

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/132358/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Moodley, Yoshan, Westbury, Michael V., Russo, Isa-Rita M. ORCID: <https://orcid.org/0000-0001-9504-3633>, Gopalakrishnan, Shyam, Rakotoarivelo, Andrinajoro, Olsen, Remi-Andre, Prost, Stefan, Tunstall, Tate, Ryder, Oliver A., Dalen, Love and Bruford, Michael W. ORCID: <https://orcid.org/0000-0001-6357-6080> 2020. Interspecific gene flow and the evolution of specialisation in black and white rhinoceros. *Molecular Biology and Evolution* 37 (11) , pp. 3105-3117. 10.1093/molbev/msaa148  
filefilefilefilefilefile

Publishers page: <http://dx.doi.org/10.1093/molbev/msaa148>  
<<http://dx.doi.org/10.1093/molbev/msaa148>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



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

2

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

4

5 **Short title:** Pliocene gene flow in African Rhinoceros

6

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

10

11 #equally contributing first authors

12

13 **Affiliations:** <sup>1</sup> Department of Zoology, University of Venda, Private Bag X5050, Thohoyandou  
14 0950, Republic of South Africa, <sup>2</sup> Section for Evolutionary Genomics, GLOBE institute, University  
15 of Copenhagen, Denmark, <sup>3</sup> School of Biosciences, Sir Martin Evans Building, Cardiff University,  
16 Museum Avenue, Cardiff, CF10 3AX, United Kingdom, <sup>4</sup> Natiora Ahy Madagasikara, Lot IIU57K  
17 Bis, Ampahibe, Antananarivo 101, Madagascar (ARR), <sup>5</sup> Science for Life Laboratory, Department  
18 of Biochemistry and Biophysics, Stockholm University, Box 1031, SE-17121 Solna, Sweden, <sup>6</sup>  
19 LOEWE-Centre for Translational Biodiversity Genomics, Senckenberg Museum, Frankfurt,  
20 Germany, <sup>7</sup> South African National Biodiversity Institute, National Zoological Garden, Pretoria,  
21 Republic of South Africa, <sup>8</sup> San Diego Zoo Institute for Conservation Research, San Diego Zoo  
22 Global, Escondido, California, United States of America, <sup>9</sup> Centre for Palaeogenetics, Svante  
23 Arrhenius Väg 20C, SE-10691 Stockholm, Sweden, <sup>10</sup> Department of Bioinformatics and  
24 Genetics, Swedish Museum of Natural History, SE-10405 Stockholm, Sweden, <sup>11</sup> Sustainable  
25 Places Research Institute, Cardiff University, Cardiff CF10 3BA, United Kingdom

26

27 \* Corresponding author. Email: [yoshan.moodley@univen.ac.za](mailto:yoshan.moodley@univen.ac.za)

28

29 **Abstract**

30

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

47

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

## 50 Introduction

51

52 Although the age of Pleistocene mammalian megaherbivores is largely over, Africa is the  
53 only continent to still harbour significant wild populations of its late-tertiary megafauna. Africa's  
54 black (*Diceros bicornis*) and white (*Ceratotherium simum*) rhinoceros are relics of this bygone  
55 "golden age" of large mammals, yet, because of long-term historical demand for their products,  
56 man has succeeded in driving this iconic group to the brink of extinction across the world (Hillman  
57 1980, Chilvers 1990). Although rhinoceros are among the world's most endangered mammals,  
58 the two African species have fared slightly better than their three Asian counterparts, owing mainly  
59 to intensive conservation interventions during the second half of the 20<sup>th</sup> Century, resulting in a  
60 global population of approximately 20,000 white and 5,000 black rhinoceros (Emslie *et al.* 2016).  
61 However, the unfortunate consequence of these population gains is that the most recent poaching  
62 epidemic, driven by increasing demand for rhinoceros horn in East and South-East Asia (Milliken  
63 and Shaw 2012, Kennaugh 2015), is targeting the more common African species.

64 A rich fossil record shows that rhinoceros species have endured a series of severe Plio-  
65 Pleistocene climatic and tectonic upheavals, to which the majority of their megafaunal  
66 contemporaries succumbed (Barnosky *et al.* 2004). In Africa, the paleoclimate during this time  
67 fluctuated between warmer, wetter, forest-rich interglacial periods that favoured browsers, and  
68 cooler, drier grassland-rich glacial periods that benefited grazers. Although black and white  
69 rhinoceros are closely related, they have evolved divergent feeding strategies. The black  
70 rhinoceros is the smaller of the two species and primarily a browser, holding its head horizontally  
71 to the ground in order to feed on leaves and twigs using a hooked upper lip. In contrast, the white  
72 rhinoceros is an obligate grazer with hypsodont teeth, a heavy, elongated skull that is held  
73 vertically and lower to the ground, with squared-off lips to enable efficient grazing. Both feeding  
74 strategies evolved within the last 6-7 million years (Ma), from about the end of the warm Miocene  
75 epoch, as global CO<sub>2</sub> levels decreased, leading to the more arid, seasonal and shifting  
76 paleoclimates of the Plio-Pleistocene (WoldeGabriel *et al.* 2001, deMenocal 2004). Although  
77 grazing rhinoceros such as *Teleoceras* had already evolved during the Miocene, the stem lineage  
78 leading to modern African rhinoceros was a mixed-feeder, represented by the morphologically  
79 intermediate *C. neumayri*, which inhabited late Miocene southern Europe from the Balkans to Iran  
80 approximately 7-9 Ma (Zeuner 1934, Geraads 2005, Geraads and Spassov 2009). This stem  
81 lineage diverged in Africa into *D. praecox* and *C. mauritanicum*, the direct ancestors of black and  
82 white rhinoceros respectively (Geraads 2005, Geraads 2017). The earliest fossil appearance of  
83 *D. praecox* is at Kuseralee in the Middle Awash Valley of Ethiopia (Giaourtsakis *et al.* 2009,  
84 Geraads 2017) and the *Ceratotherium* lineage at Langebaanweg in South Africa (Hooijer *et al.*  
85 1972, Geraads 2005), both sites dating to about 5.2 Ma. The initial divergence between *D.*  
86 *praecox* and *C. mauritanicum* must therefore have occurred no later than around the Mio-Pliocene  
87 boundary about 5.3 Ma.

88 While changing paleoclimates provide a means for the evolution and fixation of different  
89 adaptations, they may also eventually bring speciating populations into secondary contact, where  
90 gene flow might bring their diverging evolutionary trajectories back into line (Mayr 1942) and/or  
91 promote the introgression of adaptive features between populations (Pardo-Diaz *et al.* 2012,  
92 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  $\delta^{13}\text{C}$  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-

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

139

140

## 141 **Results**

142

### 143 *De novo assembly and annotation of the black rhinoceros reference genome*

144 We present a high quality reference genome for the black rhinoceros (SAMN14912225)  
145 from an ear tissue sample of an individual sampled in KwaZulu-Natal, South Africa, at the  
146 southern end of the species range. We reconstructed the assembly using a combination of paired-  
147 end, mate paired, and chromatin-based sequencing libraries. First, we generated a baseline  
148 assembly using a combination of short and long insert libraries using the *de novo* assembler  
149 Allpaths-LG (Gnerre *et al.* 2011). We subsequently carried out super-scaffolding using chromatin-  
150 based Chicago libraries (Putnam *et al.* 2016) and the Hi-Rise pipeline. This resulted in a 34.6-fold  
151 coverage genome with a total assembly length of 2.33GB, with a scaffold N50 of 28.5MB and  
152 4,264 scaffolds. A BUSCO assessment (Simão *et al.* 2015) of the gene content of the assembly  
153 revealed only 13 (0.3%) duplicated, 88 (2.1%) fragmented and 116 (2.8%) missing mammalian  
154 single copy orthologs. We then annotated the assembly using *ab initio* gene prediction and  
155 homology-based gene identification, which resulted in 19,914 transcripts.

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

162

### 163 *Genome-wide heterozygosity*

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

171

### 172 *Divergence and mutation rates*

173 Based on the autosomal sequences, we calculated an average pairwise divergence of  
174 0.0093 between the two species. Within each species, lower divergences were estimated  
175 between eastern and southern black rhinoceros (0.0011) and northern and southern white  
176 rhinoceros (0.0010). Using a conservative estimate (latest possible occurrence) for the split, or  
177 end of panmixia, between black and white rhinoceros lineages at the Mio-Pliocene boundary  
178 ~5.3Ma (Geraads 2005, 2017), these pairwise distances were translated into approximate within-

179 species divergence times of 641 and 578 thousand years ago (Ka) for the black and white  
180 rhinoceros, respectively (Figure 2A). Using the pairwise distance between the white and black  
181 rhinoceros, we calculated an autosomal mutation rate for African rhinoceros of  $8.8 \times 10^{-10}$   
182 substitutions per year, which is only slightly lower than the commonly implemented human  
183 mutation rate of  $1 \times 10^{-9}$  (Li and Durbin 2011), and refutes the commonly held view that evolutionary  
184 rates in rhinoceros genomes are substantially lower than the mammalian average (Gissi *et al.*  
185 2000). Furthermore, we calculated the per generation mutation rate for each rhinoceros species  
186 independently, assuming a generation time of 24 years for the black rhinoceros (Moodley *et al.*  
187 2017), giving a mutation rate of  $2.1 \times 10^{-8}$ , and a generation time of 27 years for the white  
188 rhinoceros (Moodley *et al.* 2018), giving a mutation rate of  $2.4 \times 10^{-8}$ .

189

### 190 *Demographic reconstruction*

191 We reconstructed the demographic histories of both African rhinoceros species over the  
192 second half of the Pleistocene (<1.4 Ma) using a pairwise sequentially Markovian coalescent  
193 (PSMC) model. Both species show a gradual reduction in effective population size ( $N_e$ ) to less  
194 than half their original size until about 520-540 Ka ago in black rhinoceros and 440-460 Ka in  
195 white rhinoceros (yellow stars, Figure 2B). The demographic trajectories of both species also  
196 diverged at this low point, indicating the approximate times at which the ancestral populations of  
197 black rhinoceros and white rhinoceros divided, signalling the end of panmixia within each species.  
198 After this point, all four genomes then appear to follow independent Middle Pleistocene population  
199 expansions. Interestingly, the southern populations of both species reach their highest size at  
200 about 230 Ka, earlier than their northern counterparts the eastern black rhinoceros at 200 Ka and  
201 the northern white rhinoceros at 180 Ka. All four populations then contract to Holocene levels of  
202  $N_e$  below 5,000, although both southern African populations show a secondary but minor  
203 population expansion at about 50 Ka for the southern white rhinoceros and within the last 20 Ka  
204 for the southern black rhinoceros. It is important to note that PSMC-inferred demographic  
205 trajectories are often difficult to interpret literally (Beichman *et al.* 2017). We also investigated  
206 whether different sequencing depths, especially in the case of the northern white and eastern  
207 black rhinoceros (16-fold coverage) may have influenced the observed demographic trajectories.  
208 Based on a comparison between PSMC trajectories reconstructed using our newly sequenced  
209 ~35-fold coverage southern black rhinoceros genome, and the same genome downsampled to  
210 16-fold coverage (Supplementary Figure S1), we deduced that the differences caused by  
211 differential coverage was negligible.

212

### 213 *Post-divergence gene flow*

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

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

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

#### 254 255 *Reproductive isolation and the cessation of post-divergence gene flow*

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



265 changes in ancestral  $N_e$  as gradual transitions, one cannot apply a purely qualitative approach to  
266 estimating divergence times. Therefore, we ran simulations specifying various divergence times  
267 between the individuals of interest. Simulations were run using the `hPSMC_quantify_split_time.py`  
268 python script from the hPSMC tool suite specifying pre-divergence  $N_e$ , time windows for  
269 divergence, and default parameters. Results from the real data as well as simulations based on  
270  $N_e$ 's calculated before its exponential increase to infinity, indicated that reproductive isolation  
271 between black and white rhinoceros lineages occurred between 3.3 and 4.1 Ma (Figure 3A), much  
272 more recently than the initial divergence time of the two species at ~5.3 Ma or earlier (Geraads  
273 2005, 2017). This result was the same regardless of which of the two genomes of each species  
274 were compared and which species was used as the mapping reference.

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

282

### 283 *Evaluating D-statistics in the presence of ancestral gene flow*

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

292

293

## 294 **Discussion**

295

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

302

### 303 *Genomic diversity, Pleistocene declines and expansion*

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

308 the earth at the beginning of the Pliocene. However, a subsequent population expansion of all  
309 genomes is not consistent with the paleoclimatic record, since this was just after the time when  
310 glacial cycles became more severe (<800 Ka, Figure 2C). Rather, the increase in effective  
311 population size occurred at the point at which panmixia in both species ended (Figure 2). It is  
312 possible that the 520 Ka expansion in black rhinoceros was associated with an interglacial cycle,  
313 whereas the 440 Ka white rhinoceros expansion could be associated with an interglacial, but both  
314 species effective sizes were inferred to have expanded, regardless of subsequent glacial cycles,  
315 until about 240 Ka. Alternatively, the early evolution of additional genetic substructure within  
316 diverging regional populations of each species, followed by their isolation during unfavourable  
317 climatic periods, could also have inflated effective population sizes, even if census sizes remained  
318 stable (Mazet *et al.* 2015, Mazet *et al.* 2016). Although most mtDNA lineages to have evolved in  
319 the white rhinoceros were already extinct by the Holocene (Moodley *et al.* 2018), in the black  
320 rhinoceros, mtDNA is highly structured with both eastern and southern lineages (Moodley *et al.*  
321 2017), lending weight to this interpretation.

322 At approximately 240 Ka, southern black and white rhinoceros underwent a population  
323 decline, followed by northern white and eastern black rhinoceros at 180 Ka. Until now, only the  
324 genomes of bonobos, Nigeria-Cameroon chimpanzees, the spectacled bear and east African  
325 baboons show a similar Middle Pleistocene decline around 150-200 Ka (Prado-Martinez *et al.*  
326 2013, Kumar *et al.* 2017, Rogers *et al.* 2019). It is interesting that these declines in both African  
327 rhinoceros coincide with the emergence of modern humans. A similarly sharp demographic  
328 decline was also inferred for the Sumatran rhinoceros, but more recently at about 100 Ka (Mays  
329 *et al.* 2018), coinciding with the appearance of humans in Asia. It is also intriguing that southern  
330 populations of both rhinoceros species decline before populations in the north and east, as it could  
331 imply differential levels of population pressure across Africa.

332  
333 *Divergence, the end of panmixia and the cessation of gene flow within each species*

334 Within-species autosomal divergence times for black (641 Ka) and white (578 Ka)  
335 rhinoceros were consistently about 100 Ka older than the times at which panmixia is inferred to  
336 have ceased for each species using PSMC. Although divergence times do not account for  
337 demographic events or gene flow, values are remarkably similar considering their different  
338 methods of inference. The reported divergence times are also within the confidence limits of  
339 mtDNA data for white rhinoceros (Harley *et al.* 2016, Moodley *et al.* 2018), but not for black  
340 rhinoceros, where mtDNA divergence between southern and eastern black rhinoceros was  
341 inferred to be significantly more ancient (920-3,575 Ka, Moodley *et al.* 2017). It is possible that  
342 values for the end of panmixia may have been downwardly biased as populations of both species  
343 underwent very similar demographic expansion trajectories after their PSMC curves became  
344 dissociated (Figure 2B). However, divergence and the end of panmixia occurred long before the  
345 final cessation of gene flow within both species (Figure 3B/C), indicating ongoing secondary  
346 contact, potentially during phases of demographic expansion during the last 400,000 years. While  
347 the two white rhinoceros populations appear to have come into secondary contact less recently  
348 (100-220 Ka), this estimate is still consistent with gene flow during the last glacial period, as  
349 recently inferred from microsatellite data (Moodley *et al.* 2018). The eastern and southern black  
350 rhinoceros on the other hand, appear to have come into more recent genetic contact across the

351 Zambezi valley (Figure 3C). The Zambezi's paleo-upper and -lower reaches were joined by river  
352 capture between 125–150 Ka (Moore and Larkin 2001), and while our results appear to contrast  
353 with a strong mtDNA and microsatellite discontinuity on either side of this river (Moodley *et al.*  
354 2017), at least one East African mtDNA haplotype was sampled on the southern bank of the  
355 Zambezi, and one southern African haplotype was sampled north of the Zambezi, hinting that  
356 although the river may have acted to maintain the genetic integrity of populations to its north and  
357 south, it was also periodically fordable for black rhinoceros. In summary, these results suggest  
358 that the period required between divergence and the end of panmixia to the cessation of gene  
359 flow is dependent on how frequently climatic changes were able to bring diverging populations  
360 into secondary contact, and in Africa, the expansion and contraction of habitats with glacial cycles  
361 appears to have maintained gene flow long after population divergence. Although post-divergence  
362 gene flow was observed previously in other taxa, including rhinoceros (Wang *et al.* 1997, Won  
363 and Hey 2005, Lee and Edwards 2008, Moodley *et al.* 2018), our results provide yet another  
364 cautionary note in evolutionary and conservation inference, that estimated times of divergence do  
365 not necessarily correlate with the cessation of genetic contact.

366

#### 367 *Gene flow between African rhinoceros species*

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

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

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

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

#### 419 420 *The evolution of specialisation*

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

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

434 *C. mauritanicum* prior to 4 Ma may have continually undermined the diverging evolutionary trends  
435 of both lineages towards browsing and grazing, respectively. The rhinoceros-rich fossil record of  
436 the mid-late Pliocene provides some evidence to support this view, because despite over a million  
437 years since their initial divergence, both species maintained their ancestral mixed feeding state  
438 throughout most of the Pliocene (Geraads 2017). So phenotypically similar were *D. praecox* and  
439 *C. mauritanicum* during this period, that palaeontologists often misidentified one species for the  
440 other (Geraads 2005, Geraads 2010). On the other hand, there is also extensive theoretical  
441 (Barton 1979, Barton 1987) and empirical (McCracken *et al.* 2009, Hohenlohe *et al.* 2012,  
442 Poelstra *et al.* 2014) evidence that adaptation can occur even in the face of gene flow, when  
443 introgressing alleles confer a selective advantage, with advantageous loci often in tight linkage  
444 disequilibrium, or if hybrid fitness is low (Barton and Hewitt 1985). The fact that both *D. praecox*  
445 and *C. mauritanicum* had developed some level of specialisation prior to the onset of reproductive  
446 isolation, suggests that adaptation may have been occurring despite Pliocene gene flow.  
447 However, eventual reproductive isolation between the two species likely resulted from an  
448 accumulation of larger numbers of loci under selection (Barton and Hewitt, 1989). The evolution  
449 of fully specialised browsing and grazing African rhinoceros species could only have occurred  
450 during the critical phase after reproductive isolation between them was established (3.3-4.1 Ma),  
451 but before the internal splits within each species (500-600 Ka, Figure 4). The fossil emergence of  
452 phenotypically modern black (2.5 Ma) and white (1.8 Ma) rhinoceros falls exactly within this  
453 interval. The timing of these fossil emergences suggests that the more severe and shifting  
454 paleoclimates of the Pleistocene provided the heterogeneity of environments that ultimately drove  
455 the evolution of obligate feeding strategies in African rhinoceros species.  
456

## 457 **Methods**

458

### 459 *Establishing the black rhinoceros reference assembly*

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

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

487

### 488 *Genome annotation*

489 Next, we carried out repeat and gene annotation. To do so, we first masked repeats in the  
490 genome using a combination of *ab initio* repeat finding and homology-based repeat annotation  
491 using RepeatModeler (<http://www.repeatmasker.org>) and RepeatMasker ([http://www.repeat](http://www.repeatmasker.org)  
492 [masker.org](http://www.repeatmasker.org)), respectively. For homology-based repeat annotation we used the mammal repeat  
493 consensus sequences from Repbase (Bao, *et al.* 2015). For the gene annotation, we did not mask  
494 simple repeats beforehand to improve mapping during the homology-based annotation  
495 implemented in Maker2 (Holt and Yandell 2011). The gene annotation was performed using a  
496 combination of *ab-initio* gene prediction (using SNAP (Korf 2004) and Augustus (Stanke and  
497 Waack 2003)) and homology-based gene annotation using Maker2. We used protein annotations  
498 of the horse (EquCab2.0; GCF\_000002305.2), the white rhinoceros (CerSimSim1.0;

499 GCF\_000283155.1) and human (GRCh38; GCA\_000001405.37) for the homology-based gene  
500 annotation step. This resulted in the annotation of 19,914 genes.

501  
502 *Raw data processing and mapping*

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

516 Raw reads were all treated comparably before being mapped to a specific reference  
517 genome. We used Cutadapt v1.8.1 (Martin 2011) to trim Illumina adapter sequences from the  
518 ends of reads and remove reads shorter than 30bp. We then merged overlapping read pairs using  
519 FLASH v1.2.1 (Magoč and Salzberg 2011). We mapped the resultant reads of the five individuals  
520 used in the study to their respective reference sequences, unless otherwise specified, using  
521 BWA v0.7.15 (Li and Durbin 2009) and processed the mapped reads further using SAMtools  
522 v1.3.1 (Li *et al.* 2009). We mapped both the East African black rhinoceros and South African black  
523 rhinoceros to the newly assembled black rhinoceros genome, both the northern white rhinoceros  
524 and the southern white rhinoceros to the published southern white rhinoceros genome, and the  
525 Sumatran rhinoceros to the Sumatran rhinoceros genome (GCA\_002844835.1).

526  
527 *Genetic variation*

528 We estimated autosomal heterozygosity from each of the four African rhinoceros  
529 individuals. To determine which scaffolds were most likely autosomal in origin, we found putative  
530 sex chromosome scaffolds for each of the rhinoceros reference genomes and removed them from  
531 future analyses. We found putative sex chromosome scaffolds through synteny by aligning the  
532 rhinoceros reference genomes to the Horse X (Genbank accession: CM000408.2) and Human Y  
533 (Genbank accession: NC\_000024.10) chromosomes. Alignments were performed using satsuma  
534 synteny (Grabherr *et al.* 2010) and utilising default parameters. To adjust for biases in  
535 heterozygosity levels that could arise due to different global coverages between the genomes of  
536 the individuals being investigated, we subsampled all of the resultant alignments down to that of  
537 the lowest coverage individual, 16-fold, using SAMtools. We then estimated the autosomal  
538 heterozygosity from all scaffolds above 100 KB in length, using sample allele frequencies in  
539 ANGSD v0.913 (Korneliussen *et al.* 2014), taking genotype likelihoods into account and specifying  
540 the following filters -minq 25 -minmapq 25 -uniqueOnly 1 -baq 1 -remove\_bads 1. We calculated  
541 the standard deviation for each of the heterozygosity estimates by performing the realSFS

542 function in the ANGSD package in independent 20MB windows of covered bases (-nSites  
543 20,000,000).

544

#### 545 *Genome divergence and mutation rate*

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

564

#### 565 *Demographic analyses*

566 We ran demographic analyses on the diploid genomes of all four African rhinoceros  
567 individuals using Pairwise Sequentially Markovian Coalescent model (PSMC) (Li and Durbin  
568 2011). Using this method, it is possible to infer changes in effective population size through time  
569 for diploid (high coverage) genomes from the distribution of its heterozygous sites across the  
570 genome. We called diploid genome sequences using SAMtools and bcftools (Narasimhan et al.  
571 2016) specifying a minimum quality score of 20 and minimum coverage of 10. We removed  
572 scaffolds found to align to sex chromosomes in the previous step and scaffolds shorter than  
573 100KB. We ran PSMC specifying atomic intervals previously shown to be suitable for human  
574 datasets (4+25\*2+4+6) and performed 100 bootstrap replicates to investigate support for the  
575 resultant demography. We overlaid the resultant PSMC plots as the point in time in which the  
576 demographic trajectories of two individuals diverges can be interpreted as a rough measure of  
577 the end of panmixia in that species. Moreover, as one of our individuals (northern white rhino)  
578 was only ~16-fold coverage, we downsampled our (~35-fold coverage) southern black rhinoceros  
579 genome to 16-fold coverage to investigate the effect this may have on our inferences. We ran a  
580 PSMC analysis on the downsampled genome and compared it to the results recovered for the  
581 same genome using the much higher coverage data.

582



583 *Inter and intra-specific post-divergence gene flow*

584 For the gene flow analyses, we mapped the raw reads from the four African species to the  
585 Sumatran rhinoceros following the same methods mentioned above to avoid any ascertainment  
586 bias that may occur when mapping to an ingroup African rhinoceros species (Westbury *et al.*  
587 2019). We performed multiple different analyses to test for post divergence gene flow between  
588 African rhinoceros. First, we implemented the four-taxon ABBA/BABA or D-statistics approach  
589 (Durand *et al.* 2011) with ANGSD. We called bases using a random base call (-doAbbababa 1),  
590 only considered scaffolds over 100KB in length, specified the Sumatran rhinoceros as outgroup,  
591 and applied the following filters; -minMapQ 25, -minQ 25, -uniqueOnly 1, -remove\_bads 1. We  
592 also adjusted quality scores around indels (-baq 1) (Li 2011). ANGSD performs all possible  
593 combinations but we only investigated the output with conspecifics in the H1 and H2 positions  
594 and an individual from the other species in the H3. Any other combination would go against the  
595 species tree and therefore produce invalid signs of admixture driven by more recent common  
596 ancestry as opposed to true admixture. To investigate the significance of our result, we performed  
597 a weighted block jackknife test using 5MB non-overlapping blocks. D-values more than three  
598 standard errors different from zero ( $-3 < Z > 3$ ) were considered as statistically significant.

599 Following the D-statistics, we implemented Dfoil (Pease and Hahn 2015), a more detailed,  
600 expanded version of D-statistics using five-taxa to test for gene flow, with the Sumatran rhinoceros  
601 as an outgroup. Dfoil implements four independent D-statistics in a sliding window fashion which  
602 are then combined before inferences are made. This has the advantage over the four-taxon test  
603 in that it can infer the direction of gene flow and uncover whether gene flow occurred between  
604 ancestral lineages. For this analysis, we also mapped all rhinoceros to the Sumatran rhinoceros.  
605 We then constructed fasta files for each individual using ANGSD and specifying maximum  
606 effective base depth (-doFasta 3) and the following parameters: -minMapQ 25, -minQ 25, -  
607 uniqueOnly 1, -remove\_bads 1. Additionally, we removed all scaffolds shorter than 1MB and  
608 trimmed the ends of the remaining scaffolds down to the nearest 100KB, leaving us with 937.2MB.  
609 The resultant fasta files were converted into an mvf file (ConvertFasta2MVF) which was then  
610 converted into three independent Dfoil input files (CalcPatternCount) of window sizes 100KB,  
611 50KB, and 20KB with mvftools (Pease and Rosenzweig 2018). Regions showing signs of  
612 admixture between the ancestral white rhinoceros and either the southern or eastern black  
613 rhinoceros were extracted and compared between window sizes. We cross referenced the  
614 introgressed genomic segments consistently showing signs of admixture despite window size  
615 against the white rhinoceros annotation to uncover putative protein coding genes in these regions.  
616 We then tested for GO enrichment terms with GOrilla (Eden *et al.* 2009). We further investigated  
617 the contiguity of introgressed regions by extracting all regions showing any signs of introgression,  
618 regardless of direction, and investigating how long the stretches of these introgression windows  
619 were for all window sizes independently.

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

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

634

#### 635 *Timing of reproductive isolation and the cessation of gene flow*

636 To add a temporal element to the onset of reproductive isolation and the cessation of gene  
637 flow between African rhinoceros, we used the F1 hybrid pairwise sequentially Markovian  
638 coalescent model, hPSMC (Cahill *et al.* 2016). To address whether ascertainment bias may have  
639 played a role in our results, we performed this analysis twice independently for the between  
640 species comparisons, white vs black rhinoceros, once using the black rhinoceros as reference  
641 genome, and once using the white rhinoceros as reference genome. Within species comparisons  
642 were only computed once using the conspecific genome as reference. We constructed haploid  
643 consensus sequences for the four individuals using ANGSD by considering the base with the  
644 highest effective depth, the following quality filters; -minQ 25, -minmapq 25, -uniqueonly 1, -  
645 remove\_bads 1, and only considering autosomes and scaffolds over 100KB. We merged these  
646 resultant haploid consensus sequences together into a pseudo diploid sequence using the  
647 hPSMC tool suite. These were then run through PSMC and plotted using an intermediate mutation  
648 rate per generation and generation time. When comparing the black and white rhinoceros we  
649 used a generation time of 25.5 years and a mutation rate of  $2.2 \times 10^{-8}$  mutations per generation.  
650 When comparing within species, we used intraspecific mutation rates and generation times. From  
651 this, we manually estimated the pre-divergence  $N_e$  by outputting the text file (-R) using the plot  
652 script from the PSMC tool suite and looking into the output text file. Using the pre-divergence  $N_e$   
653 estimated from this output, we then ran simulations to infer the confidence intervals using Ms  
654 (Hudson 2002) with the hPSMC\_quantify\_split\_time.py python script from the hPSMC tool suite,  
655 while specifying the time windows we wanted to simulate, and the remaining parameters as  
656 default. When comparing black and white rhinoceros, we estimated a pre-divergence  $N_e$  of  
657 60,000 and ran simulations using divergence times between 3,000,000-7,000,000 years in  
658 100,000 year intervals. When comparing northern and southern white rhinoceros, we estimated  
659 a pre-divergence  $N_e$  of 7,000 and ran simulations using divergence times between 50,000-  
660 450,000 years in 10,000 year intervals. When comparing eastern and southern black rhinoceros,  
661 we estimated a pre-divergence  $N_e$  of 13,000 and ran simulations using divergence times between  
662 0-400,000 years in 10,000 year intervals. Results were plotted and the simulations with an  
663 exponential increase in  $N_e$  closest to the real data, within 1.5x and 10x of the pre-divergence  $N_e$ ,  
664 were taken as the time interval in which gene flow stopped. We considered the portion between  
665 1.5x and 10x of the pre-divergence  $N_e$  as suggested by the original manuscript. This was  
666 suggested in order to capture the portion of the hPSMC plot most influenced by the divergence  
667 event. The lower bound is set to control for pre-divergence increases in population size and the

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

670

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

672 In order to evaluate the influence of ancestral gene flow on D-statistics results, we ran a simple  
673 model simulation in MSMS (Ewing and Hermisson, 2010) specifying gene flow between the  
674 ancestral lineages as shown in Figure 4. This was done using the following command: msms 82  
675 500 -l 5 2 20 20 20 20 0 -t 1760 -r 352 -ej 0.2375 5 4 -ej 0.2375 125 3 2 -em 1.875 4 2 {migration  
676 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  
677 1MB, an effective population size of 20,000 for all five populations, with constant population sizes,  
678 a generational mutation rate of  $2.2 \times 10^{-8}$  and a recombination rate one fifth of the mutation rate  
679 ( $4.4 \times 10^{-9}$ ), three independent runs of 20,000 windows, each with different migration rates ( $m =$   
680 0.5, 1, and 2), a divergence time of 200,000 generations between the black and white rhinos, the  
681 end of gene flow between the black and white rhinoceros as 150,000 generations, the within  
682 species divergence as 19,000 generations and assuming a generation time of 25.5 years. The  
683 output of the simulations were then run through a custom python script which calculated the D-  
684 score for each 1MB window independently. Finally, we performed a block jackknifing approach  
685 with the resample library in R v3.6.0 to test for significance of the results.

686

### 687 **Acknowledgements**

688

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

701

### 702 **References**

703

- 704 Bao W, Kojima KK., & Kohany O (2015). Repbase Update, a database of repetitive elements in  
705 eukaryotic genomes. *Mobile DNA*, **6**(1), 11.
- 706 Barnosky AD, Koch PL, Feranec RS, Wing SL, Shabel AB (2004) Assessing the causes of late  
707 Pleistocene extinctions on the continents. *Science* **306**(5693):70-5.
- 708 Barton NH (1979) Gene flow past a cline. *Heredity*. Dec;43(3):333-9.
- 709 Barton NH (1987) The probability of establishment of an advantageous mutant in a subdivided  
710 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  $\beta$ A 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-Díaz 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



895 **Figure legends**

896

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

904

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

910

911 **Figure 3.** Inferring the cessation of gene flow between and within black and white rhinoceros  
912 using hPSMC and simulations. **A.** hPSMC plot between black and white rhinoceros and  
913 simulations of different divergence times spanning 3-7 Ma in 100,000-year intervals. **B.** hPSMC  
914 between northern and southern white rhinoceros and simulations of divergence times spanning  
915 50,000-450,000 years in 10,000-year intervals. **C.** hPSMC between eastern and southern black  
916 rhinoceros and simulations of divergence times spanning 0-400,000 years ago in 10,000-year  
917 intervals. Greyed out regions represent 1.5x and 10x the pre- divergence effective population  
918 size. Bold red lines represent the hPSMC results based on the real data. Thin grey lines represent  
919 the simulated data while thin black lines represent the simulations closest to the real data without  
920 overlapping it, which was used to infer the time interval when gene flow ceased.

921

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