX, United Kingdom

[4]Department of Zoology, Swedish Museum of Natural History, Frescativägen 40, 114 18 Stockholm, Sweden

[5]Centre of Excellence for Omics-Driven Computational Biodiscovery (COMBio), Faculty of Applied Sciences, AIMST University, Kedah, Malaysia

[6]Centre for Palaeogenetics, Svante Arrhenius väg 20C, 10691 Stockholm, Sweden

[7]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Frescativägen 40, 114 18 Stockholm, Sweden

[8]Section for Ecology and Evolution, Department of Biology, University of Copenhagen, 2100 Copenhagen, Denmark

[9]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 650223, Kunming, China

[10]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, 650223, Kunming, China

[11]BGI-Shenzhen, 518083, Shenzhen, China

[12]National Centre for Genomic Analysis–Centre for Genomic Regulation, Barcelona Institute of Science and Technology, 08028 Barcelona, Spain.

[13]Institucio Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain.

[14]Department of Zoology, University of Venda, University Road, 0950 Thohoyandou, Republic of South Africa

[15]Norwegian University of Science and Technology, University Museum, 7491 Trondheim, Norway

[16]Sustainable Places Research Institute, Cardiff University, Cardiff CF10 3BA, United Kingdom

[17]Lead Contact

[18]Senior Authors

[*]Correspondence: fatima@palaeome.org (F.S.B), tgilbert@sund.ku.dk (M.T.P.G.)

Abstract

The black rhinoceros (*Diceros bicornis* L.) is a critically endangered species distributed historically across an extensive range in Subsaharan Africa. Hunting and habitat disturbance have diminished both its numbers and range of distribution since the 19th century, but a poaching crisis in the late 20th century drove them to the brink of extinction. Today, ~5500 individuals remain in Africa, mostly in managed populations. Genetic and genomic assessments can greatly increase our knowledge of the species and inform management strategies. However, when a species has been so severely reduced to a fragmentary version of what it was just a century ago, surveying present populations will only yield an incomplete picture. To gain understanding of the species-level and range-wide population genomic patterns, we retrieved and analysed whole-genome data from 64 black rhinoceros museum specimens collected between 1775 and 1981. Our results indicate that the black rhinoceros was structured into six genomic groups largely mirroring geography, particularly a north-south cline. We detected, however, varying degrees of admixture among groups, and that some geographical barriers created discontinuities in the population structure, mainly the Zambezi river and the Kenyan Highlands. The individual-level genomic diversity featured a maximum in the middle of the range, and decayed toward the periphery. Based on these results, we suggest the occurrence of two waves of expansion from Central Africa in the black rhinoceros' evolutionary history, one toward the south and a more recent one toward the north. This comprehensive historical portrait allowed us to ascertain the genomic ancestry of 27 newly re-sequenced genomes from extant populations of black rhinoceros. Lastly, we propose how to employ the insight gained from this temporal dataset to inform current management strategies, with the goal of preserving what is left of the black rhinoceros diversity.

Introduction

The black rhinoceros (*Diceros bicornis* L.) is a critically endangered African megaherbivore [1]. As of the end of 2017, some 5,500 black rhinoceroses were left across the continent [2], reflecting a modest, yet positive demographic recovery after the lowest recorded census size of 2,354 animals in the early 1990's. This historical low was the result of a ca. 98% decline in their wild population between 1960 and 1995, owing principally to intense poaching on them for the rhinoceros horn trade [1]. Prior to 1960, it had been the most abundant extant rhinoceros species, although its population had started to decline in the 19th century due to habitat clearance and hunting [1].

Their historical distribution encompassed a vast, continuous area across Subsaharan Africa, that spanned a broad range of habitats, from bushland and grassland, to desert, only avoiding areas of dense tropical rainforest [3] (Figure 1). Currently outside of zoos, the species survives almost solely

in protected areas, mostly in South Africa, Namibia and Kenya [2], and their persistence depends heavily on active conservation efforts, including population management. These practices should ideally take the species' natural population structure, and subspecies-level taxonomy, into account. However both of these are not straightforward. The latter has been a bone of contention for rhinoceros experts for decades, and although until recently the most commonly used classification refers to four ecotypes or management units, often described as subspecies: *D.b. bicornis* or southwestern black rhinoceros, *D.b. minor* or south-central, *D.b. michaeli* or eastern black rhinoceros, and the extinct *D.b. longipes* from Western Africa [1] - not all experts are in agreement [4].

With regards to population structure, genetic assessments are key sources of information. In this regard, although a substantial body of work exists for the black rhinoceros, these studies mostly focus on either single subspecies, or a subset of the managed populations [5–7], and the species- and range-wide understanding of the population structure and diversity has been less well explored. Additionally, given the major population extirpations and bottlenecks that the black rhinoceros has been through, it is clear that its status today may be a far from accurate reflection of how it looked just half a century ago. Yet a better understanding of its pre-decline status will be essential for not only expanding our knowledge of its ecology and evolution, but ultimately aid conservation efforts. A natural solution to gathering a truly representative sample of the black rhinoceros is nonetheless possible, thanks to the wealth of historical black rhinoceros specimens preserved in museum collections.

This temporal sampling approach to study black rhinoceros genetics was explored for the first time by [8], where the authors investigated the species-level population structure, phylogeny, and genetic erosion through time. However, their analyses were limited to molecular data of mitochondrial and microsatellite markers. Many of their conclusions are therefore based on the history of the mitochondrial control region, which is after all just one, maternally inherited locus, and only a small fraction of their historical samples yielded enough microsatellite sequence data.

In this study we aimed to expand the resolution and scope of this previous work through taking advantage of the recent developments in palaeogenome sequencing techniques. Specifically, we generated whole-genome re-sequencing data for a set of historical black rhinoceros specimens collected between 1775 and 1981, and representing most of their historical range of distribution. We furthermore supplemented this with extra genomic data from a number of individuals from extant black rhinoceros populations. Ultimately our goal was to better resolve the patterns of population structure, gene flow, phylogeny and diversity in the black rhinoceros prior to their abrupt decline. In parallel, we aimed to evaluate where modern samples fall on this updated portrait of the historical

black rhinoceros.

Specifically, we generated and analysed a dataset including re-sequenced genomes of 64 historical and 27 modern black rhinoceroses to investigate: a) the range-wide population structure of the black rhinoceros in historical times, and the relative placement of modern black rhinoceroses within it; b) the phylogenomic history of the black rhinoceroses in our dataset; c) the patterns of gene flow among historical populations of black rhinoceros; and d) the distribution of individual genomic diversity across geography.

Results

We generated shotgun, re-sequence data for 100 individual black rhinoceroses originally sampled from across the historical and contemporary range of the species. The historical specimens ranged in collection date between 1775 and 1981. We mapped the raw sequence data against both the published black rhinoceros mitochondrial genome (NC_012682 [9]), and against a soon-available black rhinoceros whole-genome assembly ([10], in press). After excluding samples of depth of coverage <1x, and one putatively identical sample (see *Filtered sample sets* in Methods), our final dataset consisted of complete mitochondrial genomes from all 100 specimens, and whole-genome data of at least 1x depth of coverage for 91 of them. Sixty four of the re-sequenced genomes derive from historical specimens, of which 54 were from samples whose associated metadata included coordinates indicating geographical origin (Figure 1). The 27 modern samples with whole-genome coverage were all georeferenced, and derive from extant populations: one Namibian (Etosha National Park), three Kenyan (Maasai Mara, Nairobi National Park and Ol Pejeta), and two South African (iMfolozi and Mkhuze) reserves (Figure 1).

Historical samples were named with the following structure: the alpha-2 code of the country of origin, the year of collection, and an index number (to distinguish samples of identical country and year). Country of origin was not known for only two samples, which we indicated by replacing the country code by ‘un’. Modern samples are labelled with simpler identifiers that include the country code (for South Africa and Namibia) or reserve code (for the Kenyan individuals) followed by an index number (see Table 1 for further details on the distribution of samples across countries in our dataset).

As expected, the DNA sequence data from the historical specimens showed signals characteristic of ancient DNA, including cytosine deaminations, shorter library insert sizes and sizeable fractions of non-endogenous DNA (Figure S1). As such, the average depth of coverage for the nuclear genomes

were generally lower and more variable among the historical specimens (ranging between 1.27x and 20.11x), compared to modern samples (ranging between 7.37x and 22.78x) (Table S1). Levels of aDNA damage consistent with cytosine deamination among historical samples ranged from 0.5% to 5%, and endogenous DNA content ranged from 5% to 62% (Table S1 and Figure S1).

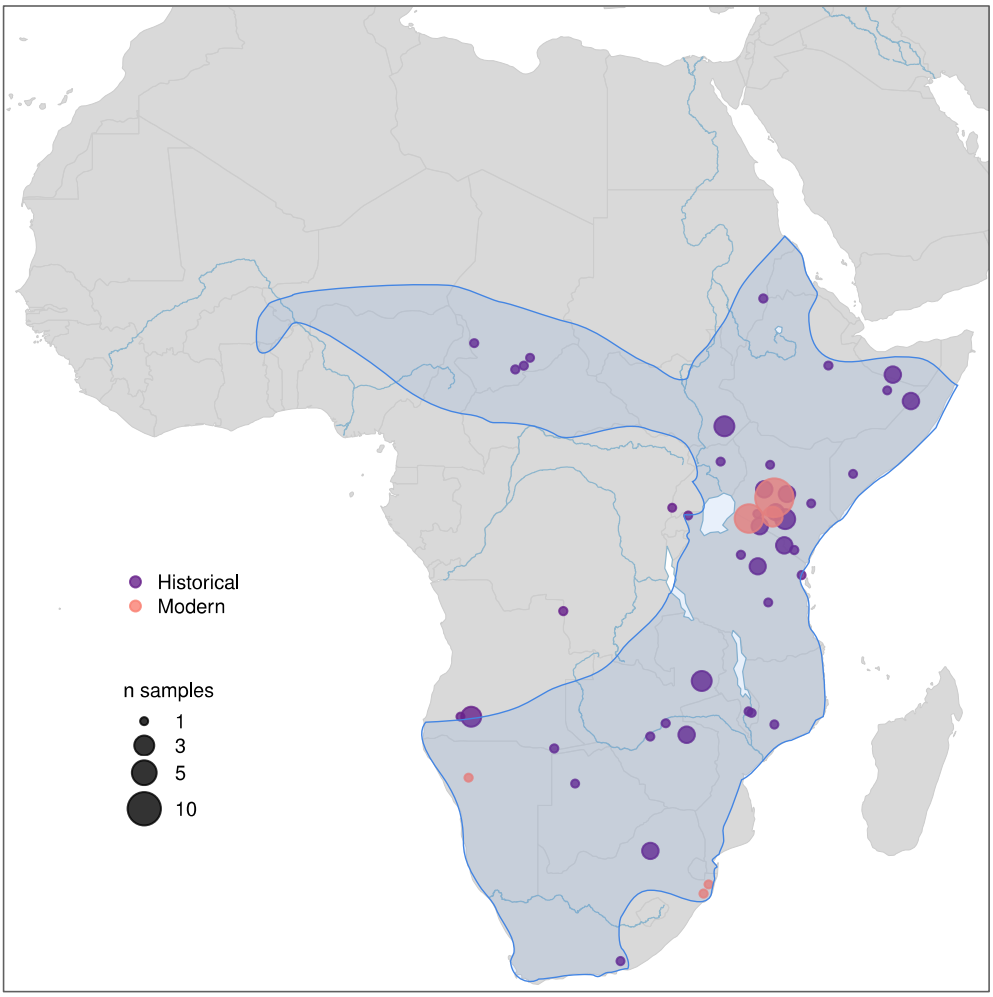


Figure 1. *Historical range of distribution of the black rhinoceros and sampling locations*. The blue shade indicates the estimated historical range of distribution of the black rhinoceros (from [3]). Coloured dots represent the sampling locations of 90 georeferenced samples in our dataset; an additional eight samples lacked coordinates, but their country of origin is known; two samples were of unknown origin (Table 1 and Table S1). Dot size represents the number of samples collected at each particular location.

Country / reserve	Code	n (samples)	n (re-sequenced genomes)	Genomic group
Historical				
Angola	AO	6	5	S
Botswana	BW	1	1	S
Chad	TD	4	4	NE, NW
DRC	CD	2	2	EA
Ethiopia	ET	3	2	NE
Kenya	KE	19	19	EA, CE
Malawi	MW	2	2	RU
Mozambique	MZ	1	1	CE
Nigeria	NG	1	1	NW
Somalia	SO	6	5	NE
South Africa	ZA	4	2	S
South Sudan	SS	3	3	EA, CE
Sudan	SD	1	0	-
Tanzania	TZ	10	8	CE, RU
Uganda	UG	1	1	EA
unknown	un	2	2	CE, S
Zambia	ZM	3	3	CE
Zimbabwe	ZW	4	3	S
TOTAL		73	64	
Modern				
Maasai Mara Game Reserve	MA	7	7	modern-CE
Nairobi National Park	NNP	3	3	modern-CE
Ol Pejeta	OP	14	14	modern-CE
Namibia	NA	1	1	modern-S
South Africa	ZA	2	2	modern-S
TOTAL		27	27	

Table 1. *Overview of the number and origin of the samples in the dataset*. For historical and modern samples separately, the countries of origin and their corresponding alpha-2 codes are specified. Although full mitochondrial genomes were recovered from 100 samples, whole-genome data of sufficient quality was only recovered from a subset of samples in each country/reserve, as indicated by the column headers *n (samples)* and *n (re-sequenced genomes)*, respectively. For each country, the genomic groups present are listed under *Genomic group*.

Mitogenome-based reassessment of black rhinoceros population structure and phylogeny

We first examined the maternal population structure of the black rhinoceroses in our dataset with a Bayesian phylogenetic analysis of 100 full mitochondrial genomes. Our results recapitulated the nine mitochondrial haplogroups previously reported by [8], based on 403 control region sequences of historical and modern individuals, both with regards to geographical and phylogenetic structure.

Following the nomenclature of this previous study, we termed these haplogroups West Africa (WW), Chari-Victoria (CV), Northeast (NE), East Africa (EA), Central Africa (CE), Ruvuma (RU), South-North (SN), Southwest (SW) and Southeast (SE) [8]. Each of these forms a monophyletic clade that largely mirrors geography, except in the case of CV, where samples from Chad and Tanzania share the same mitochondrial haplogroup, despite the vast geographical distance between their locations of origin (Figure 2).

The modern samples from Kenya feature haplotypes within the EA and CE diversity, while the modern Namibian sample falls among SW, and the two modern South Africans among SE (Figure 2). This is unsurprising given their geographical origins, if one assumes these individuals are not descendants of animals translocated from distant regions.

A dated phylogeny with the white rhinoceros as an outgroup also supports the findings of [8] regarding the estimates of split times between mitochondrial lineages in the black rhinoceros (Figure S5). The basal WW haplotype is the oldest lineage, separating from the rest ca. 700,871 ya (95% HPD interval 552,138-852,689 ya). The branch leading to NE and CV split ca. 611,333 ya (95% HPD interval 489,807-752,925 ya), and the subsequent separation of the Southern African clades (SW, SN, SE) from the branch leading to the CE and EA clades happened ca. 334,328 ya (95% HPD interval 264,830-415,617 ya) (Figure S5).

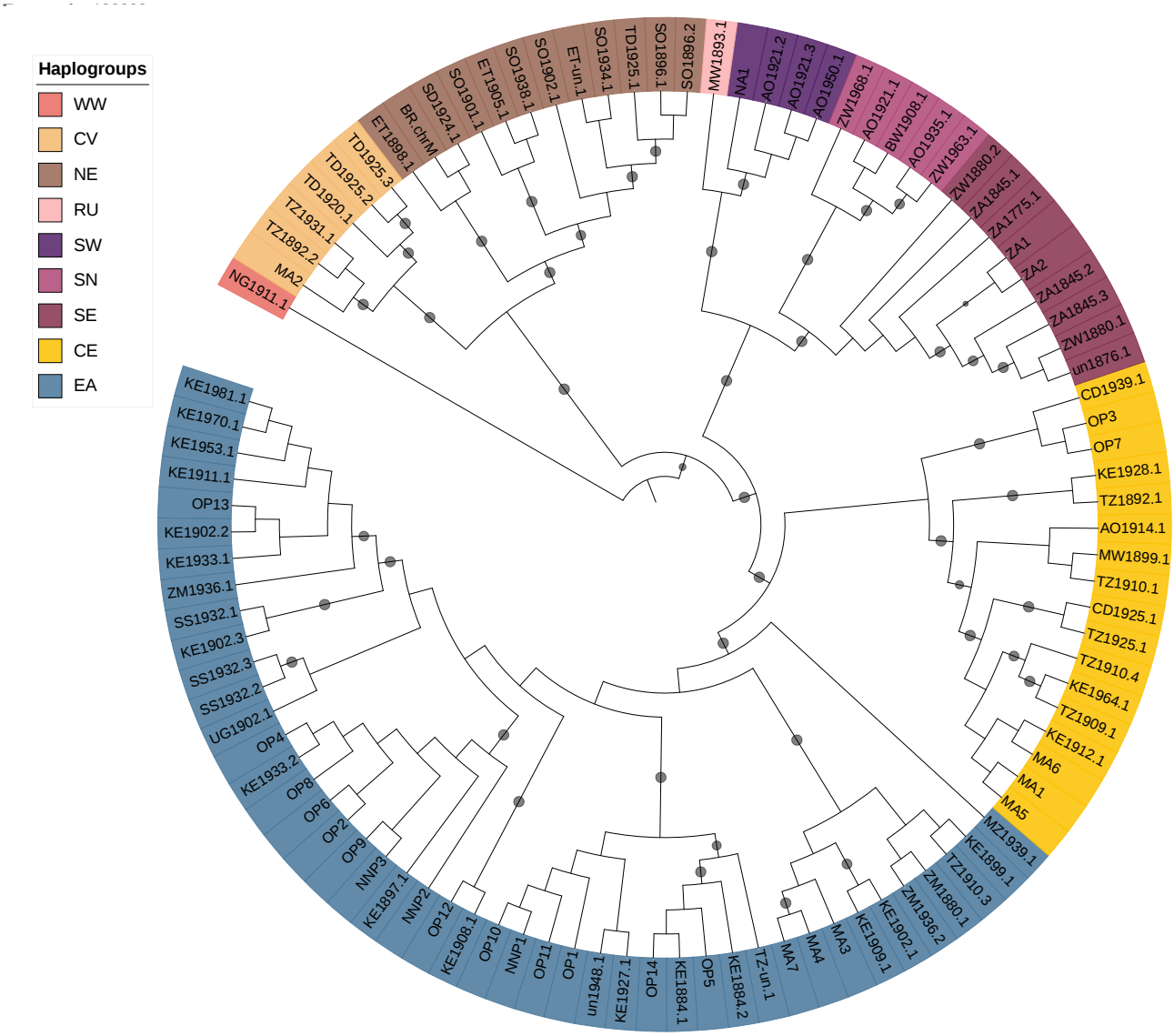


Figure 2. *Phylogenetic analysis of black rhinoceros full mitochondrial genomes*. Intraspecies, unrooted cladogram based on a Bayesian phylogenetic analysis of mitochondrial genomes in the black rhinoceros. Grey dots indicate branches with >90% bootstrap support. Coloured ranges highlight geographically-informed mitochondrial haplogroups, with nomenclature following [8], whose authors performed a similar assessment based on the mtDNA control region.

Although both our mitogenome analysis and the former control region-based analysis yielded consistent results, these results reflect the history of just one locus (moreover a haploid, maternally-inherited one). Thus we subsequently explored the signal within the genome-wide nuclear data from many of these samples.

Black rhinoceroses exhibited geography-driven population structure

Of the 91 samples from which genome-wide data was available in our dataset, 64 originated from historical specimens, and were used to portray the population structure of the species before the decline of the late 20th century. We used genotype likelihoods of >900,000 variant sites as input for each of these population structure analyses (see *Computing genotype likelihoods* in Methods).

We first performed a Principal Component Analysis (PCA) considering three sample sets separately: the 64 historical samples, and the 20 (out of 27) modern samples that we could identify as not being closely related (see *Relatedness test* in Supplementary Material), and these 84 samples combined. The first two principal components (PCs) showed that the strongest axis of variation in the dataset separates southern samples from the rest. Among the historical samples, individuals from south of the Zambezi river (Angola, Botswana, Namibia, South Africa and Zimbabwe) formed a clear group apart from central and northern samples (Figure 3A).

The PC2 revealed that populations north of the Zambezi river also showed structure that largely followed a latitudinal gradient. Specifically, the samples from Chad and Nigeria clustered closely in the west, as did those from Ethiopia and Somalia in the east. Moving southwards, individuals from Sudan, Uganda and Northern Kenya formed a group, as did individuals from Central Africa (Southern Kenya, Tanzania, Zambia, Mozambique), with a small cluster of two samples from Malawi, one Tanzanian and one from DRC at the extreme of the cline (Figure 3A).

Modern samples showed the same north-south divide in PC space, as the South African and Namibian samples fell distantly from the Kenyan along PC1 (Figure 3B). When included alongside the historical samples, the origin of the modern samples became apparent (Figure 3C). Modern Kenyan samples clearly fell among the historical diversity of Sudan, Kenya, Tanzania and other Central African countries. However, those from Nairobi National Park (NNP) and Ol Pejeta (OP) fell in between different clusters of historical samples, while those from Maasai Mara (MA) clustered tightly.

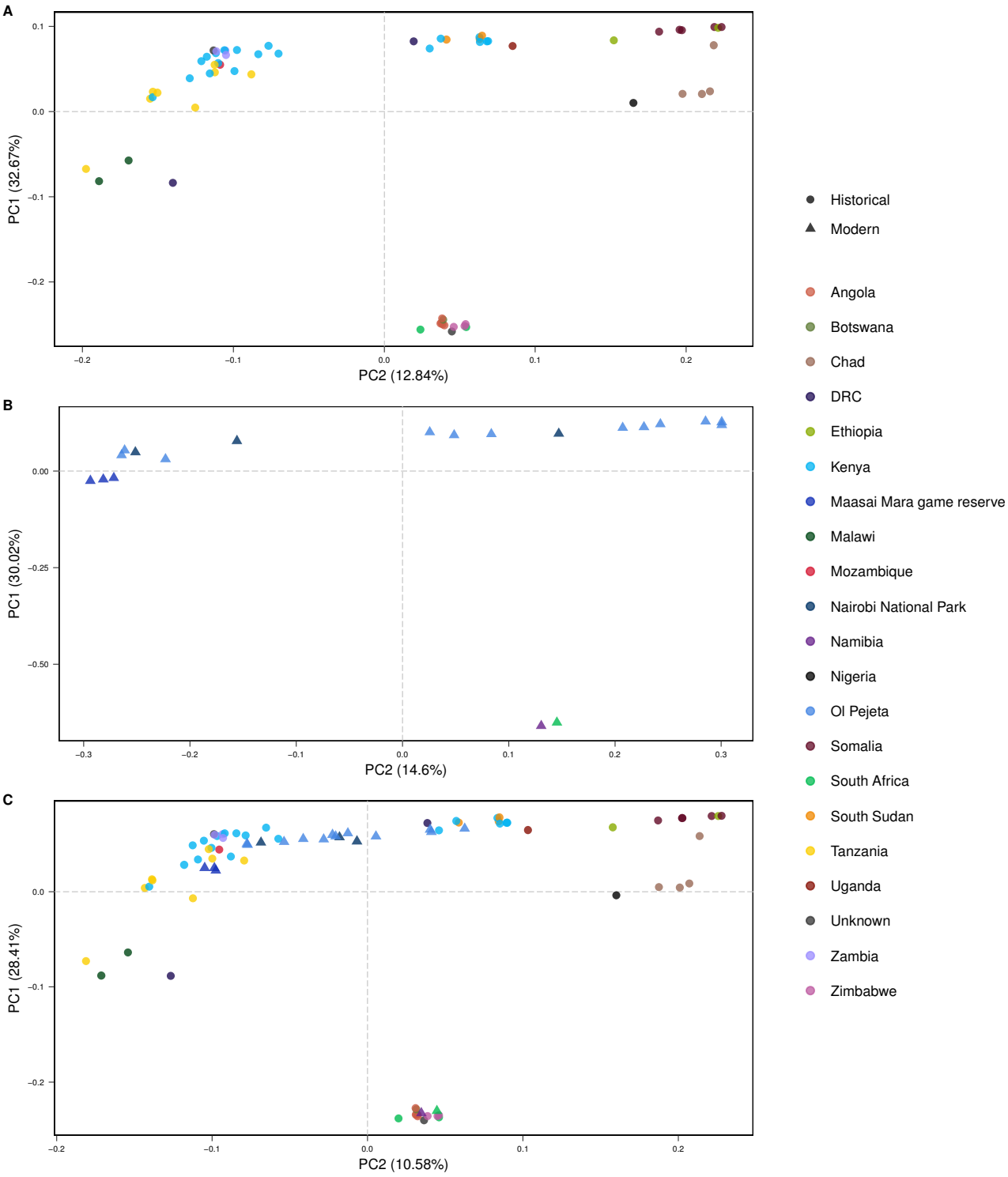


Figure 3. *PCA analysis of black rhinoceros historical and modern sample sets.* The first two principal components of PCA analyses based on genome-wide data are visualised for A) 64 historical black rhinoceroses, B) 20 modern, and C) all 84 samples combined. PC1 is represented on the *y* axis, and PC2 on the *x* axis to aid visualising the north-south geographical pattern driving most of the variation in the data.

We also conducted a non-linear reduction of dimensionality of the data with UMAP [11]. This method encompasses more of the variability in the dataset than the first few components of a PCA, but is sensitive to sample size. Therefore we investigated only the output of UMAP for the combined sample set of historical and modern unrelated samples (n = 84). The result of this approach largely supported the aforementioned geographical patterns (Figure 4). The north-south trend was replicated, and samples were sorted into the same groups, although intra-group resolution is higher thus showing further substructure, particularly in Central Africa (Figure 4). Southern African samples also appeared more structured, with historical Western Angolan and modern Namibian samples separating from the rest of the region. Modern samples from South Africa fell very close to historical counterparts. Kenyan samples from NNP and OP seemed to draw their mixed origin from different genomic pools in East and Central Africa, while MA samples aligned with a single origin (Figure 4).

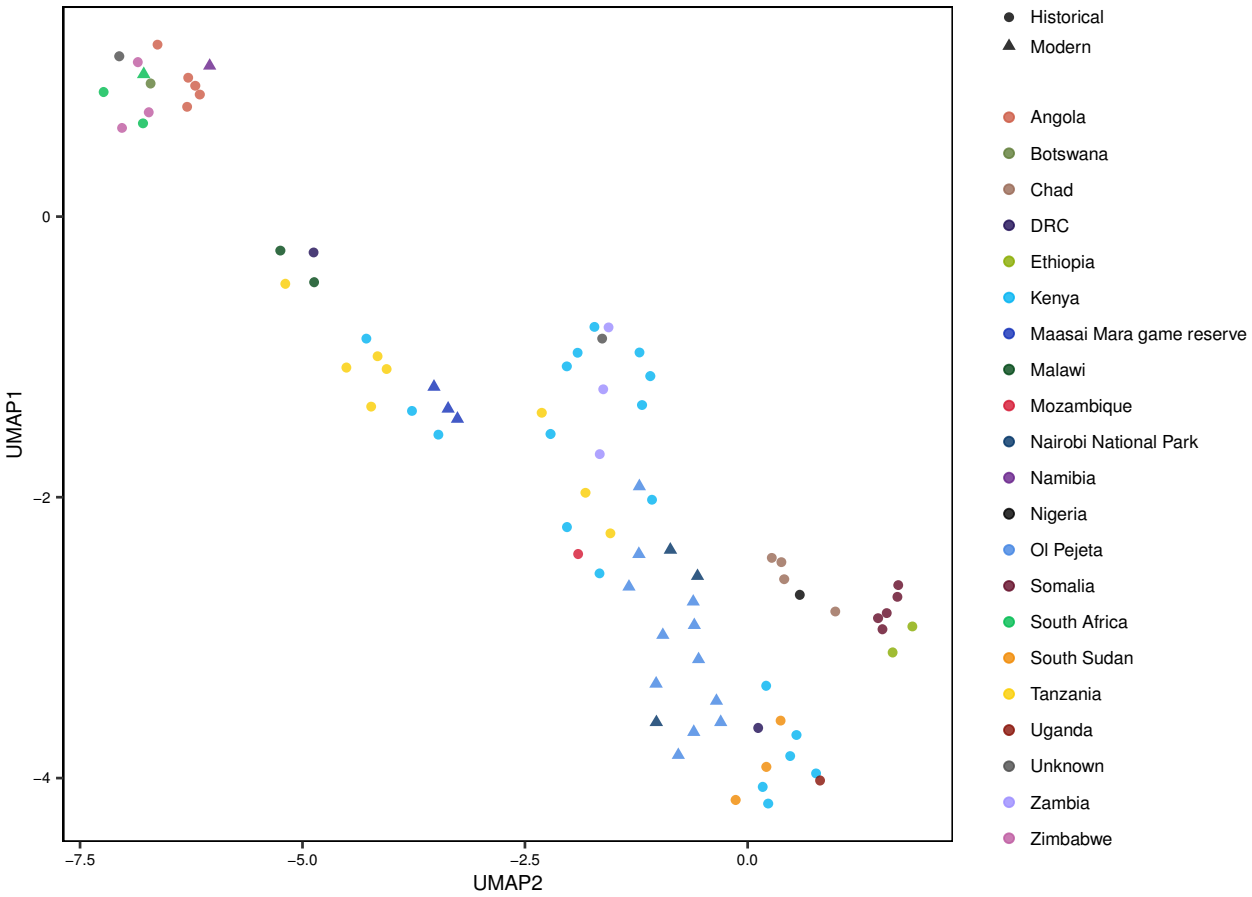


Figure 4. *UMAP analysis of historical and modern black rhinoceroses.* Visualisation of the two-dimensional output of UMAP based on genome-wide data for 84 historical and unrelated modern samples combined. The first dimension (UMAP1) is represented on the *y* axis, and the second dimension (UMAP2) is on the *x* axis, to aid visualisation of the north-south geographical pattern driving most of the variation in the data.

To complement these two model-free approaches, we next conducted an admixture analysis for the set of historical and unrelated modern samples (n = 84). At a value of assumed ancestral population of two (K = 2), southern samples separated from the rest. As K grew to three and four, northeastern and northwestern groups of samples became distinct clusters (Figure 5). At K = 6, another group appeared, gathering samples from the southern part of Central Africa, and at K = 7, South Sudan, Uganda and some Kenyan samples established a group (Figure 5). Modern samples fell among historical samples from similar geographical origins: i.e. among the southern historical samples in the case of NA1 and ZA1, and among Central African historical diversity for MA, NNP and OP (Figure 5). Reassuringly, these trends were largely consistent with those revealed by PCA and UMAP analyses.

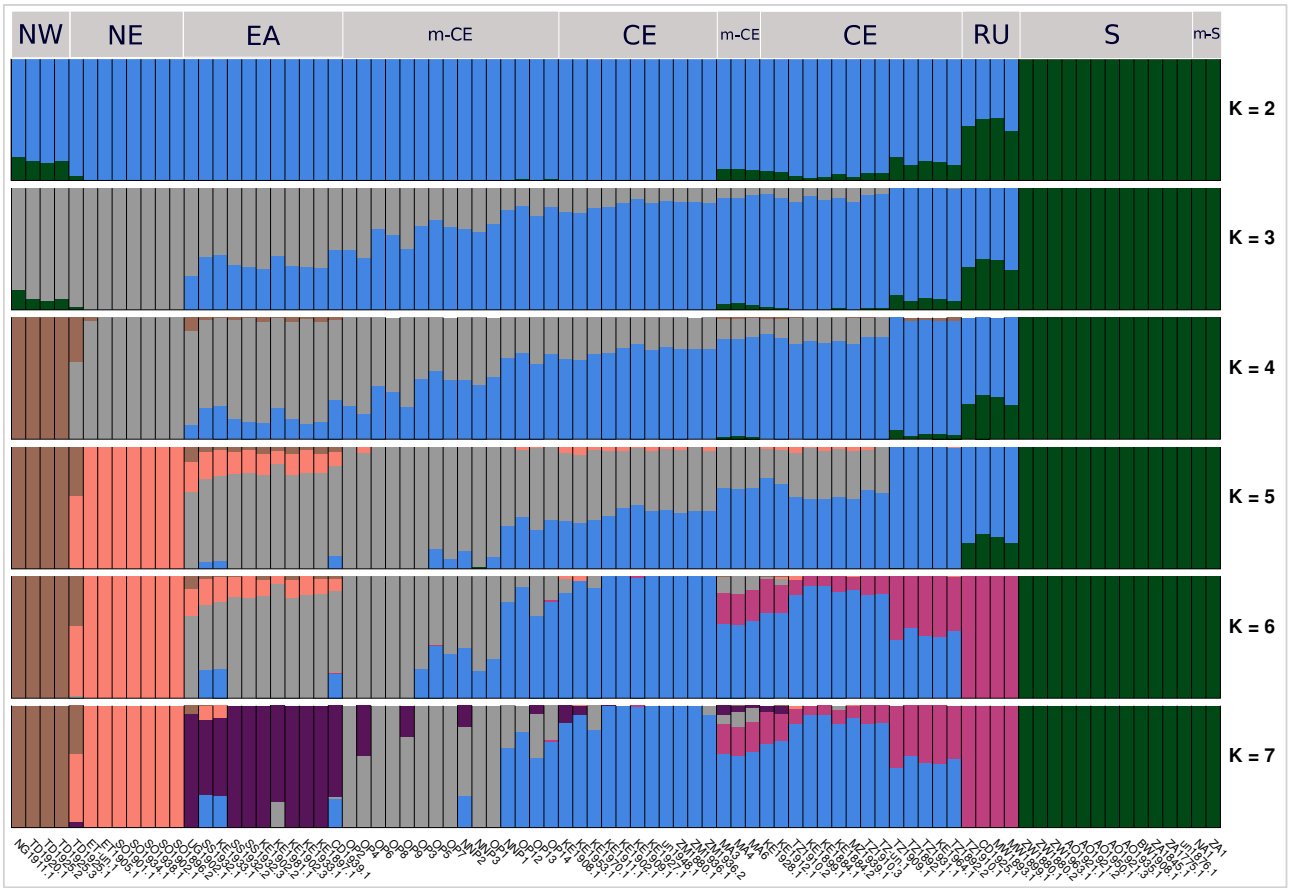


Figure 5. *Admixture analysis of historical and modern black rhinoceroses.* For 84 unrelated individuals and values of K between two and seven, the run of highest likelihood of a total of 50 runs per K is displayed. Grey labels at the top indicate geographically-informed genomic groups among the historical samples, and genomic-geographic origin in the case of the modern samples. The abbreviations *m-CE* and *m-S* stand for 'modern-CE' and 'modern-S'.

The portrait drawn by these population structure analyses suggested the existence in historical times of six genomic clusters distributed according to geography. Although these groups were not fully consistent with those previously reported based on microsatellite data [8], we followed their nomenclature for consistency. Thus we propose the following geographically-informed genomic clusters (broadly understood as populations): Northwest (NW), Northeast (NE), Eastern Africa (EA), Central Africa (CE), Ruvuma (RU) and Southern Africa (S) (Figure 6).

As suggested by the structure analyses, varying degrees of admixture existed between these historical groups. The case of the sample TD1925.1 was particularly striking, since it displayed a higher affinity for NE than for NW samples, despite its geographical origin in Chad. We are cautious, however, because in a cross-contamination analysis this sample showed less than 95% authentic mitochondrial DNA (see *Cross-contamination assessment* in Methods). Other, broader admixture patterns were observed between the geographically overlapping EA and CE groups in Northern Kenya, and between CE and RU in Southern Central Africa (Figures 5 and 6). In contrast, the groups at the extremes of the range, S, NW and NE, appeared as more isolated and showed little to no admixture (Figure 5).

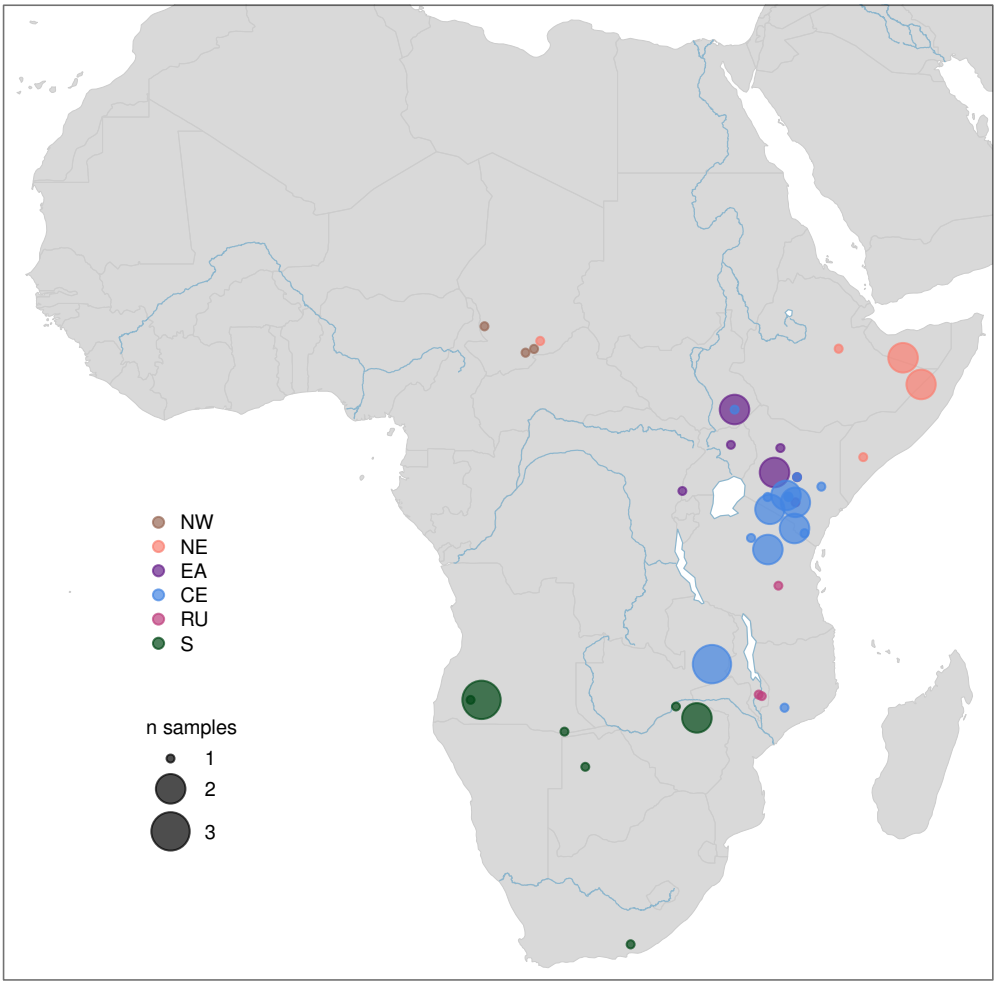


Figure 6. *Summary of the inferred historical population structure in the black rhinoceros*. Based on the

results of the genome-wide analyses detailed above (see Figures 3-5), the six geographically-informed genomic groups are placed on a map via labelling of georeferenced samples. Only historical and georeferenced samples are depicted (n = 54). Dot size represents the number of samples from each particular location.

The phylogenomic structure of the black rhinoceros was (mostly) congruent with its population structure

We explored the phylogenomic relationships within our full whole-genome dataset (n = 91) by computing a rooted maximum-likelihood tree with IQ-Tree v 1.6.8 [12]. For this, we picked at random 100 regions of 200 kbp across 47 autosomal scaffolds of the nuclear genome (see *Choice of scaffolds for variant site finding* in Methods), and we used the white rhinoceros as an outgroup.

The phylogenomic structure recovered largely matched the population structure observed. Samples that belonged to a given geographically-informed genomic group formed monophyletic clades (Figure 7). Regarding their relative phylogenomic relations, the first branch split gave rise to the RU and S groups, while the complementary branch led to CE. The groups NW and EA were nested within CE, and NE was nested within EA (Figure 7).

Some striking particularities are worth mentioning. The modern samples from Namibia and South Africa were placed where it would be expected, among other southwestern and other southeastern samples respectively. However, in contrast to their behaviour in the population structure analyses, the modern Kenyan samples formed a single monophyletic clade amidst CE (Figure 7), instead of showing distinct origins for MA versus NNP and OP samples.

An additional interesting pattern was that of the group RU. In the assessment of population structure, RU individuals appeared in between CE and S, but closer to CE. At a phylogenomic level, however, they shared their origin with S (Figure 7). This incongruence prompted us to conduct a test of gene flow particularly on the affinities of RU, as detailed in the section below.

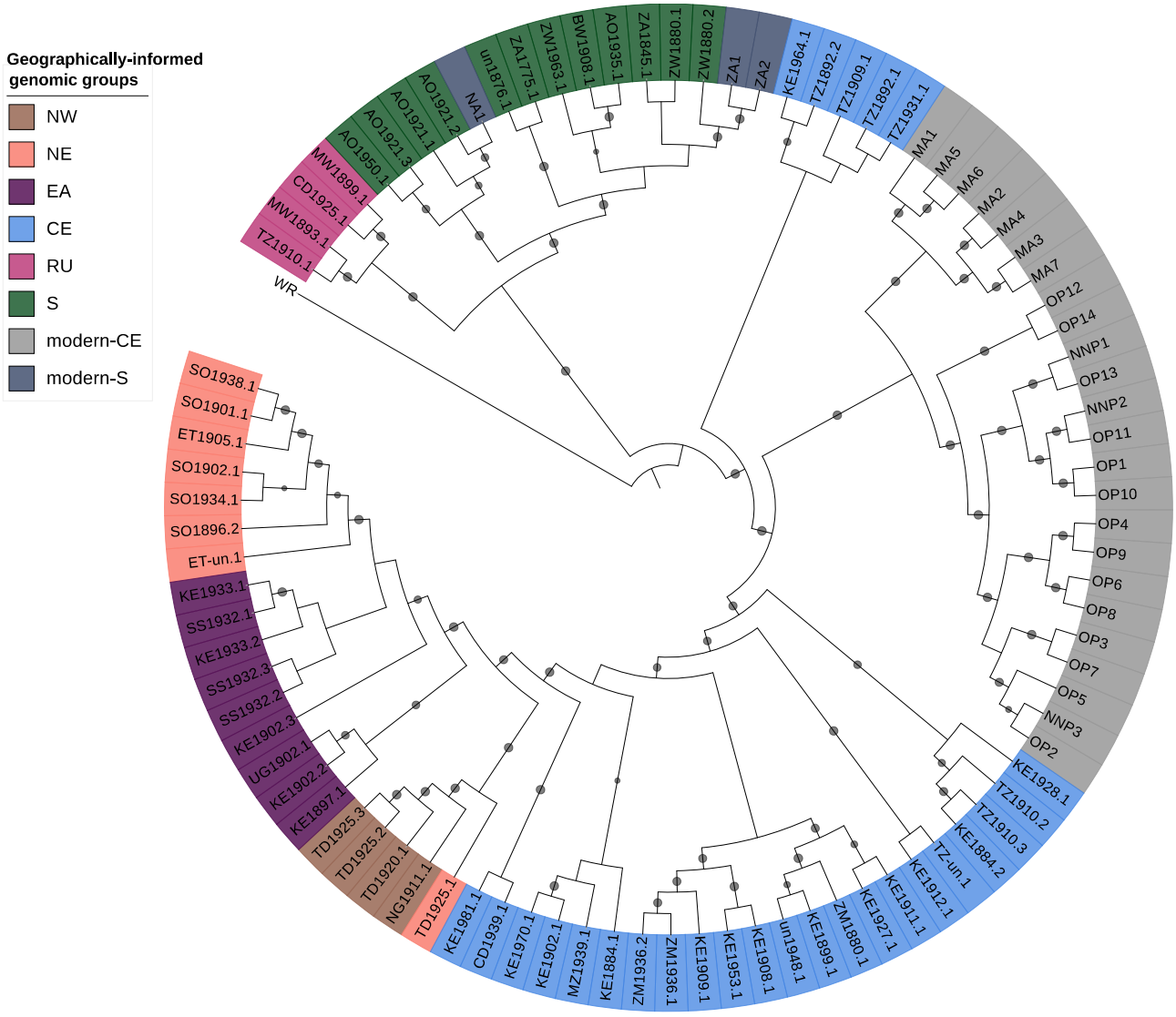


Figure 7. *Phylogenomic analysis of historical and modern black rhinoceroses*. Cladogram of a maximum likelihood assessment of 91 black rhinoceros genomes using a white rhinoceros as an outgroup (labelled as WR). Grey dots indicate branches with bootstrap support >90%. Colour ranges represent the geographically-informed genomic clusters inferred from the structure analyses (see Figures 3-6).

Genomic differentiation was not uniform across the historical range of the black rhinoceros

The outcome of the structure and phylogenomic investigation of the black rhinoceros led us to hypothesise that various degrees of connectivity among the different groups existed in the past. We therefore explored potential patterns of gene flow, first for the particular case of RU, and then more generally for the full historical range of distribution.

To clarify the history of the RU group, we computed D-statistics for relevant topologies of the shape (((H1,H2)H3)O) where the outgroup (O) was the white rhinoceros, H3 the test sample and H2

and H1 were samples from two other groups. Negative D-statistics imply that H3 shares more derived alleles with H1 than with H2, whereas positive D-statistics denote that H3 shares more derived alleles with H2 than with H1. Tests are significant if the accompanying Z-score > |3|.

The D-statistics revealed that the four samples in the RU cluster exhibited a significantly stronger link to CE than to EA and NE individuals, as expected. However, they were equally distant to CE and to S (Figure 8). Moreover, they do not show differential levels of affinity to samples from different geographic origins within S, i.e. they are equally distant from Zimbabwean, South African or Botswanan individuals (Figure 8).

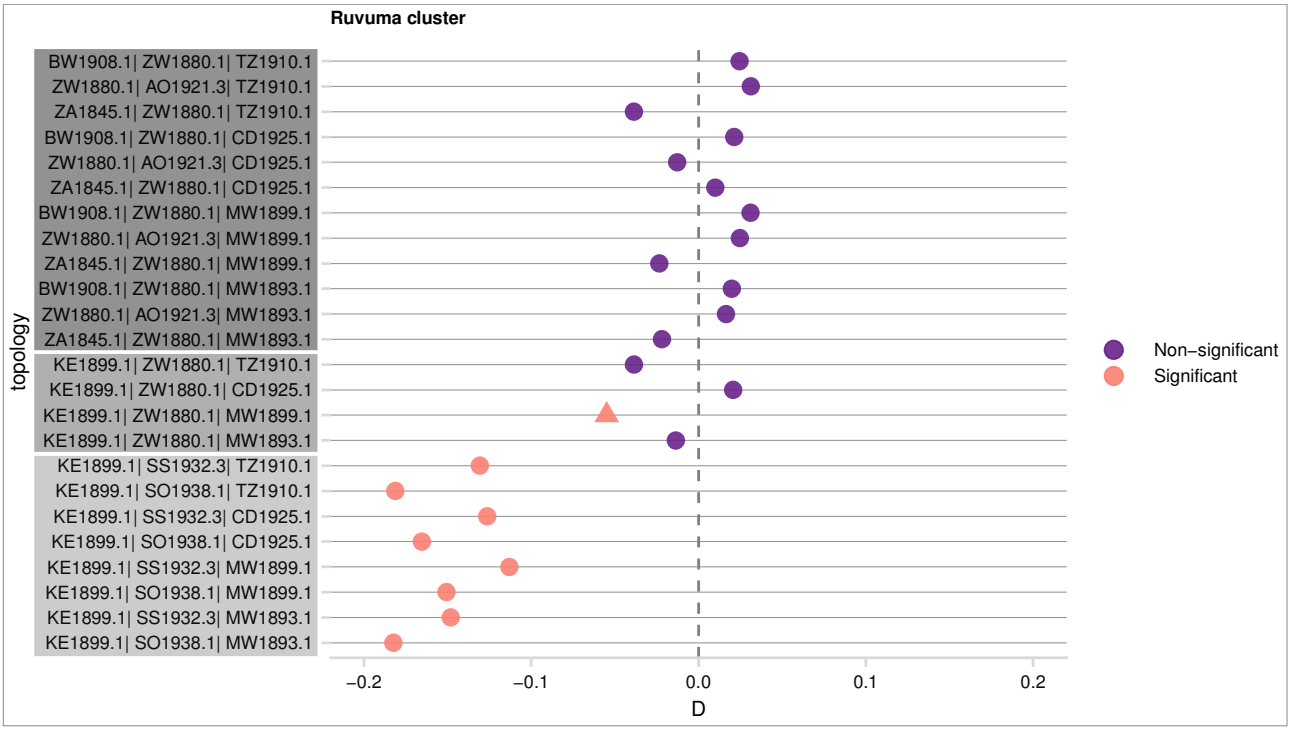


Figure 8. *D-statistics to detect the strongest affinities of the Ruvuma (RU) group*. On the *y* axis, tree topologies are written as H1 | H2 | H3, where H3 is one of the four members of the RU group in our dataset. In the H1 and H2 positions lie samples from two other groups in each test. The outgroup is a white rhinoceros in all cases. The *x* axis represents the value of the D-statistic; legend colour indicates the significance of the test (when Z-score >|3|). Different shades of grey indicate sets of D-statistics that support a particular conclusion about the affinities of RU, i.e. the lightest block shows that RU is significantly closer to CE than to EA or NE; the medium block suggests that RU is not significantly closer to CE than to S; the darkest block indicates that RU is not differentially closer to some S samples than to others. The triangular datum is only marginally significant (Z-score = -3.10).

To explore range-wide gene flow (or barriers to it) we computed effective migration rates with EEMS [13] based on genomic data from the 54 georeferenced historical samples. The resulting effective migration surface pinpoints areas where genetic differentiation decays quickly with distance (higher than average effective migration), versus areas where genetic differentiation remains high even in relatively close geographical space (lower than average effective migration). Importantly, low effective migration might be due to an actual barrier to gene flow, or to low population density in the area [13].

We first observed a major area of low effective migration across the tropical rainforest region, which was expected given that this area was not part of the range of distribution of the black rhinoceros. This area of low effective migration, however, extended toward the east to overlap the course of the Zambezi river. Below this long stretch, to the west lay an area of high effective migration (Namibia, Angola, Botswana, Zimbabwe), while Southeastern South Africa appears as an area of low migration (Figure 9). This contrast between Western and Eastern Southern Africa is surprising, given how similar all southern samples appeared in structure analyses. It must be cautioned, though, that the only historical georeferenced individual from Southeastern South Africa (ZA1775.1) featured the lowest depth of coverage in our whole-genome dataset (1.27x) (Table S1), therefore missing data might have artifactually inflated genomic differentiation.

Another, smaller area of low effective migration traversed Southern Kenya and Uganda, and Northern Tanzania, separating two areas of high effective migration: one to the south (Southern Kenya, Tanzania and Northern Mozambique) and one to the north (Northern Kenya, Uganda and South Sudan). Lastly, the northeastern (Somalia, Ethiopia) and northwestern (Chad, Nigeria) extremes of the historical range appeared to be two separate regions of high effective migration, without an apparent connection between them (Figure 9).

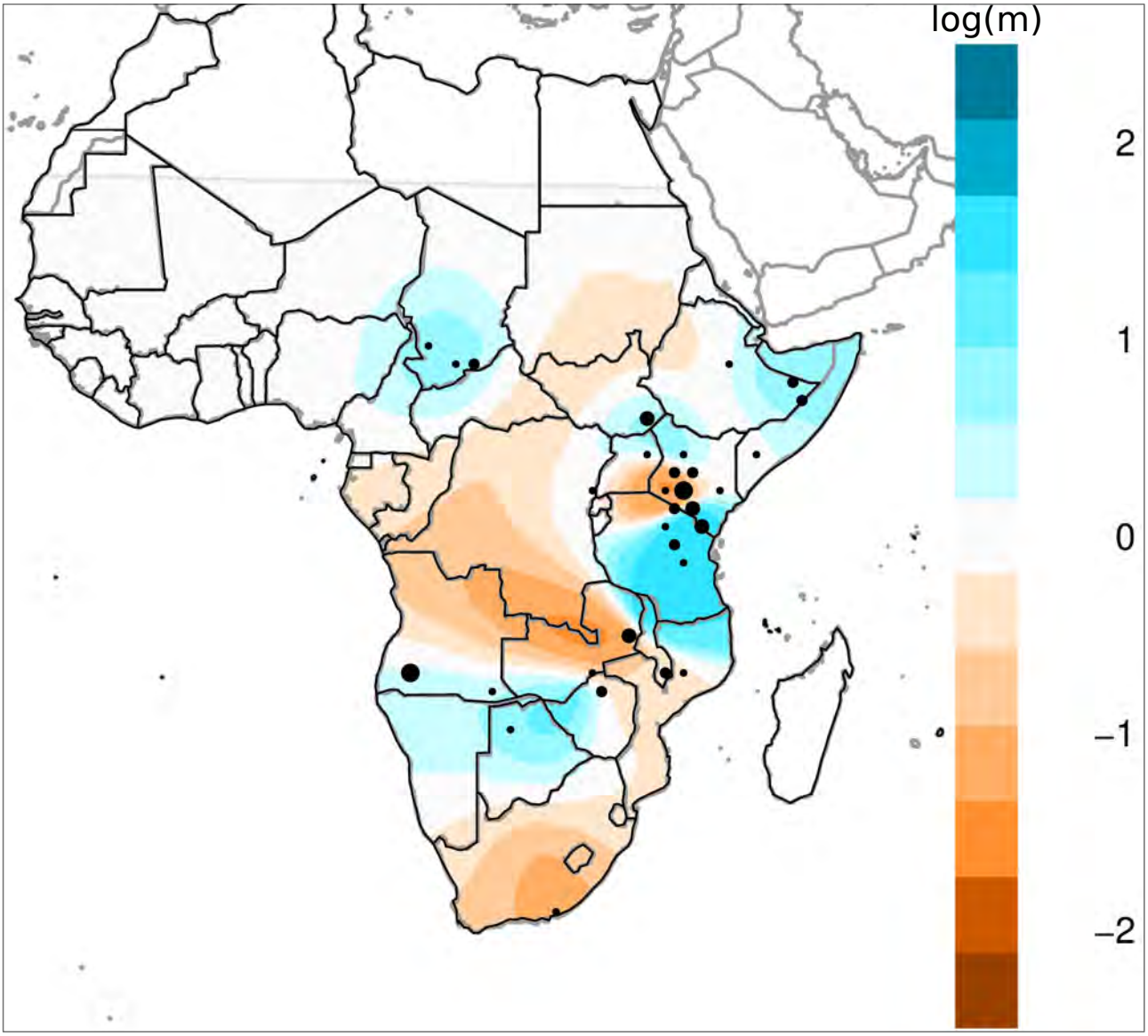


Figure 9. *Effective migration across the historical range of the black rhinoceros*. The effective migration surface was inferred with EEMS [13] based on genome-wide data from 54 georeferenced historical samples. The colour gradient represents effective migration rates in logarithmic scale; blue shades indicate rates higher than average, while brown colours represent rates lower than average.

Individual genomic diversity decayed toward the extremes of the range of distribution

We next scrutinised whether patterns of individual genomic diversity varied across geography. In this regard, we first estimated the genome-wide heterozygosity per sample (GWhet) by means of the individual folded SFS (site frequency spectrum) for transversions only, and then applied a correction for depth of coverage (see Figure S6 and *Individual metrics of genomic diversity* in Methods). The obtained estimates were sorted into the geographically-informed genomic clusters for visualisation. Among the historical samples, GWhet was highest in CE (median_GWhet$_{CE}$ = 3.19 x 10^{-4}) and EA (median_GWhet$_{EA}$ = 3.17 x 10^{-4}), and lowest in S (median_GWhet$_{S}$ = 2.23 x 10^{-4}) (Table 2). Overall, GWhet varied along a radial geographical pattern, with maximum values in

the centre of the range, and decreasing values toward the periphery (Figure 10A-B).

Among the modern individuals, the Kenyan samples (modern-CE) showed lower GWhet (median_GWhet$_{\text{modern CE}}$ = 2.54 x 10^{-4}) than their presumed historical sources, EA and CE. However their levels still fell within the range of historical ones, being comparable to NE and RU. Conversely, the modern individuals from South Africa and Namibia featured remarkably low GWhet, below all historical groups (Figure 10A).

To explore what the levels of inbreeding might have been across groups, we repurposed the same per-individual SFS employed in GWhet estimation, to calculate local estimates of heterozygosity (for transversions only) in sliding windows of 1 Mbp, with a 0.5 Mbp slide, without correction for depth. Based on the per-sample distribution of local heterozygosity, we chose the average 0.1 quantile of local heterozygosity across samples (7.2 x 10^{-5}) as a threshold to flag runs of homozygosity (RoH) (Figure S7). We thus defined RoH as regions of the genome where ≥2 adjacent windows showed a heterozygosity estimate below the chosen cutoff.

Although median values of RoH length were very similar across groups (Table 2), we noted that the longest RoH recorded among the historical black rhinoceroses occurred in the NW (max_RoH_length$_{\text{NW}}$ = 46.5 Mbp) and in the NE (max_RoH_length$_{\text{NE}}$ = 41.5 Mbp), followed by S (max_RoH_length$_{\text{S}}$ = 36.0 Mbp) (Table 2). This agrees with the decline of individual diversity toward the extremes of the range of distribution seen in GWhet.

We then calculated an inbreeding coefficient (F_{RoH}) as the fraction of genomic length in RoH over the total length considered. In this case, RU showed the highest median coefficient (median_F_{RoH} = 0.185), and CE the lowest (median_F_{RoH} = 6.19 x 10^{-2}) (Table 2). Although not as sharp, the specular image of the GWhet pattern occurred for F_{RoH}: the geographic extremes of the range of distribution showed higher values of this coefficient, while the groups in the centre displayed lower values (Figure 10C-D). Interestingly, the modern-CE samples featured the lowest level of estimated inbreeding across all groups (median_F_{RoH} = 6.43 x 10^{-2}), while the modern-S individuals showed the highest (median_F_{RoH} = 0.2843) (Table 2).

	median GWhet	median RoH length (Mbp)	max RoH length (Mbp)	median F_{RoH}
NW	0.000245	2.00	46.5	0.2079
NE	0.000280	2.50	41.5	0.1212
EA	0.000317	2.00	23.0	0.0736
CE	0.000319	2.00	30.5	0.0739
RU	0.000263	2.00	20.0	0.2034
S	0.000224	2.00	36.0	0.1665
modern-CE	0.000254	2.00	29.0	0.0643
modern-S	0.000125	2.25	32.0	0.2843

Table 2. *Summary of individual metrics of genomic diversity across groups*. For each group, median values of individual genome-wide (GW) heterozygosity, RoH length and F_{RoH} are shown, as well as the maximum length of RoH recorded.

We caution that the low depth of coverage (1.27x) of the oldest sample in the dataset (ZA1775.1, dating to 1775), yielded a negative value of GWhet after correction for depth, highlighting the lack of sensitivity of our correcting approach at depths below 2x. Moreover, almost certainly as a result of this missing data, local estimates of heterozygosity were zero across many of its windows, thus inflating the number and length of RoH. This sample was therefore not displayed in Figure 10, nor accounted for in the calculation of summary statistics of GWhet, RoH length and F_{RoH}. As for other potential effects of low depth of coverage, since the local estimates of heterozygosity were not corrected, the four samples that appear as outliers of F_{RoH} in the CE, RU and S groups coincide with those of lowest depth after ZA1775.1 (3.7x in TZ1909.1, 4.42x in un1876.1, 4.03x in CD1925.1 and 3.27x in TZ-un.1).

The areas occupied by EA and CE seemed to have been the hotspots of black rhinoceros diversity, whereas individual diversity decayed with increasing latitude both northward (groups NW and NE) and southward (group S). Unfortunately, as the reduced and uneven sample sizes across groups were not amenable to robust statistical testing, we caution that our observations should be viewed as merely a qualitative assessment.

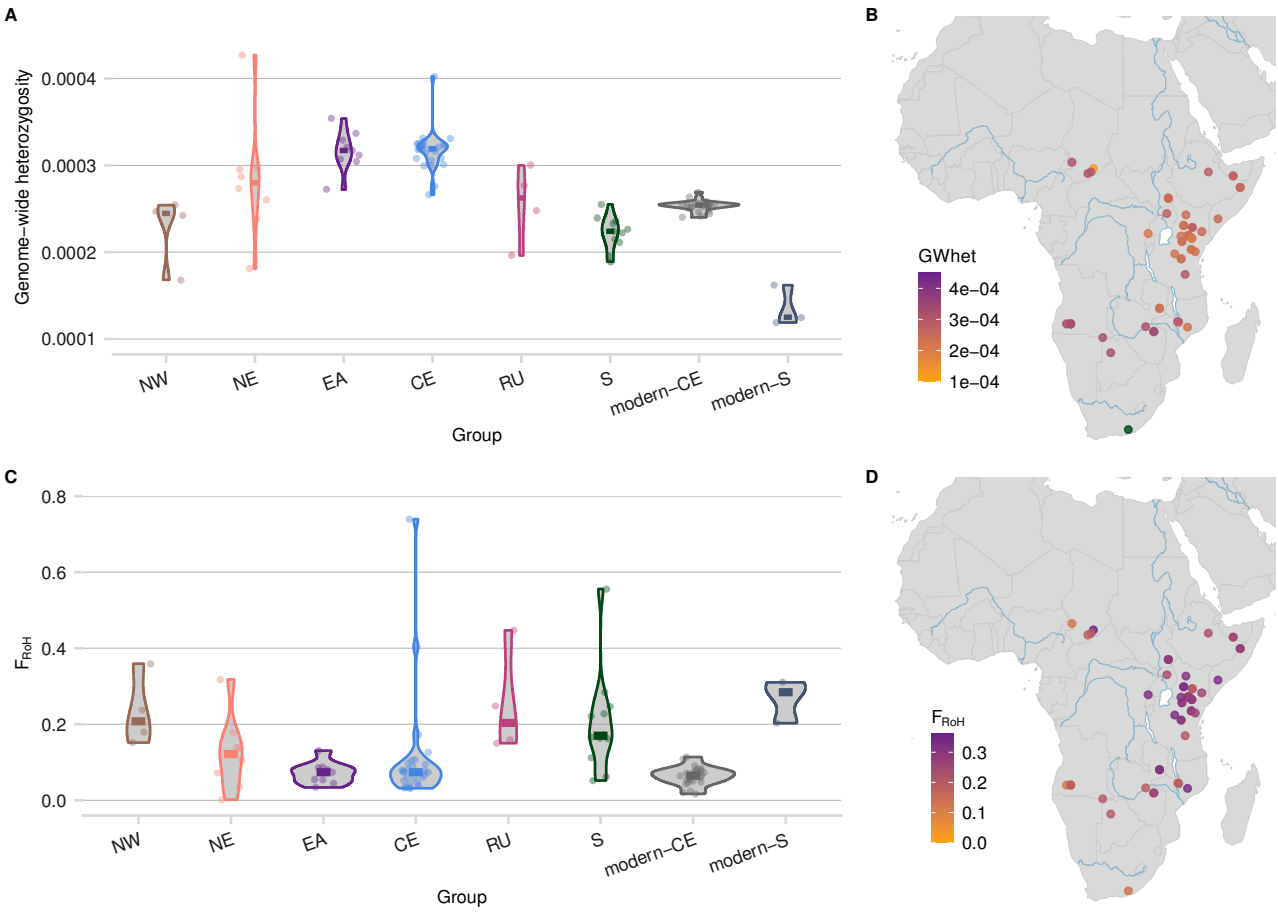


Figure 10. *Individual genomic diversity across geographically-informed genomic groups of black rhinoceros*. A) Individual genome-wide heterozygosity (GWhet) is shown as its distribution within each group, and in B) as its geographical distribution for 53 georeferenced historical samples. C) Distribution of individual F_{RoH} values per group is visualised. D) The geographic distribution of F_{RoH} is shown for 53 historical, georeferenced samples. In A) and C), horizontal bars within violins indicate the median.

Discussion

The aim of this study was to characterise the population structure, phylogeny and diversity in the black rhinoceros before its range-wide collapse in the last decades of the 20th century. For that purpose, we generated DNA sequence data from 73 historical samples from museum specimens representing most of the range of distribution of the species. Additionally, we also generated data from 27 individuals from extant populations. Mitochondrial genomes were retrieved from all 100 historical and modern samples, and more relevantly, we could also produce whole-genome data for 64 of the historical samples and for the 27 modern individuals. In the following paragraphs we put in context the main observations we extracted from this rich dataset.

A high resolution portrait of the historical black rhinoceros

The first study of the species-level structure of the black rhinoceros with molecular data and temporal sampling [8] did so with the mitochondrial control region and nuclear microsatellites as markers. We corroborated their findings regarding the maternal population and phylogenetic structure using full mitochondrial genomes, proving that the control region is a representative marker of the entire mitochondrial locus in the case of the black rhinoceros.

At the nuclear level, however, our genome-wide survey of polymorphism produced better resolved patterns than those seen with microsatellites [8]. In brief, our analyses indicated that the black rhinoceros was structured according to a latitudinal gradient in historical times, but with certain discontinuities that strengthened genomic differentiation in some parts of the range. More specifically, the Zambezi river seems to have acted as a barrier, isolating the populations to the south and west of its course (Angola, Botswana, Namibia, South Africa) from the rest of the populations. Individuals from the meridional end of the range showed little substructure, hence why we grouped them under the same label, S, and belonged to an early diverging branch in the black rhinoceros phylogeny. We detected, nonetheless, some weak differentiation between the west (Western Angola and Namibia) and the east (South Africa and Zimbabwe).

North of the Zambezi river, five more populations were identified and named as follows: Northwest (NW), Northeast (NE), East Africa (EA), Central Africa (CE) and Ruvuma (RU). Genomic distances among them mirrored geographical distances. The groups NW from Chad and Nigeria, and NE, from Somalia and Ethiopia, are most closely related, and NE in turn is also close to EA, that inhabited South Sudan, Uganda and Northern Kenya. A discontinuity traversed Eastern Kenya and Northern Uganda, possibly driven by the high altitude around the Kenyan Highlands, where population density might have been low. The CE group lay on the southern side of this barrier,

spanning Southern Kenya, Tanzania, Zambia and Mozambique. Interestingly, the populations to the north of the Kenyan discontinuity, NW, NE and EA, are monophyletic clades nested within the CE branches in the phylogeny of these black rhinoceroses.

The last group, RU, from Southern Tanzania and Malawi, seems to represent an instance of secondary contact over the evolutionary history of the black rhinoceros. Geographically their range overlaps with CE, but phylogenomically they are closer to S. Our analyses suggest that RU originated along with the S group, but because they remained on the northern bank of the Zambezi river after it was established as a barrier, extensive introgression from CE occurred.

Two waves of expansion in the evolutionary history of the black rhinoceros

This portrait, if accurate, would imply a story where two major waves of expansion occurred: one toward the south, followed by isolation, and a more recent one toward the north, giving rise in successive founder effects to the populations of the northern periphery of the range. Specifically, we speculate that an ancestral black rhinoceros population in Central Africa would have expanded toward the south causing a split into a proto-S and a proto-CE populations. Proto-S became isolated probably owing to a reconfiguration of the drainage system of the Zambezi river in the late Pleistocene [14]. On the other hand, proto-CE expanded all across Central and East Africa, and it founded NW in the far west, and EA which, in turn, expanded northeastward and founded NE. Once established, CE entered into secondary contact with the proto-S individuals that remained on the north of the Zambezi river, giving rise to RU. Of course, further analyses would be required to confirm the veracity of this story and to discern the timing of these events.

Paleontological evidence suggests that the species could have indeed originated in Central Africa [15], and in [8] the highest levels of genetic diversity are found in that region. Our assessment of geographical distribution of genomic diversity also supported this view, since the highest levels of individual-level diversity were recorded in CE, while the lowest occured at the geographical extremes of the range (within NW, NE and S). A long history of isolation in the case of S, and relatively recent founder effects in the case of NW and NE would be concordant with the gradient of diversity observed.

The relationship of modern to historical black rhinoceros populations

Disentangling the range-wide structure and phylogeography in the black rhinoceros was only feasible owing to the availability of a representative historical sample set. Modern populations of black rhinoceros represent the surviving fragments of a much richer fabric of past diversity.

Furthermore, current populations are often the product of management practices, including translocations of individuals from different geographical and genomic origins [1]. Nonetheless, having samples from modern populations allowed us to verify that our historical dataset draws a genomic-geographic map of the black rhinoceros that is complete enough to trace the likely origin of modern individuals.

Among the modern samples, those from the reserves NNP and OP in Kenya show signs of being admixed individuals whose genetic ancestry draws from the historical EA and CE, while those from MA are more similar among them and clearly from a single genomic origin within CE. Interestingly, this modern-CE group shows lower levels of genome-wide heterozygosity than their historical sources, but also lower inbreeding estimates than any of the historical groups. This might reflect a background of management practice that attempts to avoid inbreeding and maximise diversity. The former is achieved by hybridising animals from different stock populations, but the latter might be limited by the effect of the recent bottlenecks and the drift in small-sized populations, which erode genomic diversity.

The modern Southern African individuals in our dataset, on the other hand, feature the lowest heterozygosity and highest inbreeding across all groups. We acknowledge, however, that this small sample (n = 3) may well be far from representative of their source populations, thus future studies including more individuals should help validate these conclusions.

Cautionary notes on potential sources of bias

Throughout the analyses based on genome-wide data, we flagged a few samples whose low depth of coverage (ZA1775.1, TZ1909.1, un1876.1, CD1925.1 and TZ-un.1) might undermine the robustness of the results observed for them, especially regarding the individual measures of genomic diversity, where they stand out as outliers. Additionally, in a test of cross-contamination among samples based on mitochondrial sequence reads (see *Cross-contamination assessment* in Methods), we identified four samples whose authentic content fell below 95%: TD1925.1 (77.5%), KE1902.2 (0.14%), KE1909.1 (89.6%), NNP2 (94.7%). This needs further assessment to confirm that these samples are contaminated and should therefore be excluded, but at least one of them, TD1925.1, showed an odd position in structure analyses, since it originates from Chad but appears genomically much closer to NE than to NW.

We also caution that a particular source of bias might have certain influence over all analyses: reference bias. To maximise breadth and depth of coverage of the raw sequence data from museum samples (which are ancient DNA samples after all), we chose to align them against a black

rhinoceros whole-genome assembly ([10] in press), which derives from an individual from South Africa. This might affect the number of variable sites found across samples from different populations, an effect to which estimates of individual genomic diversity might be particularly sensitive [16].

Nonetheless, reference bias probably has a substantial effect only in analyses that rely on low-frequency variable sites [17]; it is likely that the broad species-level patterns we systematically obtained with different analytical approaches are true biological signals. Whether reference bias is distorting some of our findings, and if so to what extent, will require that the raw data is aligned against the reference assembly of a closely related species (i.e. the white rhinoceros), at the cost of lowering the depth of coverage across samples, and probably sacrificing those of lowest endogenous content.

The relevance of historical population genomics for the study and management of the black rhinoceros

Despite those potential biases, the dataset generated in this study still greatly increased the resolution of our insight on the status of the black rhinoceros in the recent past. The results hereby presented lay the ground for further investigation into the ecology and evolutionary history of the species, but moreover they are a meaningful source of information for the conservation of the species.

This work could help tackle the question of the subspecies-level taxonomy in the black rhinoceros, a matter of heated debate that has strongly influenced management practice [4]. From the 1980s until very recently, management was guided by a classification into four ecotypes also used as subspecies, while several taxonomic revisions since the 1960s have named as many as 17 subspecies [18], with the most recent one contemplating eight subspecies, four of which are extinct [19]. The IUCN African Rhino Specialist Group updated the list of management units in their latest 2020 report, in light of the discoveries of the first temporal, range-wide genetic survey of the black rhinoceros [8]. Based mostly on mitochondrial population structure, the following units are now in official use: *EA*, *CE* and *CV* in East Africa, *SW* in Namibia and *SE-SN* in South Africa and Zimbabwe [1].

In brief, the current management units and our six historical populations compare as follows. The historical populations NW and NE in our dataset are extinct, while EA would be equivalent to the homonymous management unit, and CE would include the units *CE* and *CV*. The population RU is probably extinct, although Tanzanian authorities claim that a few individuals remain in Selous

Game Reserve (Yoshan Moodley, personal communication). Lastly, the population S in our dataset would encompass both *SW* and *SE-SN* management units.

The correspondence of units across the different taxonomic and management classifications are not always straightforward, but having historical population genomics data at hand has shed light on important questions. Firstly, we have corroborated the losses of some local pools of historical genomic diversity (NW and NE). Secondly, our results emphasise that EA and CE should be managed separately while all southern units could be mixed if needed (e.g. to rescue very small populations). Thirdly, if some remnants of the RU population still exist, preserving this pot of diversity should be a high-ranking priority in black rhinoceros conservation. Overall, our results support and highlight the importance of updating management strategies with genetic/genomic, range-wide information.

Beyond these conclusions, having genome-wide data available opens promising new avenues for conservation-related research on the black rhinoceros. Our map of black rhinoceros genomic diversity could be leveraged to develop molecular tools to identify the provenance of black rhinoceros material seized from the black market. Also, with genomic information we could venture into the potential phenotypic effects of the intra-species diversity observed in order to guide management actions. For instance, gaining insight into inbreeding and outbreeding depression might greatly enhance the success of breeding programs. In addition, it seems plausible that populations of black rhinoceros could have been locally adapted, given the wide variety of habitats where they thrived. Putative locally adapted genotypes and their phenotypic effects could be identified with genomic tools. Since translocation of animals is relatively common, this information might turn invaluable for the future management of the black rhinoceros.

Acknowledgements

This work was supported by ERC Consolidator Grant 681396 ‘Extinction Genomics’ to M.T.P.G. and by EMBO Short-Term Fellowship 7578 to F.S.B. The authors would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure (NGI), Sweden, the Knut and Alice Wallenberg Foundation and UPPMAX for providing assistance in massively parallel DNA sequencing and computational infrastructure.

The authors are very grateful to all the museums who contributed samples to this study: the Natural History Museum London, the Museum of Natural History Berlin, the Powell-Cotton Museum, the Natural History Museum Vienna, the Natural History Museum at the National Museum Praha, the Natural History Museum Of Zimbabwe, the Swedish Museum of Natural History, the Royal Museum for Central Africa Tervuren, the Senckenberg Museum Frankfurt, the United States National Museum Washington (or Smithsonian Arts and Industries Building) and the Bavarian State Collection of Zoology.

Lastly, the authors would like to thank David Duchene Garzon for guidance regarding the phylogenomic analysis.

Methods

Data generation procedures

Manipulation of sampled tissue

Our historical sample collection included material obtained from 73 museum specimens. Collection dates ranged between 1775 and 1981. All samples consisted of keratinous material (pieces of skin, horn powder or hairs), except for one, ZA1845.1 which was a piece of bone from a skull. Samples from museum specimens were stored and processed in facilities dedicated to ancient DNA work at the Swedish Museum of Natural History (Stockholm), and the Natural History Museum of Denmark (Copenhagen).

As detailed in Chapter 2 of the present thesis [20], “skin samples were manually cut with disposable scalpels in order to generate 20-80 mg of fragmented tissue. Dry skin tissue is highly absorbent, therefore the biggest pieces of material were hydrated for 2-3 hours at 4 ºC by adding 0.5-1 mL of molecular biology grade water to each tube. Water was discarded, and the tissue was briefly washed with 0.5 mL of a 1% bleach solution, followed by two washes with molecular biology grade water. Bone material was crushed with a small hammer, and small pieces amounting to 150-200 mg were used for extraction”.

Our collection also included 27 modern samples in the form of keratinous material either preserved in ethanol or dry. Dry samples were hydrated prior to manipulation, and then each piece of skin was cut with a disposable scalpel. For extraction, 20 mg of material were used. Molecular work for 24 of these samples took place at the modern DNA facilities of the Natural History Museum of Denmark, while for the remaining three, it took place at the Swedish Museum for Natural History.

DNA extraction, library preparation and sequencing

The laboratory procedures followed to generate genomic data from black rhinoceros museum samples were the ones specified in Chapter 2 of this thesis [20] (see *DNA extraction, library preparation and sequencing* in Methods).

With regards to the modern samples, extraction of DNA was done with the KingFisher™ Duo Prime instrument and its associated Cell and Tissue DNA Kit, following the manufacturer’s guidelines. Concentration of DNA extracts was measured with a Thermo Scientific™ Qubit dsDNA high-sensitivity (HS) assay. A 20 µL aliquot of each extract was fragmented in a Covaris® focused-ultrasonicator with a customised program to reduce fragment length to 400 bp. Size distribution upon fragmentation was assessed with a TapeStation 2200 (Agilent, Santa Clara, CA). For 24 out of

27 samples, library build, qPCR and indexing amplification were performed following the same procedures as for historical samples (detailed in Chapter 2 of this thesis [20]). Resulting indexed libraries were distributed in pools containing equimolar proportions of eight indexed libraries each. Each of these pools was given one lane of BGISEQ 500 PE150 sequencing.

For samples NA1, NA2 and ZA1, however, sequencing libraries were built using the Illumina® TruSeq® Nano DNA Library Prep Kit for NeoPrep™ on DNA inserts that were 350 bp in average fragment length and following the manufacturer's guidelines. Libraries were then sequenced on an Illumina® HiSeq X platform, giving 0.5 lanes per sample in PE150 mode.

Bioinformatic data processing and quality assessment

Quality check and mapping of DNA sequence data

We generated shotgun sequencing data for a total of 101 newly re-sequenced black rhinoceros samples. Of these, 74 belonged to museum specimens, and 27 to modern samples. We followed the same procedures of initial data processing as in Chapter 2 of this dissertation [20] (see *Quality control, trimming and mapping of DNA sequence data* in Methods), including a quality check per sample with fastqc v0.11.7 [21], and the pipeline PALEOMIX 1.2.13.2 [22] with the same settings.

Each sample was aligned separately against two reference sequences: a) the publicly available mitochondrial genome of *Diceros bicornis* (NC_012682 [9]), and b) an unpublished whole-genome assembly of *Diceros bicornis* of 2.33 Gb of size and N50 = 6.0 Kb ([10] in press). Details on the assembly procedure of the whole-genome reference are not included here. As a result we generated two alignment files per sample, one against the mitochondrion and one against the whole genome.

Filtered sample sets

Of the total 101 sequenced samples, some were excluded from all or some analyses based on the following criteria. For analyses based on mitochondrial data, TZ1925.2 was discarded based on its comparatively low depth of coverage (1.04x) on the mitochondrial genome, so the final sample set included 100 individuals, 73 historical and 27 modern.

Regarding the whole-genome dataset, from the 101 sequenced, 10 samples were discarded due to low depth of coverage (<1x), systematic failure to align against the whole-genome assembly or suspected identical specimen of origin (see *Relatedness test* in Methods and Supplementary material and Figure S3). This left a final 91 samples with whole-genome data; of those, 64 were

historical and 27 were modern.

Among the modern samples, however, some show signals of relatedness (see *Relatedness test* in Methods and Supplementary material and Figure S3), so for certain analyses the sample set had to be further reduced: of 27 total modern samples, seven were ignored, leaving a modern set of 20 samples. For PCA, UMAP and admixture analyses, these were the sample sets used: 64 historical individuals, 20 modern unrelated individuals, and the combination of both including 84 samples.

Generation of sequence consensus files

A consensus sequence file in *.fasta* format was generated for each of the *.bam* files with ANGSD v 0.921 [23]. In the case of the data mapped to the mitochondrial reference, the consensus base per site was chosen, (option -doFasta 2), while for whole-genome mapped data, a random base per site was drawn (option -doFasta 1). The following quality filtering parameters were used: -minQ 20 -minMapQ 30 -baq 1 -C 50. Parameter baq refers to per-Base Alignment Quality described in [24].

Cross-contamination assessment

We evaluated whether there was a signal of cross-contamination among samples in the dataset with ContamMix, described in [25], and broadly used in human palaeogenomic studies. This method is based on the consensus mitochondrial sequences of ancient samples to detect sequences from a possible source of contamination, but the contamination levels should not be >50% for this method to work. ContamMix assumes that the sequenced mtDNA reads originate from a mixture of the reconstructed consensus sequence and other black rhinoceros whole mitochondrial genomes, representing possible contaminant sequences. By comparing the mapping affinity of each ancient mtDNA read to the reconstructed consensus sequence and to the other 99 possible contaminants (n = 100 samples tested), ContamMix reports the fraction of non-endogenous reads as the contamination rate with a Bayesian estimate of the posterior probability of the contamination proportion.

We used the data mapped against the mitochondrial reference genome (see *Quality check and mapping of DNA sequence data*) to construct the endogenous mtDNA consensus sequence of the historical samples required for ContamMix. We used the "bcftools mpileup | call" pipeline for calling genotypes and the "bcftools consensus" command for creating the mtDNA consensus sequences. Sites with base quality <20 and reads with mapping quality <30 were discarded. Only SNPs and sites with at least 10x depth of coverage were considered for consensus calling. The resulting estimates showed a fraction of authentic content above 95% for all samples but five: TD1925.1 (0.7751), KE1902.2 (0.0014), KE1909.1 (0.8961), NNP2 (0.9472).

Choice of scaffolds for variant site finding

To limit computational memory usage, we restricted statistical analysis to scaffolds >14 Mbp in the black rhinoceros reference assembly ([10] in press). The 47 scaffolds above this cutoff represent 72.83% of the total length of the assembly, and were subject to an analysis of normalised depth of coverage across samples to assess if they belonged to the sex chromosomes. Scaffolds showing a 0.5x normalised depth in male samples can be assumed to belong to the X chromosome. However none of the 47 scaffolds under inspection displayed that pattern (Figure S2), and therefore all were used for finding informative sites.

Computing genotype likelihoods

We identified putative single-nucleotide polymorphisms (SNPs) and calculated genotype likelihoods using ANGSD v 0.921 [23] across the chosen 47 scaffolds. Different panels of genotype likelihoods were computed for different sets of samples: only historical samples (n = 64), only modern unrelated samples (n = 20) (see *Relatedness test* in Methods and Supplementary material, and Figure S3), and both combined (n = 84).

In every case, the minimum number of individuals in which a variant site must be present (-minInd) was 95%. Minimum and maximum global depth per site were based on a global depth assessment with ANGSD -doDepth: 500 and 1500 respectively, except for the panel containing only modern samples where the minimum depth was lowered to 200. Additionally, we included the following quality filtering and output choice parameters: -remove_bads 1 -uniqueOnly 1 -baq 1 -C 50 -minMapQ 30 -minQ 20 -doCounts 1 -GL 1 -doGlf 2 -doMajorMinor 1 -doMaf 1 -doHWE 1 -dosnpstat 1 -HWE_pval 1e-2 -SNP_pval 1e-6

Transitions were removed *a posteriori* from genotype likelihoods files with custom-made code. Numbers of retrieved transversions were the following: n_{tv}*64historical* = 2,540,487 sites; n_{tv}*20modern* = 925,317 sites; n_{tv}*84historical-modern* = 2,715,503 sites.

Statistical analyses of genomic data

Mitochondrial haplotype network and phylogenetic analysis

Input data for mitochondrial analyses consisted of 100 black rhinoceros mitochondrial genomes (in *.fasta* format) aligned with MAFFT v 7.453 [26]. First, an intra-species phylogenetic network analysis was conducted with PopART [27] using the Median Joining Network algorithm.

A dated phylogeny was computed in a Bayesian coalescent framework with BEAST2 v 2.6 [28] including 100 black rhinoceroses plus two white rhinoceroses: the sample of northern white rhinoceros UG1905.2 in Chapter 2 of this thesis [20], and a southern white rhinoceros represented by the mitochondrial sequence within the publicly available white rhinoceros whole-genome assembly (GCA_000283155.1 with accession code PRJNA74583). As a calibration time point we chose the estimated split between white and black rhinoceroses (6 ± 0.5 Mya) [29]. Nucleotide substitution model choice was determined with jModelTest v 2.1.10 [30], where the highest-ranking model was HKY. Lineage coalescence model was the default Yule model; a strict clock was chosen as branch rate model, and no partitions of the data were specified. The MCMC chain iterated 10 M times, and branch support was assessed with 100 iterations of bootstrap resampling. With Treeannotator (within the BEAST2 package), a consensus tree was produced, and then visualised with FigTree v 1.4 [31] (the dated phylogeny in Figure S5) and the online software iTOL [32] (the cladogram in Figure 2).

Relatedness test

We ran a pairwise analysis of relatedness based on genotype likelihoods of 837,527 transversion sites across 92 individuals with ngsRelate v2 [33]. The computation of this panel of genotype likelihoods followed the procedure detailed in *Computing genotype likelihoods* except for the parameters -setMaxDepth and -setMinDepth, which were set to 900 and 200 respectively. The *.beagle* and *.mafs* files were reformatted with custom code to match the input requirements of ngsRelate v2. The output of ngsRelate v2 was visualised following the approach described in [34]. The degree of relatedness between each pair of samples was assessed visually based on the relative values of coefficient of relatedness R1 versus coefficients KING and R0.

We found one extreme outlier pair (SO1896.1-SO1896.2) indicating that these two samples were likely from the same specimen; SO1896.1 was removed from further analyses. Besides, we found 11 pairs of individuals showing a relatedness signal, all among the modern samples. To avoid all pairwise relatedness, seven of those samples were excluded in analyses sensitive to this bias (i.e. PCA, UMAP, admixture): MA1, MA2, MA5, MA7, OP10, OP11, ZA2 (Figure S3). As a criterion to exclude samples from a related pair, the sample of lowest depth of coverage was discarded.

Principal Component Analysis (PCA)

We used PCangsd v 0.973 [35] to compute covariance matrices on genotype likelihoods files including only transversions from across the 47 largest scaffolds of the assembly (see *Choice of scaffolds for variant site finding*). This was done for both historical and modern samples combined (n = 84), and separately after removal of related individuals ($n_{historical}$ = 64, n_{modern} = 20; see *Filtered sample sets*). Standard packages in R v 3.4.4 [36] were used for decomposition of each matrix in eigenvectors and eigenvalues, and *ggplot2* [37] for visualisation of principal components (PCs).

Dimensionality reduction with UMAP

To capture more of the variability in the data, we used a non-linear method to reduce dimensionality called UMAP [11]. We converted genotype likelihoods from across the 47 largest scaffolds of the assembly (see *Choice of scaffolds for variant site finding*) from *.beagle.gz* format into an array of samples (rows) by sites (columns) with custom code. We then reduced dimensionality with the UMAP *fit_transform* function, specifying a number of nearest neighbours equal to 18.75% of the sample size in each case, i.e. 12 for the set of 64 historical samples, and 16 for the set of 84 historical and modern samples combined. Code and command lines used were modified from the UMAP online tutorial on the iris dataset [38]. The resulting embedding (two-dimensional output) was visualised using *ggplot2* [37] in R v 3.4.4 [36].

Admixture

Assessment of admixture proportions in our dataset was done by means of ngsAdmix v 32 [39]. Genotype likelihoods of transversion sites across the 47 largest scaffolds of the assembly (see *Choice of scaffolds for variant site finding*) were given as input for a sample set of 84 historical and modern unrelated samples. Values of ancestral clusters, K, ranged between two and seven, and for each value of K, we ran ngsAdmix 50 times. For each value of K, the run of highest log-likelihood was chosen for visualisation with the software pong [40].

D-statistics

We used ANGSD v 0.921 [23] to run D-statistics with the option -doAbbababa 1 (sampling a random base at each site) for all triplets of samples among our 64 historical individuals with sites from the 47 largest scaffolds of the assembly (see *Choice of scaffolds for variant site finding*). We used as an outgroup a re-sequenced white rhinoceros from our own database (sample ZA2012.1 in Chapter 2 of this thesis [20]), aligned against the black rhinoceros reference assembly. In the ANGSD command line, the following quality filtering parameters were also included:

-remove_bads 1 -uniqueOnly 1 -baq 1 -C 50 -minMapQ 30 -minQ 20, and only transversion sites were considered with the flag -rmTrans 1. The D-statistic and Z-score per topology were calculated based on the resulting counts of derived alleles with the script jackKnife.R included in ANGSD. Tests were considered significant if Z > |3|.

Estimates of gene flow with EEMS

To infer potential gene flow among historical populations of black rhinoceros, we used EEMS [13], a tool that links genetic and geographic data to estimate effective migration surfaces. Of the 64 re-sequenced genomes in our dataset, we used for this purpose the 54 that were georeferenced.

As input, EEMS takes a pairwise distance matrix which we calculated with ANGSD v 0.921 [23] and the options -doIBS 1 (choose a random base per site) and -makeMatrix 1 across the 47 largest scaffolds of the assembly (see *Choice of scaffolds for variant site finding*). The following quality filters and output options were applied: -remove_bads 1 -uniqueOnly 1 -baq 1 -C 50 -minMapQ 30 -minQ 20 -minInd 51 -setMaxDepth 1500 -setMinDepth 500 -doCounts 1 -GL 1 -doMajorMinor 1 -minFreq 0.01 -doIBS 1 -makeMatrix 1. The maximum missing data per site was set to 5%, i.e. -minInd 95% of samples.

The matrix included 132,280,027 sites and was fed as input to 10 independent runs of EEMS with an MCMC chain of 2 M iterations and assuming 1000 underlying demes (a specification of grid size). The geographic area of interest was outlined by hand with an online tool [41]. Visualisation of the output estimated migration surface was conducted in R v 3.4.4 [36] with code included in EEMS.

Phylogenomic analysis

We used the software IQ-Tree v 1.6.8 [12] to perform a maximum likelihood phylogenomic analysis of 91 black rhinoceros samples, with a white rhinoceros from our own database as an outgroup (sample ZA2012.1 in Chapter 2 of this thesis [20]). First, we picked 100 regions of 200 kbp at random across the 47 biggest scaffolds of the black rhinoceros assembly with BEDTools v 2.27.1 [42]. For each of those regions, a consensus sequence in *fasta* format was generated per sample from the corresponding *.bam* files. Resulting *.fasta* files per region were merged with SeqKit v 0.7.1 [43], and the merged output was provided as input for IQtree.

Each region was considered a different partition [44], and for each partition IQtree was set to find the best fitting nucleotide substitution model with the implemented ModelFinder (-m MFP) [45]. Ultra-fast bootstrap and a SH-aLRT test for branch support were both set to 1000 replicates (-bb

1000 -alrt 1000). The output consensus tree was visualised with the online tool iTol [32].

Individual metrics of genomic diversity

The procedures described here are largely the same as the ones detailed in Chapter 2 of the present thesis [20] (see *Genome-wide heterozygosity*, *Heterozygosity correction* and *Heterozygosity in sliding windows and inbreeding measured as* F_{RoH} of the Methods). We present here a brief summary and the modifications specific to this study.

For each sample we computed the folded site-frequency spectrum (SFS) from the corresponding alignment file (*.bam*) with the ANGSD [23] option -dosaf 1 across the 47 largest scaffolds of the assembly (see *Choice of scaffolds for variant site finding*). The resulting *.saf.idx* files were used as input for calculating global and local estimates of heterozygosity.

Genome-wide heterozygosity was calculated from the actual estimate of the SFS, generated with realSFS [23] in chunks of 1×10^8 sites for transversions only. These per-sample estimates were corrected for depth of coverage by downsampling all *.bam* files corresponding to samples of depth of coverage >14x (n = 19 samples) to 12x, 10x, 8x, 6x, 4x, 2x using SAMtools v 1.9 [46,47]. For each downsampled file we estimated heterozygosity as detailed above. As described in Chapter 2, "For each sample, downsampled estimates of heterozygosity were divided by the estimate at the original depth. In this way we obtained the proportion of the original heterozygosity estimate at each of the decreasing values of depth of coverage. These proportions were then plotted against the depth of coverage, and we observed that estimates of heterozygosity decreased with depth in a non-linear fashion".

Estimates of heterozygosity for all samples were adjusted with the following equation: $gw\text{-}heterozygosity = -0.394 + 0.241 * x - 0.017 * x^2 + 6.1E\text{-}04 * x^3 - 9.36E\text{-}06 * x^4$ where x was replaced by the corresponding value of depth of coverage. The oldest sample in our dataset, ZA1775.1, featured also the lowest depth of coverage (1.27x), its original estimate of heterozygosity is very small, and its corrected estimate is <0. The correcting equation shows low sensitivity to such low values of depth of coverage, thus creating this artefact.

To identify RoH, local estimates of heterozygosity were computed from the same *.saf.idx* files generated with ANGSD -dosaf 1 for each of a total of 3,327 sliding windows of 1 Mbp (0.5 Mbp slide) along the largest 47 scaffolds of the assembly. The output estimates were not corrected by depth of coverage, as it would be computationally very heavy and time-consuming.

The arbitrary cutoff of heterozygosity to declare RoH was set to 7.22×10^{-5}, which is the average 0.1 quantile of local heterozygosity across samples (Figure S7). Custom code was used to identify RoH (*findRoH_v3.py* [48]). As elaborated in Chapter 2 of the present thesis: "windows below the chosen threshold are assigned a value of '1', whereas remaining windows are assigned a '0'. The output of this script is a list of regions of low heterozygosity of different lengths, represented by the number of adjacent windows with a heterozygosity value below the threshold. A RoH was declared for each of those regions if $n_{windows} \geq 2$. The length of RoHs for each sample was calculated in R v 3.4.4 in the following manner: $RoH\_length = n_{windows} * 1\ Mbp - ((n_{windows} - 1) * 0.5\ Mbp)$". The total length for calculating F_{RoH} followed the same equation: i.e. $total\_length = n_{total\_windows} * 1\ Mbp - ((n_{total\_windows} - 1) * 0.5\ Mbp)$.

All visualisations of individual metrics of genomic diversity were done in R v 3.4.4 [36] using standard packages and *ggplot2* [37].

Visualization of maps

Apart from the R v 3.4.4 [36] packages mentioned for graphical representation of various results, visualisation of maps and geographical data required the packages *maps* [49], *mapdata* [50], *maptools* [51], *rgdal* [52] and *sp* [53,54].

References

1. Richard Emslie (IUCN SSC African Rhino Specialist Group) (2020). IUCN Red List of Threatened Species: Black Rhino (IUCN) Available at: https://www.iucnredlist.org/en.

2. Knight, M.H. (2019). African Rhino Specialist Group report. Pachyderm *60*, 14–39.

3. Rookmaaker, K., and Antoine, P.-O. (2012). New maps representing the historical and recent distribution of the African species of rhinoceros: Diceros bicornis, Ceratotherium simum and Ceratotherium cottoni. Pachyderm *52*, 91–96.

4. Rookmaaker, K. (2011). A review of black rhino systematics proposed in Ungulate Taxonomy by Groves and Grubb (2011) and its implications for rhino conservation. Pachyderm *50*, 72–76.

5. Karsten, M., van Vuuren, B.J., Goodman, P., and Barnaud, A. (2011). The history and management of black rhino in KwaZulu-Natal: a population genetic approach to assess the past and guide the future. Anim. Conserv. *14*, 363–370.

6. Muya, S.M., Bruford, M.W., W.-T. Muigai, A., Osiemo, Z.B., Mwachiro, E., Okita-Ouma, B., and Goossens, B. (2011). Substantial molecular variation and low genetic structure in Kenya's black rhinoceros: implications for conservation. Conserv. Genet. *12*, 1575–1588.

7. Anderson-Lederer, R.M., Linklater, W.L., and Ritchie, P.A. (2012). Limited mitochondrial DNA variation within South Africa's black rhino (Diceros bicornis minor) population and implications for management. Afr. J. Ecol. *50*, 404–413.

8. Moodley, Y., Russo, I.-R.M., Dalton, D.L., Kotzé, A., Muya, S., Haubensak, P., Bálint, B., Munimanda, G.K., Deimel, C., Setzer, A., *et al.* (2017). Extinctions, genetic erosion and conservation options for the black rhinoceros (Diceros bicornis). Sci. Rep. *7*, 41417.

9. Diceros bicornis mitochondrion, complete genome - Nucleotide - NCBI Available at: https://www.ncbi.nlm.nih.gov/nuccore/NC_012682.1?report=genbank.

10. Moodley, Y., Westbury, M.V., Russo, I.M., Gopalakrishnan, S., Rakotoarivelo, A., Olsen, R., Prost, S., Tunstall, T., Ryder, O.A., Dalén, L., Bruford, M.W. Interspecific gene flow and the evolution of specialisation in black and white rhinoceros. Mol. Biol. Evol.

11. McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. JOSS *3*, 861.

12. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

13. Petkova, D., Novembre, J., and Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. Nat. Genet. *48*, 94–100.

14. Moore, A.E., and Larkin, P.A. (2001). Drainage evolution in south-central Africa since the breakup of Gondwana. South Afr. J. Geol. *104*, 47–68.

15. Geraads, D. (2010). Rhinocerotidae. In Cenozoic Mammals of Africa (University of California Press).

16. Armstrong, E.E., Taylor, R.W., Miller, D.E., Kaelin, C.B., Barsh, G.S., Hadly, E.A., and Petrov, D. (2020). Long live the king: chromosome-level assembly of the lion (Panthera leo) using linked-read, Hi-C, and long-read data. BMC Biol. *18*, 3.

17. Gopalakrishnan, S., Samaniego Castruita, J.A., Sinding, M.-H.S., Kuderna, L.F.K., Räikkönen, J., Petersen, B., Sicheritz-Ponten, T., Larson, G., Orlando, L., Marques-Bonet, T., *et al.* (2017). The wolf reference genome sequence (Canis lupus lupus) and its implications for Canis spp. population genomics. BMC Genomics *18*, 495.

18. Zukowsky, L. (1965). Die Systematik der Gattung Diceros Gray, 1821. Zool. Gart. *30*, 1–178.

19. Groves, C., and Grubb, P. (2011). Ungulate Taxonomy (JHU Press).

20. Sánchez-Barreiro, F., Gopalakrishnan, S., Ramos-Madrigal, J., Westbury, M.V., de Manuel, M., Margaryan, A., Ciucani, M.M., Vieira, F.G., Patramanis, Y., Kalthoff, D.C., *et al.* (2020). Historical population declines prompted significant genomic erosion in the northern and southern white rhinoceros (Ceratotherium simum). bioRxiv, 2020.05.10.086686. Available at: https://www.biorxiv.org/content/10.1101/2020.05.10.086686v1.

21. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

22. Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., Fernández, R., Kircher, M., McCue, M., *et al.* (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nat. Protoc. *9*, 1056–1082.

23. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*, 356.

24. Li, H. (2011). Improving SNP discovery by base alignment quality. Bioinformatics *27*, 1157–1158.

25. Fu, Q., Mittnik, A., Johnson, P.L.F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., *et al.* (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. Curr. Biol. *23*, 553–559.

26. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

27. Leigh, J.W., and Bryant, D. (2015). popart : full-feature software for haplotype network construction. Methods Ecol. Evol. *6*, 1110–1116.

28. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. *10*, e1003537.

29. Margaryan, A., Sinding, M.-H.S., Liu, S., Vieira, F.G., Chan, Y.L., Nathan, S.K.S.S., Moodley, Y., Bruford, M.W., and Gilbert, M.T.P. (2020). Recent mitochondrial lineage extinction in the critically endangered Javan rhinoceros. Zool. J. Linn. Soc. Available at: https://academic.oup.com/zoolinnean/advance-article-pdf/doi/10.1093/zoolinnean/zlaa004/32883943/zlaa004.pdf.

30. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods *9*, 772.

31. FigTree Available at: http://tree.bio.ed.ac.uk/software/figtree/.

32. Letunic, I. iTOL: Interactive Tree Of Life. Available at: https://itol.embl.de/.

33. Korneliussen, T.S., and Moltke, I. (2015). NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. Bioinformatics *31*, 4009–4011.

34. Waples, R.K., Albrechtsen, A., and Moltke, I. (2019). Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. Mol. Ecol. *28*, 35–48.

35. Meisner, J., and Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. Genetics *210*, 719–731.

36. R: The R Project for Statistical Computing Available at: https://www.r-project.org/.

37. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Available at: https://ggplot2.tidyverse.org.

38. How to Use UMAP — umap 0.3 documentation Available at: https://umap-learn.readthedocs.io/en/latest/basic_usage.html.

39. Skotte, L., Korneliussen, T.S., and Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. Genetics *195*, 693–702.

40. Behr, A.A., Liu, K.Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). pong: fast analysis and visualization of latent clusters in population genetic data. Bioinformatics *32*, 2817–2823.

41. Tool for Google Maps Available at: http://www.birdtheme.org/useful/v3tool.html.

42. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

43. Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One *11*, e0163962.

44. Chernomor, O., von Haeseler, A., and Minh, B.Q. (2016). Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. Syst. Biol. *65*, 997–1008.

45. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.

46. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

47. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

48. WR_temporal_genomics_code Available at: https://github.com/fatimasb/WR_temporal_genomics_code.

49. by Richard A. Becker, O.S.C., by Thomas P Minka, A.R.W.R.V. by R.B.E., and Deckmyn., A. (2018). maps: Draw Geographical Maps Available at: https://CRAN.R-project.org/package=maps.

50. by Richard A. Becker, O.S.C., and by Ray Brownrigg., A.R.W.R.V. (2018). mapdata: Extra Map Databases Available at: https://CRAN.R-project.org/package=mapdata.

51. Bivand, R., and Lewin-Koh, N. (2019). maptools: Tools for Handling Spatial Objects Available at: https://CRAN.R-project.org/package=maptools.

52. Bivand, R., Keitt, T., and Rowlingson, B. (2019). rgdal: Bindings for the “Geospatial” Data Abstraction Library Available at: https://CRAN.R-project.org/package=rgdal.

53. Pebesma, E.J., R.S. Bivand (2005). Classes and methods for spatial data in R Available at: https://cran.r-project.org/doc/Rnews/.

54. Roger S. Bivand, Edzer Pebesma, Virgilio Gomez-Rubio (2013). Applied spatial data analysis with R, Second edition (Springer, NY.).

Conclusions and perspectives

Conclusions and perspectives

Genomes... what for?

The work conducted during this PhD education delves largely on the repercussions of demographic history on the genomic makeup of species. Moreover, it does so on species of conservation concern as a means to provide useful insight for their management. Evidently, this approach assumes that genomic diversity and changes in population structure matter for the sustained preservation of a species.

Extinction can be preceded by genetic warning signs, as has been found in the case of the woolly mammoth [1], but potential genetic threats are not always detected before extinction [2]. High-throughput sequencing and population genomics of wildlife are uncovering pervasive patterns of genomic erosion across species with a catastrophic recent history, such as the Iberian lynx [3] or the crested ibis [4]. But whether these genomic signatures shift from mere consequences of a decline to an active cause of extinction might well be case-dependent. The life-history traits and ecology of the species, the speed and magnitude of the population decline, and the unavoidable role of stochasticity likely conflate in determining the fate of a species or population.

In real-life scenarios, increasing census sizes and protecting sufficient amounts of habitat might seem vastly more relevant than focusing on the erosion of genomic diversity. After all, unless a clear negative fitness effect is detected (e.g. a recessive hereditary disease), one might think that the phenotypic manifestation of genomic erosion is in many cases trivial. It might even be that species' resilience depends as much on phenotypic plasticity as on genomic diversity. There is, nonetheless, a theoretical and empirical body of work that cannot be ignored linking genetic factors to an increased extinction risk, as detailed in the Introduction and in Chapter 1 of this thesis.

Conservation biology makes extensive use of genetics (and also genomics in recent years) to delineate populations and management units, plan their management and aid enforcement of conservation policies [5]. In that framework, temporal assessments can contribute to a better understanding of the organisms we want to preserve as dynamic entities that have changed through time, and will have to be ready for future change. Until we have proof that genomic erosion is not a relevant parameter to consider for species preservation, it is probably sensible to pursue the conservative option: monitor and protect as much as is left of the genomic diversity. In other words, pack abundantly for the trip to the future, we do not know what awaits us.

Beyond that, having genomic data at our disposal entails far more than just assessing the levels of

diversity (or loss of it) and population structure [6]. Re-sequenced genomes are humongous pots of data that hold much more latent potential. Particularly, and provided that high-quality genomic annotations are developed alongside, we could venture into the functional aspects of the genomic patterns we observe: is the genomic erosion detected actually triggering detrimental phenotypic effects? Even if not seen at present, are they likely to manifest in a few generations? How much adaptive variation has been lost during the latest demographic debacle?

Linking genotype to phenotype is a complicated task, even more so in non-model species for whom genomic resources (reference genomes and annotations) are not complete and of high quality. Fortunately genomics and population genomics are fast-evolving fields, whose reach has grown immensely [7]. As analytical tools are developed, and an increasing amount of data generated, today we can only fantasise about what we might be able to accomplish in wildlife research in just five or 10 years .

Beyond its use in conservation-oriented research, this data-driven stream is already fueling a wave of new thinking in ecology and evolution. Old questions can be tackled, such as what the determinants of genomic diversity across the tree of life might be [8]. Also, theoretical paradigms are being revisited, such as the significance of the neutral theory in light of evidence supporting natural selection as a major driver of genome evolution [9]. It is exciting to think that the way we understand evolution and ecology might be astonishingly updated in the coming years, if we invest time and careful reflection into interpreting what this vast amount of data reveals.

The contents of this thesis are a contribution to this growing current. In these three and a half years of work, a relatively large dataset was generated, and some analyses performed to address a few particular questions. What has been accomplished, however, has not yet exhausted the potential of the raw data.

Entering an era of *haute couture* conservation

Under the assumption that genomic diversity matters for the long-term conservation of species, and that we are better equipped than ever to survey it, how do we actually translate these research outputs into useful practice? Conservation is evidently as much a political and social sciences discipline as it is biological. For instance, basic research in population genomics offers limited help in the dismantling of the international criminal networks that orchestrate the illegal trade of rhino horn (or ivory, or live animals for that matter). Nevertheless, population genomics is meant to be

one of the legs supporting species management.

The work presented here could not have possibly been conducted in many other research institutions in the world. The amount of resources invested in generating, storing and analysing the data was considerably higher than what many other research groups can hope to allocate to their biological system of interest. Thus, it is hard to recommend this approach become the standard practice when budgets for both wildlife research and conservation itself are often not as generous.

It is costly and time-consuming to produce data amenable to population genomics analyses, even more so if aDNA samples are involved, and conservation is an urgent matter. While some highlight the usefulness of genomics in conservation and management [6], scepticism remains that this genomics revolution can readily benefit it [10]. Systemic challenges, such as the funding strategies, create a gap between the immediate interests of those conducting research and those making policies [11].

To a young academic researcher like me, reaching out effectively to policy-makers sounds like a daunting and obscure endeavour. In fact, in interpreting the results of Chapters 2 and 3 of this thesis, I found it challenging to figure what might be useful for conservation practitioners to know, and how to convey the information appropriately to non-specialist audiences. But if someone were to ask me, this is what I would say.

First, the SWR have suffered remarkable genomic erosion, so a very careful monitoring of inbreeding is probably very necessary, but for the most part it seems that giving them time and space to breed *does the trick*. Recreating NWR individuals owing to *in-vitro* fertilisation is a project underway [12], and what has been cryo-preserved of the NWR in terms of genomic diversity seems to be plenty to found a new population. It might sound like another fanciful approach of little actual impact on conservation that, moreover, reduces the identity of an organism to its genomic makeup (another matter on which many theses could be written). However, it might become a pioneering rewilding experience that draws enough attention to benefit conservation beyond the rebirthing of the NWR.

In the critically endangered black rhino, the results detailed in Chapter 3 suggest that six populations existed in the recent past, although management since the 1980s until recently was based on four ecotypes (one of them now extinct) [13]. Has it been a regrettable homogenization to treat these many historical populations as three coarsely delimited ecotypes? And given that regret is of little help now, how do we best preserve what is left of this species' diversity? I would suggest that a genomic survey of all the extant populations of black rhino would be useful in several ways.

It would help us verify if looking at the past (the historical portrait) to ensure the future is of any help in this particular case. Second, we could use it to assess local adaptation and potential effects of inbreeding and outbreeding to then predict the outcome of management practices even a few generations down the line. Third, we could quantify genetic load to estimate the long-term evolutionary potential of the species.

It is remarkable that the black rhino, a mammal so able to thrive under so diverse environmental conditions, is seemingly so reluctant to be maintained under human management. The white rhino, on the other hand, is more cooperating in that regard. Tailored approaches seem necessary, and in providing detail, population genomics can contribute.

The costs of generating genomic data continue to decrease, so conducting population genomics studies of non-model species is becoming increasingly reachable [7]. Therefore, even if not standardized, these resources can complement species management whenever available. New data and analytical tools, and sadly also the increase of recognised organisms in need of protection, are boosting a wealth of detailed knowledge about species' biology and the roads that lead to extinction. From *prêt-à-porter* conservation approaches, such as rules of thumb for minimum viable population size [14], we might be moving toward species- and population-specific *haute couture* management.

Beyond that, genomic data is helping us understand that more than species conservation it might be relevant to switch terms to simply biodiversity conservation. The genomic diversity lost to bottlenecks and extirpations was also part of the biosphere, and although it is useful to define 'conservation units', there is a part of human construction in this process [15]. A species-oriented mindset might come at the cost of inadvertently losing vast amounts of local, subtle or specialized diversity.

Despite the alluring advances of conservation biology research, however, the ultimate dilemma still persists: how necessary is this research to actually preserve biodiversity? We are well aware of the phenomena that are eroding biodiversity, and how linked they are to human activities. Yet, we place emphasis on understanding the object of conservation instead of the actor of the destruction. Is this procrastination of an increasingly large and resource-demanding human population? Maybe also the perfect excuse for biologists to continue raising funds to research the organisms they are curious about? Should we simply divert the funds now allocated to conservation biology research to actual conservation instead?

The biologist in me would certainly lament the elimination of funding for biodiversity research, while there must be less radical, more productive options. But I would say that a debate is needed,

one that involves all parties: biodiversity researchers, political authorities, ground conservation practitioners and society at large. We need to find a common ground, make a list of sorted priorities, and clearly phrase what each party can contribute with. Those of us happy to continue receiving funding for research on biodiversity, we have plenty of biological insight to offer.

Personal remarks

At the start of this PhD project the aspect I thought about the least was in fact conservation. I often used the argument that a bias exists in which species we care to preserve, and that our attention is unevenly distributed across taxa and biomes [11]. Large, often furry, 'cute' animals are winners in the race for appeal to humans (with rhinos ranking high). I thought that instead of so blatantly exhibiting these unconscious human biases, we should simply care to preserve habitats at large. Although becoming an expert in applied conservation was never part of my PhD plan, at least I have now realised that this discipline goes beyond those simplifying arguments.

Megafaunal species sit on a paradox: although they are subject to conservation efforts more often than other taxa, and receive public attention, they are also the most menaced, for they easily enter into conflict with human interests (see Introduction). Furthermore, given the extensive roles of the megafauna as ecosystem regulators and engineers, there might be good ecological intuition behind their preferential conservation. By and large, biases are still there, but they might be more useful than I previously thought.

***

The writing of this thesis was finalised during the start of the global crisis of the covid-19 pandemic. This disease is caused by a new coronavirus (SARS-CoV-2) that was seemingly 'cooked' in some animal reservoir, probably wild, and then made the jump to humans [16]. It might sound incidental, but this and other zoonotic outbreaks are far from being anecdotal, unrelated episodes. The way we invade and exploit natural spaces, and the way we treat wildlife play key roles [17]. This crisis is unveiling, among other things, the sickly relationship that we humans have developed with nature.

The work developed during my times as a PhD student started as a mere population genomics adventure. I sought to learn how to tame a large dataset, ask some questions to it, and expand my theoretical knowledge enough to understand the answers. I have to some extent managed to do so, but other reflections sprouted on the way, mainly related to conservation and the impact of

anthropogenic activities on biodiversity.

As I reflected earlier in this text, I find the translation of academic research outputs into social awareness and policy-making an inscrutable process. Yet, there is an undeniable flow of scientific and technological advancement pouring directly into all aspects of society. It must therefore be possible to find ways to do the same regarding the preservation of biodiversity. I am currently pondering what these mechanisms might be.

The present work, along with any other about threatened biodiversity, carries a relevant meaning beyond its potential direct use in conservation: it draws attention to the topic. In doing so it increases the chances of raising common awareness and understanding of this global problem. Any step in that direction, even if small, is a meaningful one, because societal and political movement will be key to smoothing the human-environment conflict. Small steps being small, they are as necessary as big leaps; so far this work has at least converted me.

References

1. Pečnerová, P., Palkopoulou, E., Wheat, C.W., Skoglund, P., Vartanyan, S., Tikhonov, A., Nikolskiy, P., van der Plicht, J., Díez-Del-Molino, D., and Dalén, L. (2017). Mitogenome evolution in the last surviving woolly mammoth population reveals neutral and functional consequences of small population size. Evol Lett *1*, 292–303.

2. Dussex, N., von Seth, J., Knapp, M., Kardailsky, O., Robertson, B.C., and Dalén, L. (2019). Complete genomes of two extinct New Zealand passerines show responses to climate fluctuations but no evidence for genomic erosion prior to extinction. Biol. Lett. *15*, 20190491.

3. Abascal, F., Corvelo, A., Cruz, F., Villanueva-Cañas, J.L., Vlasova, A., Marcet-Houben, M., Martínez-Cruz, B., Cheng, J.Y., Prieto, P., Quesada, V., *et al.* (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. Genome Biol. *17*, 251.

4. Feng, S., Fang, Q., Barnett, R., Li, C., Han, S., Kuhlwilm, M., Zhou, L., Pan, H., Deng, Y., Chen, G., *et al.* (2019). The Genomic Footprints of the Fall and Recovery of the Crested Ibis. Curr. Biol. *29*, 340–349.

5. Van Dyke, F. ed. (2008). Genetic Diversity – Understanding Conservation at Genetic Levels. In Conservation Biology: Foundations, Concepts, Applications (Dordrecht: Springer Netherlands), pp. 153–184.

6. Supple, M.A., and Shapiro, B. (2018). Conservation of biodiversity in the genomics era. Genome Biol. *19*, 131.

7. Luikart, G., Kardos, M., Hand, B.K., Rajora, O.P., Aitken, S.N., and Hohenlohe, P.A. (2019). Population Genomics: Advancing Understanding of Nature. In Population Genomics: Concepts, Approaches and Applications, O. P. Rajora, ed. (Cham: Springer International Publishing), pp. 3–79.

8. Leffler, E.M., Bullaughey, K., Matute, D.R., Meyer, W.K., Ségurel, L., Venkat, A., Andolfatto, P., and Przeworski, M. (2012). Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? PLoS Biol. *10*, e1001388.

9. Kern, A.D., and Hahn, M.W. (2018). The Neutral Theory in Light of Natural Selection. Mol. Biol. Evol. *35*, 1366–1371.

10. Shafer, A.B.A., Wolf, J.B.W., Alves, P.C., Bergström, L., Bruford, M.W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., *et al.* (2015). Genomics and the challenging translation into conservation practice. Trends Ecol. Evol. *30*, 78–87.

11. Griffiths, R.A., and Dos Santos, M. (2012). Trends in conservation biology: Progress or procrastination in a new millennium? Biol. Conserv. *153*, 153–158.

12. Hildebrandt, T.B., Hermes, R., Colleoni, S., Diecke, S., Holtze, S., Renfree, M.B., Stejskal, J., Hayashi, K., Drukker, M., Loi, P., *et al.* (2018). Embryos and embryonic stem cells from the white rhinoceros. Nat. Commun. *9*, 2589.

13. Rookmaaker, K. (2011). A review of black rhino systematics proposed in Ungulate Taxonomy by Groves and Grubb (2011) and its implications for rhino conservation. Pachyderm *50*, 72–76.

14. Jamieson, I.G., and Allendorf, F.W. (2012). How does the 50/500 rule apply to MVPs? Trends Ecol. Evol. *27*, 578–584.

15. Coates, D.J., Byrne, M., and Moritz, C. (2018). Genetic Diversity and Conservation Units: Dealing With the Species-Population Continuum in the Age of Genomics. Frontiers in Ecology and Evolution *6*, 165.

16. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020). The proximal origin of SARS-CoV-2. Nat. Med. *26*, 450–452.

17. IPBES Guest Article: COVID-19 Stimulus Measures Must Save Lives, Protect Livelihoods, and Safeguard Nature to Reduce the Risk of Future Pandemics Available at: https://ipbes.net/covid19stimulus.

Appendix

Appendix

I hereby list the research outputs of a series of projects in which I have been involved during my PhD education, but that were not part of the main aims of my research and academic plan. Currently these studies are manuscripts in preparation or submitted manuscripts under review process. I contributed to these projects mostly with sample management, protocol optimisation and generation of genomic data from aDNA samples. These side-projects are not part of the contents to be evaluated by the assessment committee of this thesis, therefore I specify solely their title, author list and affiliations and abstract when available.

Predicting sample success for large-scale ancient DNA studies: examining factors influencing DNA yield, sequencing success and damage patterns

Xénia Keighley[1,2] , Maiken Hemme Bro-Jørgensen[1,3#], Hans Ahlgren[3#], Paul Szpak[4], Marta Maria Ciucani[1], Fátima Sánchez Barreiro[1], Lesley Howse[5], Anne Birgitte Gotfredsen[6], Peter Jordan[2], Aikaterini Glykou[3], Kerstin Lidén[3], Morten Tange Olsen[1*]

[1]Section for Evolutionary Genomics, GLOBE Institute, Natural History Museum of Denmark, University of Copenhagen, CSS Building 7, Øster Farimagsgade 5 DK-1353 Copenhagen K, Denmark

[2]Arctic Centre/Groningen Institute of Archaeology, Faculty of Arts, University of Groningen, PO Box 716, 9700 AS Groningen, The Netherlands

[3] Archaeological Research Laboratory, Department of Archaeology and Classical Studies, Stockholm University, 106 91 Stockholm, Sweden

[4]Department of Anthropology, Trent University, 1600 West Bank Drive, Peterborough, Ontario, Canada K9L 0G2

[5]Archaeology Centre, University of Toronto, 15 Russell Street, Toronto, Ontario, M5S 2S2, Canada

[6]GLOBE Institute, Natural History Museum of Denmark, University of Copenhagen

* Corresponding author: morten.olsen@bio.ku.dk

# These authors contributed equally

The genetic history and diffusion routes of early maize in North America

Jazmín Ramos-Madrigal[1*],Nathan Wales[1,2], Bruce Smith[3,4], Fátima Sánchez-Barreiro[1], Christian Carøe[1], Filipe G. Viera[1], Rute R. da Fonseca[5], Shanlin Liu[1,6], Chunxue Guo[6], Guojie Zhang[6,7], Bent Petersen[8], Thomas Sicheritz-Pontén[8], M. Thomas P. Gilbert[1,9,10*]

[1]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark

[2]Department of Plant and Microbial Biology, University of California, 111 Koshland Hall, Berkeley, CA 94720, USA

[3]Program in Human Ecology and Archaeobiology, Department of Anthropology, National Museum of Natural History, Smithsonian Institution, 10th and Constitution Avenue, Washington, DC 20560, USA

[4]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

[5]The Bioinformatics Centre, Department of Biology, University of Copenhagen, Denmark.

[6]China National GeneBank, BGI-Shenzhen, 518083, China

[7]Centre for Social Evolution, Department of Biology, Universitetsparken 15, University of Copenhagen, Copenhagen DK-2100, Denmark

[8]DTU Bioinformatics, Department of Bio and Health Informatics, Technical University of Denmark, Building 208, DK-2800 Lyngby, Denmark

[9]Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, WA 6102, Australia

[10]NTNU University Museum, 7491 Trondheim, Norway

* Corresponding authors

Abstract

Archaeological and genetic evidence from modern and ancient maize (*Zea mays*) samples indicate that maize initially reached the southwestern United States (US) by around 4,000 years ago via a highland Mexican route, followed by a second introduction via the Pacific coast, around 2,000 years ago. However, maize diffusion routes northward from the domestication center in southern Mexico, up to the US Southwest and Southeast remain contentious. To explore the potential diffusion routes, we generated high-throughput sequencing data from 24 ancient maize cobs and kernels from three archaeological sites dated to different time periods: the Romero Cave in northeastern Mexico (n=2; 2,450-2,750 BP), the Three Fir Shelter in the southwestern US (n=6; 2,075-2,500 BP) and the Ozarks Shelter in the southeastern US (n=16; 70-1,120 BP). By comparing these data to a reference dataset composed of modern maize landraces and previously published ancient maize samples, we

assess the genetic ancestry of early maize in North America and identify potential dispersion routes north from the domestication center, into northeastern Mexico and the southeastern US

Sequencing the Asian unicorn: genomics sheds light on Pseudoryx nghetinhensis *(saola) evolution and population history*

Shanlin Liu[1,2,3], Faezah Mohd Salleh[1,4], Fátima Sánchez Barreiro[1], Jazmin Ramos Madrigal[1], Mikkel-Holger Strander Sinding[1], Minh Duc Le[6], Guanliang Meng[2,3], Chentao Yang[2,3], Anders Albrechtsen[5], Guojie Zhang[2,3,5], Lei Chen[7], Wen Wang[7], Hans R. Siegismund[5], M. Thomas P. Gilbert[1*], Rasmus Heller[5*]

[1]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7,1350, Copenhagen, Denmark

[2]BGI-Shenzhen, Shenzhen, GuangDong, 518083, China

[3]China National GeneBank-Shenzhen, BGI-Shenzhen, Shenzhen, Guangdong Province 518083, China

[4]Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

[5]Department of Biology, University of Copenhagen, Copenhagen, Denmark

[6]Centre for Natural Resources and Environmental Studies, Hanoi National University, 19 Le Thanh Tong, Hanoi, Vietnam

[7]Center for Ecological and Environmental Science, Northwestern Polytechnical University, Xi'an 710072, China

*Correspondence address. M. Thomas P. Gilbert, Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark. Tel: +45 23 71 25 19; E-mail: tgilbert@snm.ku.dk; Rasmus Heller, Department of Biology, University of Copenhagen, Copenhagen, Denmark. E-mail: rheller@bio.ku.dk

Abstract

The critically endangered saola is the conservation flagship species of the Annamite Range Moist Forest Ecoregion. It remains extremely poorly understood, despite tremendous efforts directed at studying it over the last 20 years. The species is by all accounts on the brink of extinction, yet lacks any scientific description beyond the most rudimentary. We constructed a saola reference genome sequence with scaffold N50 reaching ca. 2.3 Mb, and drafted the polymorphism dataset of saola

from a unique collection of 40 samples dating to the early 1990s. Our whole genome-based phylogenetic analysis indicates that the saola is sister to the ancestor of the cattle and water buffalo lineages. We also found that there is evidence for positive selection on genes linked to brain function which may play critical roles in the saola's reported tame, shy and seclusive behaviors. In addition, our data indicate that the saola population size has largely declined throughout the past ~1 million years, which in combination with its extremely low heterozygosity (ca. 0.06%) and abundant short Runs of Homozygosity segments (total length of ca. 1/3 of the genome), suggest that the saola's demise was initiated long ago and the most recent decline in saola population size is the continuation of a millennia-old decline, possibly related to the agricultural subsistence of modern human populations.

Ancient genomics of the woolly rhinoceros reveals demographic stability prior to extinction and a genetic basis of adaptation

Lord E.[1-3*#], Dussex N.[1,2#], Kierczak M.[4], Díez-del-Molino D.[1,3], Ryder OA.[5], Stanton, DWG.[1,2], Gilbert MTP.[6,7], Sánchez-Barreiro F.[6], Zhang G.[8-11], Sinding MHS.[6,12], Lorenzen ED.[6], Willerslev E.[6], Fedorov S.[13], Bocherens H.[14,15], Nathan S.[16], Goossens B.[16-19], van der Plicht J.[20], Chan YL.[1-3], Prost S.[21-22], Potapova O.[23-24], Kirillova I.[25], Lister AM.[26], Heintzman PD.[27], Kapp J.[28], Shapiro B.[28-29], Vartanyan S.[30], Götherström A.1,[31], Dalén L.[1-3*]

[1]Centre for Palaeogenetics, Svante Arrhenius väg 20C, SE-10691 Stockholm, 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, SE-752 37 Uppsala, 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, 1352 Copenhagen, Denmark

[7]Norwegian University of Science and Technology, University Museum, 7491 Trondheim, Norway

[8]Section for Ecology and Evolution, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark

[9]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 650223, Kunming, China

[10]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, 650223, Kunming, China

[11]BGI-Shenzhen, 518083, Shenzhen, China

[12]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland

[13]Mammoth museum of North-Eastern Federal University, Yakutsk, Russia

[14]Senckenberg Centre for Human Evolution and Palaeoenvironment (S-HEP), Sigwartstr. 10, 72076 Tübingen, Germany

[15]Department of Geosciences, Biogeology, University of Tübingen, Hölderlinstr. 12, 72074 Tübingen, Germany

[16]Sabah Wildlife Department, Wisma Muis, 88100 Kota Kinabalu, Sabah, Malaysia

[17]Organisms and Environment Division, Cardiff School of Biosciences, 33 Park Place, Cardiff CF10 3BA, UK

[18]Sustainable Places Research Institute, Cardiff University, 33 Park Place, Cardiff CF10 3BA, UK

[19]Danau Girang Field Centre, c/o Sabah Wildlife Department, Wisma Muis, 88100 Kota Kinabalu, Sabah, Malaysia

[20]Center for Isotope Research, Groningen University, Groningen, The Netherlands

[21]LOEWE-Centre for Translational Biodiversity Genomics, Senckenberg Museum, Frankfurt, Germany

[22]South African National Biodiversity Institute, National Zoological Garden, Pretoria, South Africa

[23]Pleistocene Park Foundation, Philadelphia PA 19006, USA

[24]Mammoth Site of Hot Springs, SD, Inc., Hot Springs, SD 57747, USA

[25]Ice Age Museum, National Alliance of Shidlovskiy ‘Ice Age’, Moscow, Russia

[26]Department of Earth Sciences, Natural History Museum, London SW7 5BD, UK

[27]The Arctic University Museum of Norway, UiT The Arctic University of Norway, 9037 Tromsø, Norway

[28]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz CA 95064, USA

[29]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz CA 96050 USA

[30]N.A. Shilo North-East Interdisciplinary Scientific Research Institute, Far East Branch, Russian Academy of Sciences (NEISRI FEB RAS), Magadan 685000, Russia

[31]Archaeological Research Laboratory, Department of Archaeology and Classical Studies, Stockholm University, 106 91 Stockholm, Sweden

# These authors contributed equally to this study

*Corresponding authors: Edana Lord, Centre for Palaeogenetics, Svante Arrhenius väg 20C, SE-10691 Stockholm, Sweden (edana.lord@zoologi.su.se); Love Dalén Centre for Palaeogenetics, Svante Arrhenius väg 20C, SE-10691 Stockholm, Sweden (love.dalen@nrm.se)

Abstract

The Late Pleistocene in Eurasia was marked by severe climatic fluctuations and the arrival of anatomically modern humans. Climatic changes are suggested to have resulted in increased rates of faunal turnover and adaptation, particularly in the Arctic, while climate change and/or over-hunting by humans may have driven some species to extinction in the end-Pleistocene to mid-Holocene. The woolly rhinoceros (*Coelodonta antiquitatis*) was a cold-adapted species widely distributed across northern Eurasia in the Late Pleistocene, and went extinct approximately 14 ka cal BP. Here, we investigate the demographic history of this species leading up to its extinction, and the genomic basis of adaptation to an arctic environment, using a 13.6X genome from a woolly rhinoceros dated to ca. 18.5 ka cal BP, and mitochondrial genomes from 14 specimens ranging from approximately 50-14.1 ka cal BP. Our results suggest that unlike other northern megafauna, the woolly rhinoceros went through an expansion in effective population size at around 29.7 ka BP, which was followed by a period of demographic stability to at least 18.5 ka cal BP, until close to the species' extinction. Additionally, we find no evidence for increased inbreeding or reduced genome-wide heterozygosity prior to the species' extinction. The extinction may thus have been sudden, and was likely driven by rapid warming in the Bølling-Allerød interstadial. Furthermore, we show that the woolly rhinoceros had adaptations to temperature sensation similar to those of the woolly mammoth.