This submission is intended as an ARTICLE in the Discoveries section.

Full Title: Interspecific gene flow and the evolution of specialisation in black and white rhinoceros.

Short title: Pliocene gene flow in African Rhinoceros

Authors: Yoshan Moodley[1]#, Michael V. Westbury[2]#, Isa-Rita M. Russo[3], Shyam Gopalakrishnan[2], Andrinajoro Rakotoarivelo[1,4], Remi-Andre Olsen[5], Stefan Prost[6,7], Tate Tunstall[8], Oliver A. Ryder[8], Love Dalén[9,10], Michael W. Bruford[3,11]

#equally contributing first authors

Affiliations: [1] Department of Zoology, University of Venda, Private Bag X5050, Thohoyandou 0950, Republic of South Africa, [2] Section for Evolutionary Genomics, GLOBE institute, University of Copenhagen, Denmark, [3] School of Biosciences, Sir Martin Evans Building, Cardiff University, Museum Avenue, Cardiff, CF10 3AX, United Kingdom, [4] Natiora Ahy Madagasikara, Lot IIU57K Bis, Ampahibe, Antananarivo 101, Madagascar (ARR), [5] Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Box 1031, SE-17121 Solna, Sweden, [6] LOEWE-Centre for Translational Biodiversity Genomics, Senckenberg Museum, Frankfurt, Germany, [7] South African National Biodiversity Institute, National Zoological Garden, Pretoria, Republic of South Africa, [8] San Diego Zoo Institute for Conservation Research, San Diego Zoo Global, Escondido, California, United States of America, [9] Centre for Palaeogenetics, Svante Arrhenius Väg 20C, SE-10691 Stockholm, Sweden, [10] Department of Bioinformatics and Genetics, Swedish Museum of Natural History, SE-10405 Stockholm, Sweden, [11] Sustainable Places Research Institute, Cardiff University, Cardiff CF10 3BA, United Kingdom

* Corresponding author. Email: yoshan.moodley@univen.ac.za

Abstract

Africa's black (*Diceros bicornis*) and white (*Ceratotherium simum*) rhinoceros are closely related sister-taxa that evolved highly divergent obligate browsing and grazing feeding strategies. Although their precursor species *D. praecox* and *C. mauritanicum* appear in the fossil record ~5.2 million years ago (Ma), by 4 Ma both were still mixed feeders, and were even spatio-temporally sympatric at several Pliocene sites in what is today Africa's Rift Valley. Here, we ask whether or not *D. praecox* and *C. mauritanicum* were reproductively isolated when they came into Pliocene secondary contact. We sequenced and *de novo* assembled the first annotated black rhinoceros reference genome, and compared it with available genomes of other black and white rhinoceros. We show that ancestral gene flow between *D. praecox* and *C. mauritanicum* ceased sometime between 3.3 and 4.1 Ma, despite conventional methods for the detection of gene flow from whole genome data returning false positive signatures of recent interspecific migration due to incomplete lineage sorting. We propose that ongoing Pliocene genetic exchange, for up to 2 million years after initial divergence, could have potentially hindered the development of obligate feeding strategies until both species were fully reproductively isolated, but that the more severe and shifting palaeoclimate of the early Pleistocene was likely the ultimate driver of ecological specialisation in African rhinoceros.

Keywords: Reproductive isolation, ancestral gene flow, incomplete lineage sorting, rhinoceros, Pliocene, genomes

Introduction

Although the age of Pleistocene mammalian megaherbivores is largely over, Africa is the only continent to still harbour significant wild populations of its late-tertiary megafauna. Africa's black (*Diceros bicornis*) and white (*Ceratotherium simum*) rhinoceros are relics of this bygone "golden age" of large mammals, yet, because of long-term historical demand for their products, man has succeeded in driving this iconic group to the brink of extinction across the world (Hillman 1980, Chilvers 1990). Although rhinoceros are among the world's most endangered mammals, the two African species have fared slightly better than their three Asian counterparts, owing mainly to intensive conservation interventions during the second half of the 20th Century, resulting in a global population of approximately 20,000 white and 5,000 black rhinoceros (Emslie *et al.* 2016). However, the unfortunate consequence of these population gains is that the most recent poaching epidemic, driven by increasing demand for rhinoceros horn in East and South-East Asia (Milliken and Shaw 2012, Kennaugh 2015), is targeting the more common African species.

A rich fossil record shows that rhinoceros species have endured a series of severe Plio-Pleistocene climatic and tectonic upheavals, to which the majority of their megafaunal contemporaries succumbed (Barnosky *et al.* 2004). In Africa, the paleoclimate during this time fluctuated between warmer, wetter, forest-rich interglacial periods that favoured browsers, and cooler, drier grassland-rich glacial periods that benefited grazers. Although black and white rhinoceros are closely related, they have evolved divergent feeding strategies. The black rhinoceros is the smaller of the two species and primarily a browser, holding its head horizontally to the ground in order to feed on leaves and twigs using a hooked upper lip. In contrast, the white rhinoceros is an obligate grazer with hypsodont teeth, a heavy, elongated skull that is held vertically and lower to the ground, with squared-off lips to enable efficient grazing. Both feeding strategies evolved within the last 6-7 million years (Ma), from about the end of the warm Miocene epoch, as global CO_2 levels decreased, leading to the more arid, seasonal and shifting paleoclimates of the Plio-Pleistocene (WoldeGabriel *et al.* 2001, deMenocal 2004). Although grazing rhinoceros such as *Teleoceras* had already evolved during the Miocene, the stem lineage leading to modern African rhinoceros was a mixed-feeder, represented by the morphologically intermediate *C. neumayri*, which inhabited late Miocene southern Europe from the Balkans to Iran approximately 7-9 Ma (Zeuner 1934, Geraads 2005, Geraads and Spassov 2009). This stem lineage diverged in Africa into *D. praecox* and *C. mauritanicum*, the direct ancestors of black and white rhinoceros respectively (Geraads 2005, Geraads 2017). The earliest fossil appearance of *D. praecox* is at Kuseralee in the Middle Awash Valley of Ethiopia (Giaourtsakis *et al.* 2009, Geraads 2017) and the *Ceratotherium* lineage at Langebaanweg in South Africa (Hooijer *et al.* 1972, Geraads 2005), both sites dating to about 5.2 Ma. The initial divergence between *D. praecox* and *C. mauritanicum* must therefore have occurred no later than around the Mio-Pliocene boundary about 5.3 Ma.

While changing paleoclimates provide a means for the evolution and fixation of different adaptations, they may also eventually bring speciating populations into secondary contact, where gene flow might bring their diverging evolutionary trajectories back into line (Mayr 1942) and/or promote the introgression of adaptive features between populations (Pardo-Diaz *et al.* 2012, Dasmahapatra *et al.* 2012, Racimo *et al.* 2015). Secondary contact between diverging precursor

93 species could have taken place at several mid-late Pliocene sites (3.0-4.3 Ma) in what is today
94 Africa's Rift Valley, where both *D. praecox* and *C. mauritanicum* fossils co-occur within the same
95 horizon (Geraads 2010, red-grey squares Figure 1). One particularly rhinoceros-rich site is
96 Kanapoi in north-western Kenya, where middle Pliocene (4 Ma) *D. praecox* and *C. mauritanicum*
97 fossils show evidence that they had already evolved some of the adaptations to browsing and
98 grazing respectively. However, cranial morphology and stable $\delta^{13}C$ isotope ratios of these
99 precursor species from Kanapoi and other mid-Pliocene sites confirm that both precursor species
100 were still mixed feeders relative to their modern descendants (Geraads 2017).
101 The transition between these mixed feeding precursors and specialized modern species
102 occurred in East Africa in the late Pliocene or early Pleistocene, as the first fossil emergence of
103 the black rhinoceros was at Koobi Fora about 2.5 Ma, and white rhinoceros at Olduvai around 1.8
104 Ma (Hooijer 1969, Harris 1983, Geraads 2010). The Pleistocene distribution of the black
105 rhinoceros appears to have been widespread across sub-Saharan Africa, outside dense Central
106 and West African rainforests, and is similar to its Holocene distribution (Figure 1), but with a strong
107 genetic discontinuity on either side of the Zambezi River in South-Central Africa (Moodley *et al.*
108 2017). Given that observed levels of microsatellite and mitochondrial genetic diversity were much
109 higher to the north of the Zambezi, we hypothesise that black rhinoceros expanding from East
110 Africa, crossed into southern Africa prior to the existence of the river's present day course, and
111 were then restricted to the subregion by a river capture event 125-150 Ka (Moore and Larkin
112 2001) with limited gene flow connecting eastern and southern populations. In contrast, the
113 Pleistocene white rhinoceros ranged more widely than the black rhinoceros, occurring from South
114 Africa to as far north as Libya (Geraads 2010). However, this range contracted significantly into
115 two genetically distinct populations during the Holocene, with the northern white rhinoceros
116 inhabiting central African grasslands west of the Nile River, and the southern white rhinoceros
117 restricted to grasslands south of the Zambezi. Although their Holocene ranges are discontinuous,
118 microsatellite data suggest that the two white rhinoceros populations may have come into
119 secondary contact sometime during the last glacial period (14-106 Ka) when grasslands were
120 continuous between eastern and southern Africa (Moodley *et al.* 2018).
121 In this study we ask whether *D. praecox* and *C. mauritanicum* were reproductively isolated
122 when they came into secondary contact at Kanapoi and other mid-Pliocene sites, and whether
123 ongoing genetic exchange between the two precursor species could have delayed the evolution
124 of their obligate modern day feeding strategies. We attempt to answer this by estimating the time
125 at which the black and the white rhinoceros became fully reproductively isolated. Furthermore,
126 we date the divergences within each species and contrast these with the times at which
127 populations last came into secondary contact. Given their specialised feeding roles, we also
128 predicted a strong influence of fluctuating Pleistocene paleoclimates on the demographic history
129 of each species.
130 Until the advent of evolutionary genomics methods, these ideas were largely untestable,
131 mainly because of the limited resolution of Pliocene evolutionary events from the handful of
132 previously available molecular markers (Groves *et al.* 2006, Moodley *et al.* 2018). In contrast,
133 data from millions of polymorphic loci from whole genomes now offer the opportunity to
134 reconstruct patterns of genome wide diversity, divergence, and demographic history over much
135 deeper time periods. Therefore, to shed light on these questions, we established the first high-

coverage *de novo* black rhinoceros genome assembly and, together with the previously generated white rhinoceros assembly and two further resequenced rhinoceros genomes, we carried out comparative analyses of the evolution of both African species.

Results

De novo assembly and annotation of the black rhinoceros reference genome

We present a high quality reference genome for the black rhinoceros (SAMN14912225) from an ear tissue sample of an individual sampled in KwaZulu-Natal, South Africa, at the southern end of the species range. We reconstructed the assembly using a combination of paired-end, mate paired, and chromatin-based sequencing libraries. First, we generated a baseline assembly using a combination of short and long insert libraries using the *de novo* assembler Allpaths-LG (Gnerre *et al.* 2011). We subsequently carried out super-scaffolding using chromatin-based Chicago libraries (Putnam *et al.* 2016) and the Hi-Rise pipeline. This resulted in a 34.6-fold coverage genome with a total assembly length of 2.33GB, with a scaffold N50 of 28.5MB and 4,264 scaffolds. A BUSCO assessment (Simão *et. al.* 2015) of the gene content of the assembly revealed only 13 (0.3%) duplicated, 88 (2.1%) fragmented and 116 (2.8%) missing mammalian single copy orthologs. We then annotated the assembly using *ab initio* gene prediction and homology-based gene identification, which resulted in 19,914 transcripts.

To unravel the evolutionary history between black and white rhinoceros, and to capture as much of the variation within each species, we analysed our newly sequenced genome together with three other African rhinoceros genomes. The black rhinoceros was represented by genomes from its southern and eastern (SAMN14911588, 16-fold coverage) populations, and the white rhinoceros by its southern (35-fold coverage) and northern (SAMN14911569, 16-fold coverage) populations.

Genome-wide heterozygosity

The proportion of heterozygous sites in African rhinoceros genomes varied both within and between species. The highest values were from populations in the northern range of both species, with the eastern black rhinoceros and northern white rhinoceros returning the highest diversity values at 0.00075 (S.D. 0.00071-0.00079) and 0.00045 (S.D. 0.00036-0.00054) heterozygous sites per base respectively. The southern populations of both species revealed lower values with southern black rhinoceros at 0.00031 (S.D. 0.00022-0.00040) and the southern white rhinoceros at 0.00027 (S.D. 0.00022-0.00033) heterozygous sites.

Divergence and mutation rates

Based on the autosomal sequences, we calculated an average pairwise divergence of 0.0093 between the two species. Within each species, lower divergences were estimated between eastern and southern black rhinoceros (0.0011) and northern and southern white rhinoceros (0.0010). Using a conservative estimate (latest possible occurrence) for the split, or end of panmixia, between black and white rhinoceros lineages at the Mio-Pliocene boundary ~5.3Ma (Geraads 2005, 2017), these pairwise distances were translated into approximate within-

species divergence times of 641 and 578 thousand years ago (Ka) for the black and white rhinoceros, respectively (Figure 2A). Using the pairwise distance between the white and black rhinoceros, we calculated an autosomal mutation rate for African rhinoceros of 8.8×10^{-10} substitutions per year, which is only slightly lower than the commonly implemented human mutation rate of 1×10^{-9} (Li and Durbin 2011), and refutes the commonly held view that evolutionary rates in rhinoceros genomes are substantially lower than the mammalian average (Gissi *et al.* 2000). Furthermore, we calculated the per generation mutation rate for each rhinoceros species independently, assuming a generation time of 24 years for the black rhinoceros (Moodley *et al.* 2017), giving a mutation rate of 2.1×10^{-8}, and a generation time of 27 years for the white rhinoceros (Moodley *et al.* 2018), giving a mutation rate of 2.4×10^{-8}.

Demographic reconstruction

We reconstructed the demographic histories of both African rhinoceros species over the second half of the Pleistocene (<1.4 Ma) using a pairwise sequentially Markovian coalescent (PSMC) model. Both species show a gradual reduction in effective population size (Ne) to less than half their original size until about 520-540 Ka ago in black rhinoceros and 440-460 Ka in white rhinoceros (yellow stars, Figure 2B). The demographic trajectories of both species also diverged at this low point, indicating the approximate times at which the ancestral populations of black rhinoceros and white rhinoceros divided, signalling the end of panmixia within each species. After this point, all four genomes then appear to follow independent Middle Pleistocene population expansions. Interestingly, the southern populations of both species reach their highest size at about 230 Ka, earlier than their northern counterparts the eastern black rhinoceros at 200 Ka and the northern white rhinoceros at 180 Ka. All four populations then contract to Holocene levels of Ne below 5,000, although both southern African populations show a secondary but minor population expansion at about 50 Ka for the southern white rhinoceros and within the last 20 Ka for the southern black rhinoceros. It is important to note that PSMC-inferred demographic trajectories are often difficult to interpret literally (Beichman *et al.* 2017). We also investigated whether different sequencing depths, especially in the case of the northern white and eastern black rhinoceros (16-fold coverage) may have influenced the observed demographic trajectories. Based on a comparison between PSMC trajectories reconstructed using our newly sequenced ~35-fold coverage southern black rhinoceros genome, and the same genome downsampled to 16-fold coverage (Supplementary Figure S1), we deduced that the differences caused by differential coverage was negligible.

Post-divergence gene flow

Signatures of post-divergence gene flow between the two African rhinoceros lineages were inferred through a variety of approaches relying on the known topology of the African rhinoceros species tree (Figure 2A). We first implemented the four-taxon ABBA/BABA or D-statistic analysis (Durand *et al.* 2011), which showed evidence for significant levels of post-divergence interspecific gene flow between southern white rhinoceros and both black rhinoceros, as well as between eastern black rhinoceros and both white rhinoceros (Supplementary Table S1, Supplementary Figure S2). This result was unexpected as it did not follow a geographically mediated pattern, as one would expect gene flow between geographically close lineages to be

the most probable, that is, between northern white and eastern black, and between southern white and southern black rhinoceros (see Figure 1). Then, to add further levels of information, such as the direction of gene flow and whether gene flow occurred between ancestral lineages, we performed the complementary five-taxon Dfoil analysis (Pease and Hahn 2015) which utilises a system of four D-statistics to distinguish introgressions in a symmetric five-taxon phylogeny, using the Sumatran rhinoceros (*Dicerorhinus sumatrensis*) as outgroup. This analysis also indicated several instances of gene flow (Supplementary Table S2), although most of these were at a very low frequency. Dfoil analysis did, however, suggest similar levels of high frequency gene flow between the ancestral white rhinoceros lineage and both black rhinoceros. We then extracted regions consistently showing evidence for admixture, regardless of window size, from the Dfoil results and cross referenced these putatively introgressed genomic segments against the white rhinoceros annotation, revealing an exchange of 47 protein coding genes, the majority of which had no human analogue (Supplementary Table S3). Using the recovered gene codes, we ran a gene ontology (GO) enrichment test with GOrilla (Eden *et al.* 2009) to investigate whether certain biological processes may have been selectively retained from past introgression events. We found no significantly enriched GO terms. Finally, we investigated the length of contiguous introgressed windows to understand the relative timing of introgression. We found the vast majority of introgressed windows to be singletons with only very few consecutive windows detected (Supplementary Table S4).

As both D-statistics and D-foil analyses rely on the D-statistic to infer gene flow, they can both be confounded by similar caveats and biases based on the data. Therefore, we computed the D3-statistic (Hahn and Hibbins, 2019), which is a three-sample test for introgression that uses pairwise distances to estimate the presence of admixture in a triplet taxa ((A, B), C). D3 bypasses the need for an outgroup genome to polarise ancestral and derived alleles, so should be more robust than D-statistics when no suitable closely related outgroup is available. To test for significance, we ran the D3 analysis using both 100 KB, and 1 MB, non-overlapping sliding windows. Results showed no significant levels of differential gene flow between any of the African rhinoceros triplets and were consistent regardless of window size (Supplementary Table S5). To further test for admixture we also implemented Treemix (Pickrell and Pritchard, 2012) and the F3- and F4-statistics (Reich *et al*. 2009, Keinan *et al*. 2007). These analyses neither confirmed nor excluded the possibility of post-divergence interspecific gene flow suggested by D-statistics, but we include their details in Supplementary methods, Figures S3-S6 and Tables S6 an S7.

Reproductive isolation and the cessation of post-divergence gene flow

To ascertain when admixture between the speciating African rhinoceros lineages may have ceased, we conducted multiple F1 hybrid pairwise sequentially Markovian coalescent model (hPSMC) analyses using pseudo-diploidised African rhinoceros genomes, and intermediate mutation rates and generation times. This analysis is based on the premise that a pseudo F1 hybrid genome cannot coalesce more recently than the speciation event of the two parental species (Cahill *et al.* 2016). This point of coalescence is represented by a transition from an infinite population size to the population size of the shared ancestral lineage prior to divergence, thus allowing the determination of the latest time for the development of reproductive isolation between the two species. However, as hPSMC utilises PSMC, and PSMC is known to portray rapid

changes in ancestral Ne as gradual transitions, one cannot apply a purely qualitative approach to estimating divergence times. Therefore, we ran simulations specifying various divergence times between the individuals of interest. Simulations were run using the hPSMC_quantify_split_time.py python script from the hPSMC tool suite specifying pre-divergence Ne, time windows for divergence, and default parameters. Results from the real data as well as simulations based on Ne's calculated before its exponential increase to infinity, indicated that reproductive isolation between black and white rhinoceros lineages occurred between 3.3 and 4.1 Ma (Figure 3A), much more recently than the initial divergence time of the two species at ~5.3 Ma or earlier (Geraads 2005, 2017). This result was the same regardless of which of the two genomes of each species were compared and which species was used as the mapping reference.

When applying the same hPSMC and simulation analyses within each species, we found gene flow to have also continued long after the initial divergence of and cessation of panmixia within the lineages. We found that white rhinoceros last experienced north-south gene flow approximately 200-300 Ka after the species diverged into northern and southern populations, at some point in time between 100-220 Ka (Figure 3B). Gene flow between eastern and southern populations continued for even longer after divergence in the black rhinoceros (~400-500 Ka), until ceasing more recently between 30-130 Ka (Figure 3C).

Evaluating D-statistics in the presence of ancestral gene flow

To further evaluate our seemingly unlikely D-statistics results, we ran simulations in 1MB blocks based on a simple model specifying ancestral gene flow between the ancestral black and white rhinoceros lineages prior to their divergence into their respective subspecies (Figure 4) and ran D-statistics on these simulations. Although results differed based on specified ancestral migration rates (Supplementary Tables S8-S10), we found significant Z-scores indicating post-divergence gene flow between the southern black rhinoceros and both white rhinoceros subspecies as well as between the northern white and both black rhinoceros subspecies, even though we did not model subspecies-level gene flow.

Discussion

In this study we generated the first reference genome assembly for the critically endangered black rhinoceros, from an individual belonging to the species' southern-most population. We analysed this southern black rhinoceros reference genome, together with nuclear genomes from eastern black (Kenya), southern white (South Africa), and northern white (South Sudan) rhinoceros, to uncover the evolutionary history of and relationships between the two species of African rhinoceros.

Genomic diversity, Pleistocene declines and expansion

Levels of genome wide heterozygosity support recent population histories of anthropogenically-mediated decline in the black and white rhinoceros, as shown previously (Moodley *et al.* 2017, Tunstall *et al.* 2018, Moodley *et al.* 2018). All four rhinoceros genomes showed a mid-Pleistocene decline, which may have been associated with a gradual cooling of

the earth at the beginning of the Pliocene. However, a subsequent population expansion of all genomes is not consistent with the paleoclimatic record, since this was just after the time when glacial cycles became more severe (<800 Ka, Figure 2C). Rather, the increase in effective population size occurred at the point at which panmixia in both species ended (Figure 2). It is possible that the 520 Ka expansion in black rhinoceros was associated with an interglacial cycle, whereas the 440 Ka white rhinoceros expansion could be associated with an interglacial, but both species effective sizes were inferred to have expanded, regardless of subsequent glacial cycles, until about 240 Ka. Alternatively, the early evolution of additional genetic substructure within diverging regional populations of each species, followed by their isolation during unfavourable climatic periods, could also have inflated effective population sizes, even if census sizes remained stable (Mazet *et al.* 2015, Mazet *et al.* 2016). Although most mtDNA lineages to have evolved in the white rhinoceros were already extinct by the Holocene (Moodley *et al.* 2018), in the black rhinoceros, mtDNA is highly structured with both eastern and southern lineages (Moodley *et al.* 2017), lending weight to this interpretation.

At approximately 240 Ka, southern black and white rhinoceros underwent a population decline, followed by northern white and eastern black rhinoceros at 180 Ka. Until now, only the genomes of bonobos, Nigeria-Cameroon chimpanzees, the spectacled bear and east African baboons show a similar Middle Pleistocene decline around 150-200 Ka (Prado-Martinez *et al.* 2013, Kumar *et al.* 2017, Rogers *et al.* 2019). It is interesting that these declines in both African rhinoceros coincide with the emergence of modern humans. A similarly sharp demographic decline was also inferred for the Sumatran rhinoceros, but more recently at about 100 Ka (Mays *et al.* 2018), coinciding with the appearance of humans in Asia. It is also intriguing that southern populations of both rhinoceros species decline before populations in the north and east, as it could imply differential levels of population pressure across Africa.

Divergence, the end of panmixia and the cessation of gene flow within each species

Within-species autosomal divergence times for black (641 Ka) and white (578 Ka) rhinoceros were consistently about 100 Ka older than the times at which panmixia is inferred to have ceased for each species using PSMC. Although divergence times do not account for demographic events or gene flow, values are remarkably similar considering their different methods of inference. The reported divergence times are also within the confidence limits of mtDNA data for white rhinoceros (Harley *et al.* 2016, Moodley *et al.* 2018), but not for black rhinoceros, where mtDNA divergence between southern and eastern black rhinoceros was inferred to be significantly more ancient (920-3,575 Ka, Moodley *et al.* 2017). It is possible that values for the end of panmixia may have been downwardly biased as populations of both species underwent very similar demographic expansion trajectories after their PSMC curves became dissociated (Figure 2B). However, divergence and the end of panmixia occurred long before the final cessation of gene flow within both species (Figure 3B/C), indicating ongoing secondary contact, potentially during phases of demographic expansion during the last 400,000 years. While the two white rhinoceros populations appear to have come into secondary contact less recently (100-220 Ka), this estimate is still consistent with gene flow during the last glacial period, as recently inferred from microsatellite data (Moodley *et al.* 2018). The eastern and southern black rhinoceros on the other hand, appear to have come into more recent genetic contact across the

Zambezi valley (Figure 3C). The Zambezi's paleo-upper and -lower reaches were joined by river capture between 125–150 Ka (Moore and Larkin 2001), and while our results appear to contrast with a strong mtDNA and microsatellite discontinuity on either side of this river (Moodley *et al.* 2017), at least one East African mtDNA haplotype was sampled on the southern bank of the Zambezi, and one southern African haplotype was sampled north of the Zambezi, hinting that although the river may have acted to maintain the genetic integrity of populations to its north and south, it was also periodically fordable for black rhinoceros. In summary, these results suggest that the period required between divergence and the end of panmixia to the cessation of gene flow is dependent on how frequently climatic changes were able to bring diverging populations into secondary contact, and in Africa, the expansion and contraction of habitats with glacial cycles appears to have maintained gene flow long after population divergence. Although post-divergence gene flow was observed previously in other taxa, including rhinoceros (Wang *et al*. 1997, Won and Hey 2005, Lee and Edwards 2008, Moodley *et al*. 2018), our results provide yet another cautionary note in evolutionary and conservation inference, that estimated times of divergence do not necessarily correlate with the cessation of genetic contact.

Gene flow between African rhinoceros species

Both D-statistics and Dfoil suggested gene flow between the two African rhinoceros at the subspecies level, that is, within the 63 Ka gap after the divergence of eastern and southern black rhinoceros, but before the divergence of northern and southern white rhinoceros. While this scenario might be plausible, it is highly unlikely that diverging eastern and southern black rhinoceros populations both came into secondary contact with the ancestral white rhinoceros population within this short space of time. These putatively unrealistic results may have arisen due to caveats of the D-statistics analysis itself, our dataset, the biology of the individuals involved, or a combination of all three. Possible explanations for false positive signs of introgression could include ancestral population structure, which produce deviations from expectations based solely on incomplete lineage sorting (ILS, Slatkin and Pollack 2008), introgression from unsampled or extinct "ghost" lineages, differences in relative population sizes of the lineages or in the timing of gene flow events, or different evolutionary rates or sequencing errors between the H1 and H2 individuals and therefore differential divergence of the extant lineages from their common ancestor (Zheng and Janke, 2018). We also considered whether our taxa (both ingroup and outgroup) were too divergent from one another, which would lead to evolutionary signals being overwhelmed by noise caused by multiple substitutions and substitution saturation, although previous studies have shown D-statistics to be robust to these factors (Zheng and Janke, 2018). Owing to these uncertainties in interpreting our D-statistics and Dfoil results, we performed a number of additional analyses to infer gene flow including; D3, F3-statistics, F4-statistics, Treemix, and hPSMC. While each method has its own caveats, the combination of all methods provides us with a suite of information to aid in the interpretation of our results. Unlike D-statistics and Dfoil, D3 and F3-statistics found no evidence for recent, subspecies-level gene flow. F4-statistics suggested gene flow had occurred between the four African lineages but could not be

used to determine which lineages were involved, and Treemix produced ambiguous results that may reflect its unsuitability for our dataset.

One potential explanation for these contradictory results was uncovered via hPSMC analyses. hPSMC showed that gene flow between the two species ceased relatively early, during the mid-late Pliocene between 3.3 and 4.1 Ma, long before divergence into subspecies lineages. This result was consistent between genomes and regardless of which reference (black or white rhinoceros) was used for mapping. The lack of contiguous regions of gene flow inferred by Dfoil also suggest an absence of recent interspecific migration, as recent secondary contact between black and white rhinoceros subspecies would have resulted in larger and more continuous tracts of introgression (Pool and Nielsen, 2009). Instead, recombination and ILS appears to have broken up such tracts into mainly singleton windows, indicating that the last gene flow event(s) between the two species must have occurred prior to the divergence of subspecies lineages, thus corroborating hPSMC results, which suggest the development of reproductive isolation during the Pliocene.

Finally, we tested the idea that Pliocene gene flow between the ancestral black and ancestral white rhinoceros lineages, could have resulted in false positive signatures of recent gene flow. Although the jackknifing significance test should be robust to such a case, we addressed this possibility by running D-statistics on simulated data, generated using a simple model and various levels of ancestral gene flow, followed by subspecies divergence. We found significant Z-scores for subspecies level gene flow, even though the only migration events simulated were those between the ancestral lineages (Supplementary Tables S8-S10). These simulation results, together with hPSMC and a lack of contiguous gene flow tracts strongly suggest that gene flow between the black and white rhinoceros lineages ceased during the Pliocene, long before the divergence of their subspecies (Figure 4), and that genetic signatures of this ancient introgression are differentially present in our sampled African rhinoceros genomes due to ILS. We caution that future studies which infer recent interspecific gene flow employ a suite of independent analyses, including simulations, to rule out the possibility that gene flow occurred between ancestral lineages, with subsequent random genetic drift leading to ILS.

The evolution of specialisation

Our results suggest that the African rhinoceros precursors *D. praecox* and *C. mauritanicum* may still have been able to exchange genes with each other when they came into secondary contact at Kanapoi in Kenya 4 Ma (Figure 4). This was supported by the hPSMC analysis, which shows that gene flow ceased up to two million years after the initial divergence of ancestral *Diceros* and *Ceratotherium* lineages. On the other hand, if gene flow was not possible at Kanapoi, it was likely because reproductive isolation had only just become fully developed between the two species.

The evolutionary consequences of this ancestral gene flow are intriguing. We analysed the segments of DNA inferred to be exchanged between the two species, but did not find any significantly enriched GO terms, leading us to hypothesise that there was little to no evidence for adaptive introgression, as observed in other recent studies for example, Pardo-Diaz *et al.* (2012) and Dasmahapatra *et al.* (2012). Perhaps this is not surprising since a classical view (Mayr 1942) would predict that periods of secondary contact and ongoing gene flow between *D. praecox* and

C. mauritanicum prior to 4 Ma may have continually undermined the diverging evolutionary trends of both lineages towards browsing and grazing, respectively. The rhinoceros-rich fossil record of the mid-late Pliocene provides some evidence to support this view, because despite over a million years since their initial divergence, both species maintained their ancestral mixed feeding state throughout most of the Pliocene (Geraads 2017). So phenotypically similar were *D. praecox* and *C. mauritanicum* during this period, that palaeontologists often misidentified one species for the other (Geraads 2005, Geraads 2010). On the other hand, there is also extensive theoretical (Barton 1979, Barton 1987) and empirical (McCracken *et al.* 2009, Hohenlohe *et al.* 2012, Poelstra *et al.* 2014) evidence that adaptation can occur even in the face of gene flow, when introgressing alleles confer a selective advantage, with advantageous loci often in tight linkage disequilibrium, or if hybrid fitness is low (Barton and Hewitt 1985). The fact that both *D. praecox* and *C. mauritanicum* had developed some level of specialisation prior to the onset of reproductive isolation, suggests that adaptation may have been occurring despite Pliocene gene flow. However, eventual reproductive isolation between the two species likely resulted from an accumulation of larger numbers of loci under selection (Barton and Hewitt, 1989). The evolution of fully specialised browsing and grazing African rhinoceros species could only have occurred during the critical phase after reproductive isolation between them was established (3.3-4.1 Ma), but before the internal splits within each species (500-600 Ka, Figure 4). The fossil emergence of phenotypically modern black (2.5 Ma) and white (1.8 Ma) rhinoceros falls exactly within this interval. The timing of these fossil emergences suggests that the more severe and shifting paleoclimates of the Pleistocene provided the heterogeneity of environments that ultimately drove the evolution of obligate feeding strategies in African rhinoceros species.

Methods

Establishing the black rhinoceros reference assembly

To ensure a straightforward assembly of our reference genome, we undertook to sample from a more genetically depauperate black rhinoceros population where heterozygous sites are likely to be more sparsely distributed across the genome. Of the five remaining aboriginal stocks in Africa, KwaZulu-Natal (South Africa) contains the lowest levels of genetic diversity (Anderson-Lederer 2011, Moodley *et al.* 2017) owing to an early 20th Century population collapse. The KwaZulu population has since recovered to over 2,000 individuals (Emslie *et al.* 2016). We obtained ear notches taken during routine management of a male and female black rhinoceros (*D. b. minor*) from the Zululand Rhino Reserve, near the town of Mkhuze in KwaZulu-Natal, South Africa. Both samples were taken by a veterinarian under an ordinary permit (OP 4368/2015) from the provincial authority Ezemvelo KZN Wildlife and preserved in 99% alcohol. The samples were then couriered to the Naturhistoriska riksmuseet under CITES permit number 51491-15 where DNA was extracted with a Kingfisher Duo (Thermofisher Scientific) using the Cell and Tissue DNA Kit. The best quality sample, a male individual (SAMN14912225), was selected for genome sequencing. We employed an exhaustive sequencing strategy, establishing two short insert DNA libraries of 180bp and 650bp as well as three mate-pair DNA libraries of 3 KB, 5 KB and 20 KB fragment size. The libraries were sequenced on the Illumina HiSeq X platform, with one lane for each of the short-insert libraries and one lane for a pool of the three mate-pair libraries. We then *de novo* assembled these reads using Allpaths-LG v52485 (Gnerre *et al.* 2011) according to the method described by Pujolar *et al.* (2018).

We further improved our reference assembly by generating three Chicago libraries (Putnam *et al.* 2016) from the reference sample at Dovetail Genomics (Santa Cruz, CA). This method uses *in vitro* reconstituted chromatin to achieve 3D folding of the DNA. The folded DNA is then cut using an endonuclease and subsequently ligated back together. The advantage of this method is that some links are made between regions of the same DNA strand up to hundreds of KB apart, due to their close proximity in the 3D folding. The Chicago libraries were assembled with Dovetail's Hi-Rise scaffolding pipeline. To assess the gene-content of the assembly, BUSCO v3.0.2 was run using its set of 4,104 single-copy mammalian orthologs (Simão et. al. 2015).

Genome annotation

Next, we carried out repeat and gene annotation. To do so, we first masked repeats in the genome using a combination of *ab initio* repeat finding and homology-based repeat annotation using RepeatModeler (http://www.repeatmasker.org) and RepeatMasker (http://www.repeatmasker.org), respectively. For homology-based repeat annotation we used the mammal repeat consensus sequences from Repbase (Bao, *et al.* 2015). For the gene annotation, we did not mask simple repeats beforehand to improve mapping during the homology-based annotation implemented in Maker2 (Holt and Yandell 2011). The gene annotation was performed using a combination of *ab-initio* gene prediction (using SNAP (Korf 2004) and Augustus (Stanke and Waack 2003)) and homology-based gene annotation using Maker2. We used protein annotations of the horse (EquCab2.0; GCF_000002305.2), the white rhinoceros (CerSimSim1.0;

GCF_000283155.1) and human (GRCh38; GCA_000001405.37) for the homology-based gene annotation step. This resulted in the annotation of 19,914 genes.

Raw data processing and mapping

To investigate the evolutionary history of African rhinoceros, we analysed the South African black rhinoceros reference assembly together with the Broad Institute's white rhinoceros reference genome (CerSimSim1.0), obtained from a female southern white rhinoceros (*C. s. simum*, iMfolozi, Studbook# 159) which was wild caught in 1963 at iMfolozi Game Reserve, South Africa. To include as much of the variation within each species as possible, we further resequenced the genomes of a female East African black rhinoceros (*D. b. michaeli*, Sally, Studbook# 78, SAMN14911588), wild caught in 1950 in the Kibwezi District, southern Kenya; and a female northern white rhinoceros (*C. s. cottoni*, Nola, Studbook# 374, SAMN14911569), wild caught in 1974 in the Shambe Region of Sudan, now South Sudan. Both Kenyan and South Sudanese samples were obtained from the San Diego Zoo, and sequencing to 16-fold coverage was carried out at the Broad Institute. As an outgroup genome, we used the recently sequenced Sumatran rhinoceros (*Dicerorhinus sumatrensis*, Mays *et al.* 2018), which diverged from the clade containing African rhinoceros about 18 million years ago (Margaryan *et al*. 2020).

Raw reads were all treated comparably before being mapped to a specific reference genome. We used Cutadapt v1.8.1 (Martin 2011) to trim Illumina adapter sequences from the ends of reads and remove reads shorter than 30bp. We then merged overlapping read pairs using FLASH v1.2.1 (Magoč and Salzberg 2011). We mapped the resultant reads of the five individuals used in the study to their respective reference sequences, unless otherwise specified, using BWAv0.7.15 (Li and Durbin 2009) and processed the mapped reads further using SAMtools v1.3.1 (Li *et al.* 2009). We mapped both the East African black rhinoceros and South African black rhinoceros to the newly assembled black rhinoceros genome, both the northern white rhinoceros and the southern white rhinoceros to the published southern white rhinoceros genome, and the Sumatran rhinoceros to the Sumatran rhinoceros genome (GCA_002844835.1).

Genetic variation

We estimated autosomal heterozygosity from each of the four African rhinoceros individuals. To determine which scaffolds were most likely autosomal in origin, we found putative sex chromosome scaffolds for each of the rhinoceros reference genomes and removed them from future analyses. We found putative sex chromosome scaffolds through synteny by aligning the rhinoceros reference genomes to the Horse X (Genbank accession: CM000408.2) and Human Y (Genbank accession: NC_000024.10) chromosomes. Alignments were performed using satsuma synteny (Grabherr *et al.* 2010) and utilising default parameters. To adjust for biases in heterozygosity levels that could arise due to different global coverages between the genomes of the individuals being investigated, we subsampled all of the resultant alignments down to that of the lowest coverage individual, 16-fold, using SAMtools. We then estimated the autosomal heterozygosity from all scaffolds above 100 KB in length, using sample allele frequencies in ANGSDv0.913 (Korneliussen *et al.* 2014), taking genotype likelihoods into account and specifying the following filters -minq 25 -minmapq 25 -uniqueOnly 1 -baq 1 -remove_bads 1. We calculated the standard deviation for each of the heterozygosity estimates by performing the realSFS

function in the ANGSD package in independent 20MB windows of covered bases (-nSites 20,000,000).

Genome divergence and mutation rate

To estimate the mutation rate per generation for each species, we computed pairwise distances between the black and white rhinoceros autosomes twice independently: once with all four African rhinoceros mapped to the black rhinoceros reference genome, and again will the four genomes mapped to the white rhinoceros reference, and we took the average of the results. That is, the average distance between the eastern black + southern white, eastern black + northern white, southern black + southern white, southern black + northern white. We computed pairwise distances using a consensus base IBS approach (-doIBS 2) in ANGSD and applying the filters -minQ 25 -minmapq 25 -uniqueonly 1 -remove_bads 1. Using this information, we then computed the mutation rate per generation assuming a genome-wide strict molecular clock and using the following equation: mutation rate = pairwise distance x generation time/2 x divergence time. We assumed a divergence time coinciding with that of the Miocene/Pliocene boundary (5.3 Ma) as the stem lineage (*C. neumayri*) was common during the late Miocene (7-9 Ma) but had already split into *D. praecox* and *C. mauritanicum* lineages by the Pliocene (Geraads 2005, 2017). A generation time of 24 years was assumed for the black rhinoceros (Moodley et al, 2017) and 27 years for the white rhinoceros (Moodley *et al.* 2018). Moreover, we used the per year mutation rate calculated by comparing the black and white rhinoceros to estimate the within species divergence dates based on the within species average pairwise distances when mapping to the conspecific reference genome.

Demographic analyses

We ran demographic analyses on the diploid genomes of all four African rhinoceros individuals using Pairwise Sequentially Markovian Coalescent model (PSMC) (Li and Durbin 2011). Using this method, it is possible to infer changes in effective population size through time for diploid (high coverage) genomes from the distribution of its heterozygous sites across the genome. We called diploid genome sequences using SAMtools and bcftools (Narasimhan *et al.* 2016) specifying a minimum quality score of 20 and minimum coverage of 10. We removed scaffolds found to align to sex chromosomes in the previous step and scaffolds shorter than 100KB. We ran PSMC specifying atomic intervals previously shown to be suitable for human datasets (4+25*2+4+6) and performed 100 bootstrap replicates to investigate support for the resultant demography. We overlaid the resultant PSMC plots as the point in time in which the demographic trajectories of two individuals diverges can be interpreted as a rough measure of the end of panmixia in that species. Moreover, as one of our individuals (northern white rhino) was only ~16-fold coverage, we downsampled our (~35-fold coverage) southern black rhinoceros genome to 16-fold coverage to investigate the effect this may have on our inferences. We ran a PSMC analysis on the downsampled genome and compared it to the results recovered for the same genome using the much higher coverage data.

Inter and intra-specific post-divergence gene flow

For the gene flow analyses, we mapped the raw reads from the four African species to the Sumatran rhinoceros following the same methods mentioned above to avoid any ascertainment bias that may occur when mapping to an ingroup African rhinoceros species (Westbury *et al.* 2019). We performed multiple different analyses to test for post divergence gene flow between African rhinoceros. First, we implemented the four-taxon ABBA/BABA or D-statistics approach (Durand *et al.* 2011) with ANGSD. We called bases using a random base call (-doAbbababa 1), only considered scaffolds over 100KB in length, specified the Sumatran rhinoceros as outgroup, and applied the following filters; -minMapQ 25, -minQ 25, -uniqueOnly 1, -remove_bads 1. We also adjusted quality scores around indels (-baq 1) (Li 2011). ANGSD performs all possible combinations but we only investigated the output with conspecifics in the H1 and H2 positions and an individual from the other species in the H3. Any other combination would go against the species tree and therefore produce invalid signs of admixture driven by more recent common ancestry as opposed to true admixture. To investigate the significance of our result, we performed a weighted block jackknife test using 5MB non-overlapping blocks. D-values more than three standard errors different from zero ($-3 < Z > 3$) were considered as statistically significant.

Following the D-statistics, we implemented Dfoil (Pease and Hahn 2015), a more detailed, expanded version of D-statistics using five-taxa to test for gene flow, with the Sumatran rhinoceros as an outgroup. Dfoil implements four independent D-statistics in a sliding window fashion which are then combined before inferences are made. This has the advantage over the four-taxon test in that it can infer the direction of gene flow and uncover whether gene flow occurred between ancestral lineages. For this analysis, we also mapped all rhinoceros to the Sumatran rhinoceros. We then constructed fasta files for each individual using ANGSD and specifying maximum effective base depth (-doFasta 3) and the following parameters: -minMapQ 25, -minQ 25, -uniqueOnly 1, -remove_bads 1. Additionally, we removed all scaffolds shorter than 1MB and trimmed the ends of the remaining scaffolds down to the nearest 100KB, leaving us with 937.2MB. The resultant fasta files were converted into an mvf file (ConvertFasta2MVF) which was then converted into three independent Dfoil input files (CalcPatternCount) of window sizes 100KB, 50KB, and 20KB with mvftools (Pease and Rosenzweig 2018). Regions showing signs of admixture between the ancestral white rhinoceros and either the southern or eastern black rhinoceros were extracted and compared between window sizes. We cross referenced the introgressed genomic segments consistently showing signs of admixture despite window size against the white rhinoceros annotation to uncover putative protein coding genes in these regions. We then tested for GO enrichment terms with GOrilla (Eden *et al.* 2009). We further investigated the contiguity of introgressed regions by extracting all regions showing any signs of introgression, regardless of direction, and investigating how long the stretches of these introgression windows were for all window sizes independently.

To further test for admixture, we implemented D3 (Hahn and Hibbins 2019), a three taxon test for introgression that makes use of pairwise distances and does not require an outgroup genome using the topology ((A, B), C) and the equation (BC-AC)/(BC+AC). We computed pairwise distances between the four African rhinoceros based on a consensus base using ANGSD -doIBS 2 and the following parameters; -makeMatrix 1 -uniqueOnly 1 -remove_bads 1 -doMajorMinor 1 -minInd 4 -GL 1 -setMinDepthInd 5 -minmapq 25 -minq 25. We did this twice

independently specifying two different non-overlapping window sizes (100KB, and 1MB) to test the significance of our results. We calculated a p-value for each comparison to evaluate the difference from 0 by calculating the mean, standard deviation, and assuming a normal distribution in R v3.6.0 (R Core Team, 2019) using the pnorm function. We also ran Treemix v 1.13 (Pickrell and Pritchard 2012) with various migration edges as well as the threepop and fourpop tests, otherwise known as the F3- and F4-statistics (Reich *et al*. 2009, Keinan *et al*. 2007), to determine the presence or absence of gene flow among African rhinoceros using the software available in the Treemix toolsuite (see Supplementary methods).

Timing of reproductive isolation and the cessation of gene flow

To add a temporal element to the onset of reproductive isolation and the cessation of gene flow between African rhinoceros, we used the F1 hybrid pairwise sequentially Markovian coalescent model, hPSMC (Cahill *et al.* 2016). To address whether ascertainment bias may have played a role in our results, we performed this analysis twice independently for the between species comparisons, white *vs* black rhinoceros, once using the black rhinoceros as reference genome, and once using the white rhinoceros as reference genome. Within species comparisons were only computed once using the conspecific genome as reference. We constructed haploid consensus sequences for the four individuals using ANGSD by considering the base with the highest effective depth, the following quality filters; -minQ 25, -minmapq 25, -uniqueonly 1, -remove_bads 1, and only considering autosomes and scaffolds over 100KB. We merged these resultant haploid consensus sequences together into a pseudo diploid sequence using the hPSMC tool suite. These were then run through PSMC and plotted using an intermediate mutation rate per generation and generation time. When comparing the black and white rhinoceros we used a generation time of 25.5 years and a mutation rate of 2.2×10^{-8} mutations per generation. When comparing within species, we used intraspecific mutation rates and generation times. From this, we manually estimated the pre-divergence Ne by outputting the text file (-R) using the plot script from the PSMC tool suite and looking into the output text file. Using the pre-divergence Ne estimated from this output, we then ran simulations to infer the confidence intervals using Ms (Hudson 2002) with the hPSMC_quantify_split_time.py python script from the hPSMC tool suite, while specifying the time windows we wanted to simulate, and the remaining parameters as default. When comparing black and white rhinoceros, we estimated a pre-divergence Ne of 60,000 and ran simulations using divergence times between 3,000,000-7,000,000 years in 100,000 year intervals. When comparing northern and southern white rhinoceros, we estimated a pre-divergence Ne of 7,000 and ran simulations using divergence times between 50,000-450,000 years in 10,000 year intervals. When comparing eastern and southern black rhinoceros, we estimated a pre-divergence Ne of 13,000 and ran simulations using divergence times between 0-400,000 years in 10,000 year intervals. Results were plotted and the simulations with an exponential increase in Ne closest to the real data, within 1.5x and 10x of the pre-divergence Ne, were taken as the time interval in which gene flow stopped. We considered the portion between 1.5x and 10x of the pre-divergence Ne as suggested by the original manuscript. This was suggested in order to capture the portion of the hPSMC plot most influenced by the divergence event. The lower bound is set to control for pre-divergence increases in population size and the

upper bound is to avoid exploring parameter space in which little information is present (Cahill *et al.* 2016).

Evaluating the role of ancestral gene flow on D-statistics results

In order to evaluate the influence of ancestral gene flow on D-statistics results, we ran a simple model simulation in MSMS (Ewing and Hermisson, 2010) specifying gene flow between the ancestral lineages as shown in Figure 4. This was done using the following command: msms 82 500 -I 5 2 20 20 20 20 0 -t 1760 -r 352 -ej 0.2375 5 4 -ej 0.2375125 3 2 -em 1.875 4 2 {migration rate} -em 1.875 2 4 {migration rate} -ej 2.5 4 2 -ej 12.5 2 1. In brief we specified window sizes of 1MB, an effective population size of 20,000 for all five populations, with constant population sizes, a generational mutation rate of $2.2x10^{-8}$ and a recombination rate one fifth of the mutation rate ($4.4x10^{-9}$), three independent runs of 20,000 windows, each with different migration rates (m = 0.5, 1, and 2), a divergence time of 200,000 generations between the black and white rhinos, the end of gene flow between the black and white rhinoceros as 150,000 generations, the within species divergence as 19,000 generations and assuming a generation time of 25.5 years. The output of the simulations were then run through a custom python script which calculated the D-score for each 1MB window independently. Finally, we performed a block jackknifing approach with the resample library in R v3.6.0 to test for significance of the results.

Acknowledgements

We thank Dr. Mike Toft of Kifaru Wildlife Veterinary Services and Dr. Lourens Swanepoel for making Zululand black rhinoceros samples available for this study, as well as Kim Labuschagne for help with export permits. Sequencing was funded by a grant from the International Rhino Foundation (grant nr. R-2014-1) to MWB. The authors acknowledge support from Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, the National Genomics Infrastructure funded by the Swedish Research Council, and Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing as well as de novo assembly of the data, and access to the UPPMAX computational infrastructure. The construction and sequencing of Chicago libraries and the Hi-Rise assembly were funded through a Dovetail Prize awarded to LD from Dovetail Genomics, LLC. LD also acknowledges funding from FORMAS (grant nr. 2015-676). YM acknowledges funding from the University of Venda (grant nr. SMNS/15/ZOO/05).

References

Bao W, Kojima KK., & Kohany O (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, **6**(1), 11.

Barnosky AD, Koch PL, Feranec RS, Wing SL, Shabel AB (2004) Assessing the causes of late Pleistocene extinctions on the continents. *Science* **306**(5693):70-5.

Barton NH (1979) Gene flow past a cline. *Heredity*. Dec;43(3):333-9.

Barton NH (1987) The probability of establishment of an advantageous mutant in a subdivided population. *Genetics Research*. Aug;50(1):35-40.

711 Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual review of Ecology and*
712 *Systematics*. Nov;16(1):113-48.
713 Barton NH, Hewitt GM (1989) Adaptation, speciation and hybrid zones. *Nature*.
714 Oct;341(6242):497-503.
715 Beichman AC, Phung TN, Lohmueller KE (2017) Comparison of single genome and allele
716 frequency data reveals discordant demographic histories. *G3: Genes, Genomes, Genetics*.
717 Nov 1;7(11):3605-20.
718 Cahill JA., Soares AER, Green RE, Shapiro B (2016) Inferring Species Divergence Times Using
719 Pairwise Sequential Markovian Coalescent Modelling and Low-Coverage Genomic Data.
720 *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*
721 **371** (1699): 20150138.
722 Chilvers, B (1990) Rhino's last stand in Africa. *Rhino and Elephant Journal* **3**: 12-19, figs. 1-3.
723 Dasmahapatra KK, Walters JR, Briscoe AD, Davey JW, Whibley A, Nadeau NJ, Zimin AV, Hughes
724 DS, Ferguson LC, Martin SH. & Salazar C (2012) Butterfly genome reveals promiscuous
725 exchange of mimicry adaptations among species. *Nature* **487**(7405), p.94.
726 deMenocal PB (2004) African climate change and faunal evolution during the Pliocene–
727 Pleistocene. *Earth and Planetary Science Letters* **220**(1-2):3-24.
728 Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and
729 visualization of enriched GO terms in ranked gene lists. BMC bioinformatics. Dec;10(1):48.
730 Emslie RH, Milliken T, Talukdar B, Ellis S, Adcock K, Knight M. 2016. African and Asian
731 Rhinoceroses – Status, Conservation and Trade A report from the IUCN Species Survival
732 Commission (IUCN SSC) African and Asian Rhino Specialist Groups and TRAFFIC to the
733 CITES Secretariat pursuant to Resolution Conf. 9.14 (Rev. CoP15).
734 Eran E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009). GOrilla: A Tool for Discovery and
735 Visualization of Enriched GO Terms in Ranked Gene Lists. *BMC Bioinformatics* **10**
736 (February): 48.
737 Ewing G, Hermisson J (2010) MSMS: a coalescent simulation program including recombination,
738 demographic structure and selection at a single locus. *Bioinformatics*. Aug 15;26(16):2064-
739 5.
740 Geraads D (2005) Pliocene Rhinocerotidae (Mammalia) from Hadar and Dikka (Lower Awash,
741 Ethiopia), and a revision of the origin of modern rhinos. *Journal of Vertebrate Paleontology*
742 **25** (2): 451–61.
743 Geraads D (2017) Perissodactyla (Rhinocerotidae and Equidae) from Kanapoi. *Journal of Human*
744 *Evolution*, September. https://doi.org/10.1016/j.jhevol.2017.07.013.
745 Geraads D & Spassov N (2009) Rhinocerotidae (Mammalia) from the late Miocene of Bulgaria.
746 *Palaeontographica: Beiträge zur Naturgeschichte der Vorzeit. Abt. A, Palaeozoologie,*
747 *Stratigraphie*, *287*, 99-122.
748 Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, & Berlin AM. (2011).
749 High-quality draft assemblies of mammalian genomes from massively parallel sequence
750 data. *Proceedings of the National Academy of Sciences* **108**(4), 1513-1518.
751 Giaourtsakis IX, Pehlevan C, Haile-Selassie Y (2009) 14. Rhinocerotidae. In: Haile-Selassie Y,
752 WoldeGabriel G. (Eds.), Ardipithecus kadabba, *Late Miocene evidence from the Middle*
753 *Awash, Ethiopia*. University of California Press, Berkeley, pp. 429e468.

754 Gissi C, Reyes A, Pesole G, Saccone C (2000) Lineage-specific evolutionary rate in mammalian
755 mtDNA. *Molecular Biology and Evolution* **17**(7):1022-31.
756 Grabherr MG, Russell P, Meyer M, Mauceli E, Alföldi J, di Palma F, & Lindblad-Toh K (2010)
757 Genome-Wide Synteny through Highly Sensitive Sequence Alignment: Satsuma.
758 *Bioinformatics* **26** (9): 1145–51.
759 Harris, JM (1983) Family Rhinocerotidae; pp. 130–156 in JM Harris (ed.), *Koobi Fora Research*
760 *Project: Volume 2. The Fossil Ungulates: Proboscidea, Perissodactyla and Suidae.*
761 Clarendon Press, Oxford.
762 Hillman K (1980) African rhinoceros. *WWF Yearbook 1979-1980*: pp. 69-75.
763 Hahn MW, Hibbins MS (2019) A three-sample test for introgression. *Molecular biology and*
764 *evolution*. Dec 1;36(12):2878-82.
765 Hohenlohe PA, Bassham S, Currey M, Cresko WA (2012) Extensive linkage disequilibrium and
766 parallel adaptive divergence across threespine stickleback genomes. *Philosophical*
767 *Transactions of the Royal Society B: Biological Sciences*. Feb 5;367(1587):395-408.
768 Holt C & Yandell M (2011). MAKER2: an annotation pipeline and genome-database management
769 tool for second-generation genome projects. *BMC bioinformatics* **12**, 491.
770 Hooijer DA (1969) Pleistocene East African rhinoceroses; pp. 71–98 in LSB. Leakey (ed.), *Fossil*
771 *Vertebrates of Africa*, vol. 1. Academic Press, London.
772 Hooijer, DA (1972) A late Pliocene rhinoceros from Langebaanweg, Cape Province. *Annals of the*
773 *South African Museum* **59**:151–191.
774 Huaiyu M, Muruganujan A, Casagrande JT & Thomas PD (2013) Large-Scale Gene Function
775 Analysis with the PANTHER Classification System. *Nature Protocols* **8** (8): 1551–66.
776 Hudson RR (2002) Generating Samples under a Wright–Fisher Neutral Model of Genetic
777 Variation. *Bioinformatics* **18**, 2: 337–338
778 Keinan A, Mullikin JC, Patterson N and Reich D (2007) Measurement of the human allele
779 frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans.
780 *Nat Genet*, 39(10):1251–5.
781 Kennaugh A (2015) Rhino Rage: What is driving illegal consumer demand for rhino horn. *Natural*
782 *Resources Defense Council*; New York, pp. 1-23
783 Korf I (2004) Gene Finding in Novel Genomes. *BMC Bioinformatics* **5** (May): 59.
784 Korneliussen TS, Albrechtsen A & Nielsen R (2014) ANGSD: Analysis of Next Generation
785 Sequencing Data. *BMC Bioinformatics* **15** (November): 356.
786 Kumar V, Lammers F, Bidon T, Pfenninger M, Kolter L, Nilsson M, Janke A (2017). The
787 evolutionary history of bears is characterized by gene flow across species. *Scientific Reports*,
788 **7**, 46487.
789 Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF & Prohaska SJ (2011) Proteinortho:
790 Detection of (Co-)orthologs in Large-Scale Analysis. *BMC Bioinformatics* **12** (1): 124.
791 Lee JY & Edwards SV (2008) Divergence across Australia's Carpentarian barrier: statistical
792 phylogeography of the red-backed fairy wren (*Malurus melanocephalus*). *Evolution:*
793 *International Journal of Organic Evolution* Dec;62(12):3117-34.
794 Li H (2011) Improving SNP Discovery by Base Alignment Quality. *Bioinformatics* **27** (8): 1157–
795 58.
796 Li H & Durbin R (2009) Fast and Accurate Short Read Alignment with Burrows–Wheeler

797 Transform. *Bioinformatics* **25** (14): 1754–60.
798 Li H & Durbin R (2011) Inference of Human Population History from Individual Whole-Genome
799 Sequences. *Nature* **475** (7357): 493–96.
800 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R &
801 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map
802 Format and SAMtools. *Bioinformatics* **25** (16): 2078–79.
803 Löytynoja A (2014) Phylogeny-Aware Alignment with PRANK. *Methods in Molecular Biology*
804 **1079**: 155–70.
805 McCracken KG, Bulgarella M, Johnson KP, Kuhner MK, Trucco J, Valqui TH, Wilson RE, Peters
806 JL (2009) Gene flow in the face of countervailing selection: adaptation to high-altitude
807 hypoxia in the βA hemoglobin subunit of yellow-billed pintails in the Andes. Molecular Biology
808 and Evolution. Apr 1;26(4):815-27.
809 Magoč T & Salzberg SL (2011) FLASH: Fast Length Adjustment of Short Reads to Improve
810 Genome Assemblies. *Bioinformatics* **27** (21): 2957–63.
811 Margaryan A, Sinding MH, Liu S, Vieira FG, Chan YL, Nathan SK, Moodley Y, Bruford MW, Gilbert
812 MT (2020) Recent mitochondrial lineage extinction in the critically endangered Javan
813 rhinoceros. *Zoological Journal of the Linnean Society*.
814 Martin M (2011) Cutadapt Removes Adapter Sequences from High-Throughput Sequencing
815 Reads. *EMBnet.journal* **17** (1): 10–12.
816 Mazet O, Rodríguez W, Chikhi L (2015) Demographic inference using genetic data from a single
817 individual: Separating population size variation from population structure. *Theoretical*
818 *population biology* **1**:104:46-58.
819 Mazet O, Rodríguez W, Grusea S, Boitard S, Chikhi L (2016) On the importance of being
820 structured: instantaneous coalescence rates and human evolution—lessons for ancestral
821 population size inference?. *Heredity* **116**(4):362.
822 Milliken T & Shaw J (2012) The South Africa – VietNam rhino horn trade nexus: A deadly
823 combination of institutional lapses, corrupt wildlife industry professionals and Asian crime
824 syndicates. TRAFFIC. South Africa: Johannesburg.
825 Moodley Y, Russo IR, Dalton DL, Kotzé A, Muya S, Haubensak P, Bálint B, Munimanda GK,
826 Deimel C, Setzer A, Dicks K (2017) Extinctions, genetic erosion and conservation options
827 for the black rhinoceros (*Diceros bicornis*). *Scientific reports* Feb 8;7:41417.
828 Moodley Y, Russo IR, Robovský J, Dalton DL, Kotzé A, Smith S, Stejskal J, Ryder OA, Hermes
829 R, Walzer C, Bruford MW (2018) Contrasting evolutionary history, anthropogenic declines
830 and genetic contact in the northern and southern white rhinoceros (*Ceratotherium simum*).
831 *Proceedings of the Royal Society B* Nov 7;285(1890):20181567.
832 Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C & Durbin R (2016) BCFtools/RoH: A
833 Hidden Markov Model Approach for Detecting Autozygosity from next-Generation
834 Sequencing Data. *Bioinformatics* **32** (11): 1749–51.
835 Pardo-Diaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, Joron M, McMillan WO &
836 Jiggins CD (2012) Adaptive introgression across species boundaries in *Heliconius* butterflies.
837 *PLoS genetics* **8**(6), p.e1002752.
838 Pease JB & Hahn MW (2015) Detection and Polarization of Introgression in a Five-Taxon
839 Phylogeny. *Systematic Biology* **64** (4): 651–62.

840 Pease JB & Rosenzweig BK (2018) Encoding Data Using Biological Principles: The Multisample
841 Variant Format for Phylogenomics and Population Genomics. *IEEE/ACM Transactions on*
842 *Computational Biology and Bioinformatics / IEEE, ACM* 15 (4): 1231–38.
843 Pickrell J, Pritchard J. (2012) Inference of population splits and mixtures from genome-wide
844 allele frequency data. *Nature Precedings*. Mar 2:1-.
845 Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M,
846 Grabherr MG, Wolf JB (2014) The genomic landscape underlying phenotypic integrity in the
847 face of gene flow in crows. *Science*. Jun 20;344(6190):1410-4.
848 Pool JE, Nielsen R (2009) Inference of historical changes in migration rate from the lengths of
849 migrant tracts. *Genetics*. Feb 1;181(2):711-9.
850 Prado-Martinez J, Sudmant P, Kidd J, Li H, Kelley J, Lorente-Galdos B *et al.* (2013).Great ape
851 genetic diversity and population history. *Nature* **499** (7459), 471-475 DOI:
852 10.1038/nature12228.
853 Pujolar JM, Dalén L, Olsen RA, Hansen MM & Madsen J (2018) First *de novo* whole genome
854 sequencing and assembly of the pink-footed goose. *Genomics* **110**(2), pp.75-79.
855 Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley
856 PD, Sugnet CW, Haussler D (2016) Chromosome-scale shotgun assembly using an in vitro
857 method for long-range linkage. *Genome research*. **26**(3), pp.342-350.
858 R Core Team (2019). R: A language and environment for statistical computing. R Foundation for
859 Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
860 Racimo F, Sankararaman S, Nielsen R & Huerta-Sánchez E (2015) Evidence for archaic adaptive
861 introgression in humans. *Nature Reviews Genetics* **16**(6), p.359.
862 Reich, D., Thangaraj, K., Patterson, N., Price, A. L., and Singh, L. (2009) Reconstructing Indian
863 population history. *Nature*, 461(7263):489–94.
864 Rogers J, Raveendran M, Harris RA, Mailund T, Leppälä K, Athanasiadis G *w* (2019). The
865 comparative genomics and complex population history of *Papio* baboons. *Science Advances*,
866 5(1): eaau6947.
867 Rookmaaker K and Antoine PO (2012) New maps representing the historical and recent
868 distribution of the African species of rhinoceros: Diceros bicornis, Ceratotherium simum and
869 Ceratotherium cottoni. *Pachyderm* 52, 91-96.
870 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV & Zdobnov EM (2015) BUSCO:
871 assessing genome assembly and annotation completeness with single-copy orthologs.
872 *Bioinformatics* **31**(19), pp.3210-3212.
873 Slatkin M, Pollack JL. (2008). Subdivision in an ancestral species creates asymmetry in gene
874 trees. *Mol Biol Evol.* 2510:2241–2246.
875 Stanke M & Waack S (2003) Gene Prediction with a Hidden Markov Model and a New Intron
876 Submodel. *Bioinformatics* **19** Suppl 2 (October): ii215–25.
877 Wang RL, Wakeley J, Hey J (1997) Gene flow and natural selection in the origin of Drosophila
878 pseudoobscura and close relatives. *Genetics*;147(3):1091-1106.
879 Westbury MV, Petersen B, Lorenzen ED (2019) Genomic analyses reveal an absence of
880 contemporary introgressive admixture between fin whales and blue whales, despite known
881 hybrids. *PloS one*;14(9).
882 WoldeGabriel G, Haile-Selassie Y, Renne PR, Hart WK, Ambrose SH, Asfaw B, Heiken G, White
883 T (2001) Geology and palaeontology of the late Miocene Middle Awash valley, Afar rift,

884 Ethiopia. *Nature* **412**(6843):175.
885 Won YJ & Hey J (2005) Divergence population genetics of chimpanzees. *Molecular biology and*
886 *evolution* Feb 1;22(2):297-307.
887 Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and*
888 *Evolution* **24** (8): 1586–91.
889 Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in
890 global climate 65 Ma to present. *Science* **292**(5517):686-93.
891 Zheng Y and Janke A (2018) Gene flow analysis method, the D-statistic, is robust in a wide
892 parameter space. *BMC bioinformatics.* Dec 1;19(1):10.
893
894

Figure legends

Figure 1. The distribution of African rhinoceros taxa in time and space. The distribution of co-occurring Pliocene fossils of *Diceros praecox* (precursor to the black rhinoceros) and Ceratotherium mauritanicum (precursor to the white rhinoceros) are given in red-grey squares. Pleistocene fossil distributions of modern black and white rhinoceros are given in black, white or black-white squares and Holocene distributions (after Rookmaaker and Antoine, 2012) of eastern (green) and southern (yellow) black rhinoceros and northern (blue) and southern (red) white rhinoceros are depicted.

Figure 2. Evolutionary and demographic histories of the black and white rhinoceros. **A.** Species tree and intraspecific divergence times assuming an ancestral split at the Miocene-Pliocene boundary. **B.** Demographic reconstructions for each African rhinoceros species showing windows for the end of panmixia within both species. **C.** Paleoclimatic reconstruction for the last 1.4 million years, modified from Zachos *et al.* (2001).

Figure 3. Inferring the cessation of gene flow between and within black and white rhinoceros using hPSMC and simulations. **A.** hPSMC plot between black and white rhinoceros and simulations of different divergence times spanning 3-7 Ma in 100,000-year intervals. **B.** hPSMC between northern and southern white rhinoceros and simulations of divergence times spanning 50,000-450,000 years in 10,000-year intervals. **C.** hPSMC between eastern and southern black rhinoceros and simulations of divergence times spanning 0-400,000 years ago in 10,000-year intervals. Greyed out regions represent 1.5x and 10x the pre- divergence effective population size. Bold red lines represent the hPSMC results based on the real data. Thin grey lines represent the simulated data while thin black lines represent the simulations closest to the real data without overlapping it, which was used to infer the time interval when gene flow ceased.

Figure 4. Model of gene flow and the evolution of specialization in white and black rhinoceros. Our analyses indicate that ongoing gene flow between speciating *Ceratotherium* and *Diceros* lineages continued for up to two million years after initial divergence. Black asterisks indicate the first appearances of both lineages in the fossil record. The grey dashed line marks the time at which fossils of both lineages were present at Kanapoi in East Africa.