Electronic Supplement Material 2 (Methods and Figures) for

Contrasting evolutionary history, anthropogenic declines and genetic contact in the northern and southern white rhinoceros (*Ceratotherium simum*)

Yoshan Moodley*, Isa-Rita Russo*, Jan Robovský, Desiré L. Dalton, Antoinette Kotzé, Steve Smith, Jan Stejskal, Oliver A. Ryder, Robert Hermes, Chris Walzer & Michael W. Bruford

* Equally contributing first authors

#Corresponding authors: Yoshan Moodley (yoshan.moodley@univen.ac.za), Isa-Rita Russo (RussoIM@cardiff.ac.uk) and Michael W. Bruford (BrufordMW@Cardiff.ac.uk).

This file contains:

1. Samples

Samples were collected from wild and captive animals for both SWR and NWR. A total of 217 samples were obtained from eight SWR populations, including 174 specimens taken from six wild subpopulations in South Africa and three museum specimens, the oldest of which dates to 1869 (Robovský et al. 2010), prior to the SWR bottleneck of the late 1800s (electronic supplementary material (ESM1, Table S1). In addition, 42 captive samples were obtained from zoological gardens and animal parks in Europe and Africa (ESM1, Table S2). Since the NWR is extinct in the wild, our wild sample comprises 11 individuals that were obtained from museum specimens representing a large part of the former distribution from north-eastern Belgian Congo (now DRC), Sudan and Uganda. These samples were from animals hunted between 1905 and 1912 (ESM1, Table S1). We also included 11 NWR samples from captive animals (ESM1, Table S2). This population was founded by animals caught on the west bank of the Nile River in the Equatoria province, Sudan (now South Sudan) and the West Nile province, Uganda between 1950 and 1970, when wild NWR were already declining. These captive NWR were managed at the Dvůr Králové Zoo (DK) in the Czech Republic and the San Diego Wildlife and Animal Park (SD) in the USA. The population has since declined such that only three individuals remain. The captive NWR sample of 11 (7 DK, 4 SD), comprises six of the original founders (Sudan, Angalifu, Saut, Lucy, Dinka and Nola) and Nasi, Suni, Nabire and Najin which are daughters of Nasima (another founder from Uganda). Nasi is the only known SWR-NWR hybrid.

2. Molecular methods

Museum samples (~250 µg) were rehydrated for 48 hours in distilled water. DNA was then extracted in an ancient DNA laboratory at the Konrad Lorenz Institute in Vienna according to the protocol outlined in Rohland *et al.* (2010). All modern samples were extracted using a DNeasy® Tissue Kit (QIAGEN® Hilden, Germany) following the manufacturer's instructions. The 5' end of the control region using primers mt15996L (5'-TCCACCATCAGCACCCAAAGC-3')[5] and mt16502H (5'-TTTGATGGCCCTGAAGTAAGAACCA-3') were used to amplify a 477 bp fragment of the control region. Polymerase chain reactions (PCR) was carried out using 50-100 ng/µl of DNA in a 25 µl reaction containing 1x PCR buffer, 3 mM $MgCl_2$, 0.2 mM of each dNTP, 0.1 mg/µl purified BSA (New England BioLabs), 0.2 µM of each primer and 1.25 U Amplitaq Gold DNA polymerase (Promega). The following PCR conditions were used: 95°C for 5 min followed by 45 cycles of 94°C for 30 sec, 60°C for 1 min, 72°C for 1 min, and 72°C for 10 min. All museum samples were amplified/sequenced at least twice in both directions, for consistency. Sequencing reactions were run through an ABI 3130*xl* Genetic Analyser and sequences were trimmed and

aligned in CLC DNA Workbench (CLC Biotech, Qiagen Aarhus). We also followed the protocol and used the primers designed in Moodley et al. (2017) for the least tractable museum samples. DNA sequence quality was checked by eye and all mutations in the combined alignment were again checked against sequence chromatograms.

Samples were also amplified for 10 microsatellite loci (ESM1, Table S4). Amplifications were carried out in 10 μl using 50-100 ng/μl of DNA with 5μl of QIAGEN Multiplex PCR Master Mix, 0.1μl of 0.001 mg/μl purified BSA (New England BioLabs; 10 mg/ml) and 0.2μM of each of the forward and reverse primers. Primers DB44, DB49 and DB52 were amplified in multiplex using as follows: 95ºC for 15 min, followed by 40 cycles at 94ºC for 0.5 min, annealing at 64ºC for 1.0 min, 72ºC for 1.0 min and 72ºC for 30 min. Primers DB1, DB66 and BR6 were amplified in a singleplex as above with annealing temperatures of 60ºC, 55ºC and 49ºC, respectively. Primers SW35, RHI32A, RH17B and RH17C were amplified in multiplex using a touchdown PCR: 95 ºC for 15 min, followed by 10 cycles of 0.5 min at 94ºC, 0.5 min at 48ºC and 0.5 min at 72ºC; 10 cycles with an annealing temperature at 44ºC and 20 cycles with an annealing temperature at 40ºC, followed by 72ºC for 30 min. Each individual was genotyped twice for the historical samples. If both genotypes matched each other exactly we considered the alleles reliable. If alleles did not match, we performed another PCR on the same sample. The error rate (the number of times a PCR failed) was 20% in each round of PCRs.

Microsatellite fragments were scored by eye in GeneMarker v 1.91 (SoftGenetics LLC) and rechecked independently. To cross-calibrate data sets, DNA samples from published specimens were electrophoresed together with samples collected for this study and all museum specimens were re-amplified at least three times to confirm allele scores. The final data set comprised 231 individuals. Data were analysed for scoring errors, allelic dropout and null alleles using Microchecker (Van Oosterhout 2004), for deviation from Hardy-Weinberg expectation and genotypic equilibrium using FSTAT (Goudet 2001). Tests were performed for each population and locus, with Bonferroni correction.

3. Ancient changes in effective population size

Ancient demographic change in both SWR and NWR populations was using in Msvar v.1.3 (Beaumont 1999; Storz & Beaumont 2002). We used a microsatellite mutation rate (μ) of 5×10^{-4} substitutions/site/year (Peery $et\ al.$ 2012), but allowed this rate to vary between 10^{-3} and 10^{-5}. We implemented MCMC simulations, to estimate demographic parameters including current (N_0) and

ancestral (N_1) effective population sizes, the mutation rate per locus per generation (μ) and the time since the population change (T). Prior distributions for N_0, N_1, μ and T were assumed to be log-normal while the means and standard deviations of these log-normal distributions were drawn from the prior (or hyperprior) distributions. Multiple runs were performed with large variances on priors (ESM1, Table S5) in order to influence posterior distributions as little as possible. The three simulations were performed separately for SWR and NWR as follows: (i) the log-normal distribution values of N_1 and N_0 to be the same (stable population), (ii) larger priors for N_1 than N_0 (decline or bottleneck) and (iii) vice versa, smaller priors for N_1 than N_0 (expansion). Simulations were run for 1 x 10^{10} iterations, thinning every 100 000 steps to end with 100 000 thinned samples, with the first 20% discarded as burn-in. We used a generation time of 27 years, calculated as the average between ages of first and last reproduction (Bertschinger 1994). MCMC convergence between independent runs was assessed with the Gelman and Rubin's diagnostic (Gelman and Rubin 1992; Brooks & Gelman 1998). Runs and variables reached convergence when a quantile less than 1.20 was obtained (Gelman *et al.* 2004). All summary statistics were calculated in the R package "Boa" v1.1.7 (Smith 2007).

4. Anthropogenic demographic changes

To investigate potential anthropogenic effects on the recent demography of the white rhinoceros, we used an Approximate Bayesian Computation (ABC) approach (Beaumont et al. 2002) as implemented in ABCTOOLBOX v1.1 (Wegmann *et al.* 2010). Demographic histories for SWR and NWR were first tested independently by exploratory simulations of six scenarios: a null model, two expansion models, two bottleneck models and a model with two bottlenecks (Figure S1). Under expansion and bottleneck scenarios we tested whether the timing of the demographic event coincided with sub-Saharan Africa's two most important anthropogenic events - the migration of iron-age, agriculturalist Niger-Congo language speakers (Bantu) into eastern and southern Africa 400-2,000 years ago (ya; Grollemund et al. 2015) and the expansion of colonial-era European influence into the region (present to 400 ya). We used wide uniform $\log_{10}$ priors to capture a variety of credible parameter values and a Gamma distributed mutation rate prior set between 1 x 10^{-4} and 1 x 10^{-3}. For each scenario, we estimated the following parameters: the modern effective population size (N_0), the ancestral effective population size (N_1) and the time of population size change (T). Summary statistics (mean/total number of alleles per locus, heterozygosity and Garza-Williamson's index for each simulation as well as for the observed data, using a subsample of 20 and 15 individuals for the SWR and NWR populations respectively, were

calculated using Arlequin 3.5. The full set of summary statistics that best explained the variance in the model parameter space were used to calculate the Euclidean distance between observed and simulated datasets. We set the number of simulations at 1,000,000 to assess whether the models were able to explain the observed data. We also checked the fraction of retained simulations (5%) with smaller or equal likelihoods than the likelihood of the observed data (P-value). Model fit was therefore evaluated using the 5,000 simulations closest to the observed dataset. For this we used the ABC-GLM general linear post-sampling adjustment step built into ABCtoolbox to calculate the posterior probability and marginal densities for each scenario. Marginal distributions were used to calculate the posterior probability and Bayes Factors (BF) for each pairwise comparison between scenarios. We accepted the alternative hypothesis when the BF between two scenarios was greater than three. In addition, an R correlation test with Spearman's rho statistics was used to remove summary statistics with highest pairwise correlations. The best models were repeated with 1,000,000 simulations to obtain final posterior parameter estimates.

5. Secondary contact

A two-population null model was constructed (Figure S2), based on the best independent demographic scenarios for SWR and NWR determined above, and tested against several models of migration to determine the likelihood of NWR-SWR secondary contact since divergence. We set four uniform migration priors (ESM1, Table S7) with the expectation that white rhinoceros populations would more likely have come into secondary contact during times of late Pleistocene glaciation, when the grassland biome would have been continuous between eastern and southern Africa. Therefore we first tested for migration (uni- and bidirectional) at any time during the last glacial period (LGP) of the late Pleistocene (14,000 – 106,000 ya, scenarios 2-4) which followed the end of the Eemian interglacial. We then subdivided the LGP to try differentiate between recent migration during the last glacial maximum (LGM, 14,000 – 26,000 ya, scenarios 5-7) and potentially earlier migration during the LGP (26,000 – 106,000 ya, scenarios 8-10). Lastly, we tested the null hypothesis against a model of ancient (pre-Eemian) migration (130,000 – 500,000 ya, scenarios 11-13). Additional two-sample (pairwise) summary statistics (nucleotide diversity, F_{ST}) were computed for simulated and observed data sets. A wide prior, equivalent to the 95% highest posterior density (HPD) interval for the coalescence of SWR and NWR mitochondrial genomes, was used for the time to the most recent common ancestor (TMRCA) between the two populations. We tested this null model against several models of migration to determine the likelihood of NWR-SWR secondary contact since divergence.

Figure S1. Distribution, phenotype and population trends in the white rhinoceros. A. Historic distribution of northern (NWR) and southern (SWR) white rhinoceros in sub-Saharan Africa (after Rookmaaker and Antoine, 2012). B. Northern (left) and southern (right) white rhinoceros females at Ol Pejeta Conservancy, Kenya. Photographic credit: Copyright (c) Chantelle Melzer (www.chantellemelzer.com) 19 May 2017, reproduced online and in print gratuity with full permission from the copyright holder. Population trends of NWR (red) and SWR (blue) through the 20th century. D. Extent of the grassland biome (in red) in Africa during the last glacial maximum (left) when it was continuous from north to south, and at present (right) where it is fragmented.
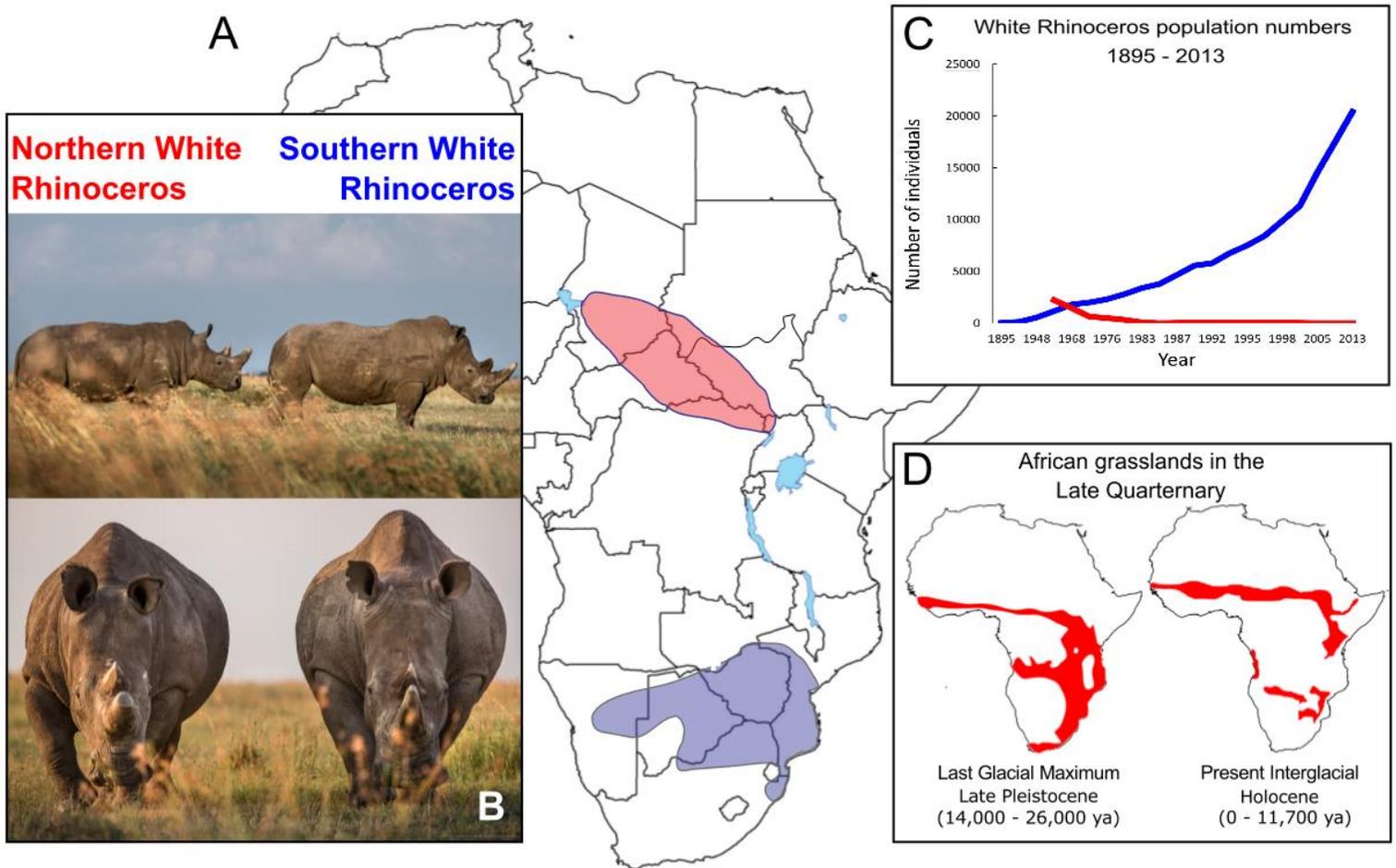
Figure S2. The six demographic scenarios tested for both the southern white and northern white rhinoceros. The first scenario assumes a single stable population and can be regarded as the null hypothesis. Scenario 2-5 modelled two expansions with the time of the expansion during the colonial period (1-15 generations ago) or before the colonial era (15-74 generations ago) and two bottlenecks with the time of the bottleneck during the colonial period (1-15 generation ago) or before the colonial era (15-74 generations ago), respectively. Scenario 6 modelled two bottlenecks without any recovery with declines during the Bantu migration and the colonial era. N_1, ancestral effective population size; N_0, current effective population size; T, time since population size change.
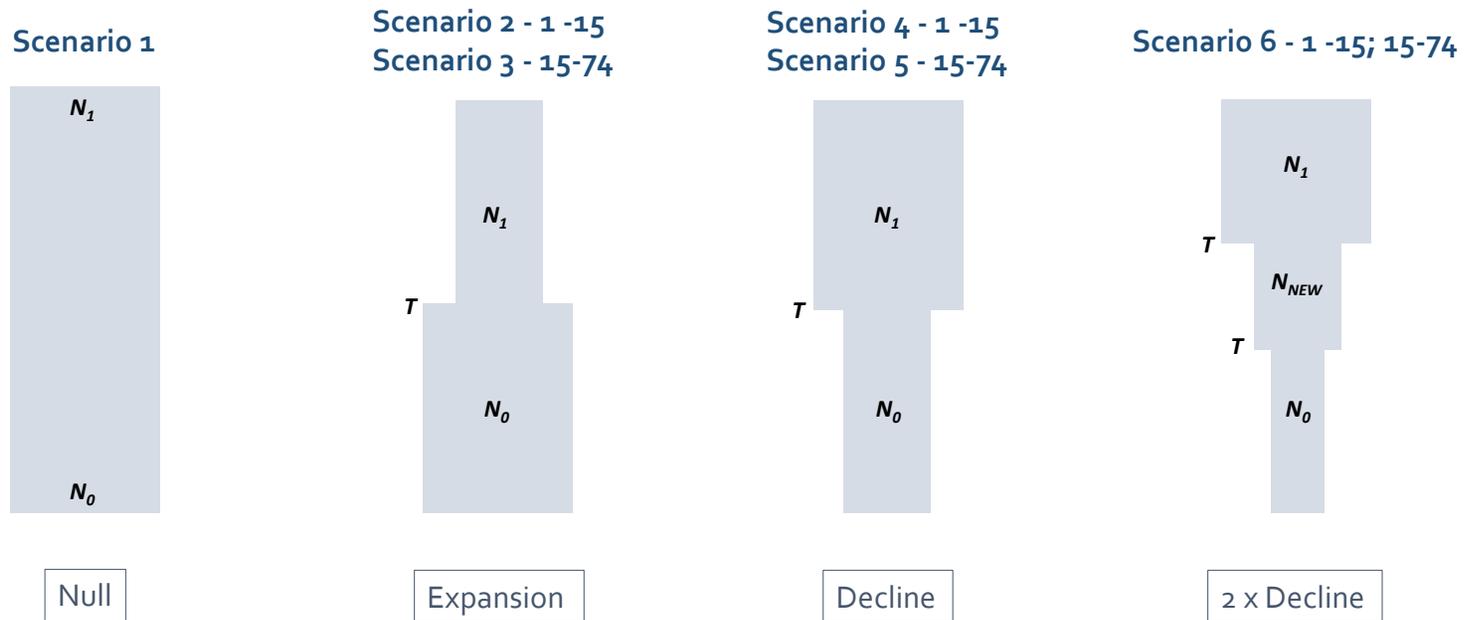
Figure S3. The four demographic scenarios tested for both the southern white and northern white rhinoceros in the combined analysis. The first scenario assumes two stable populations (Southern White Rhino = SWR; Northern White Rhino = NWR) and can be regarded as the null hypothesis. Scenarios modelled unidirectional (scenarios 2, 5, 8,11 and scenarios 3, 6, 9,12) as well as bidirectional (scenarios 4, 7, 10, 13) migration between the southern and northern populations. N_1, ancestral effective population size; N_{0_SWR}, southern white rhinoceros current effective population size; N_{0_NWR}, northern white rhinoceros current effective population size; T, time since population size change. Scenario numbers correspond to those in ESM1, Table S7.
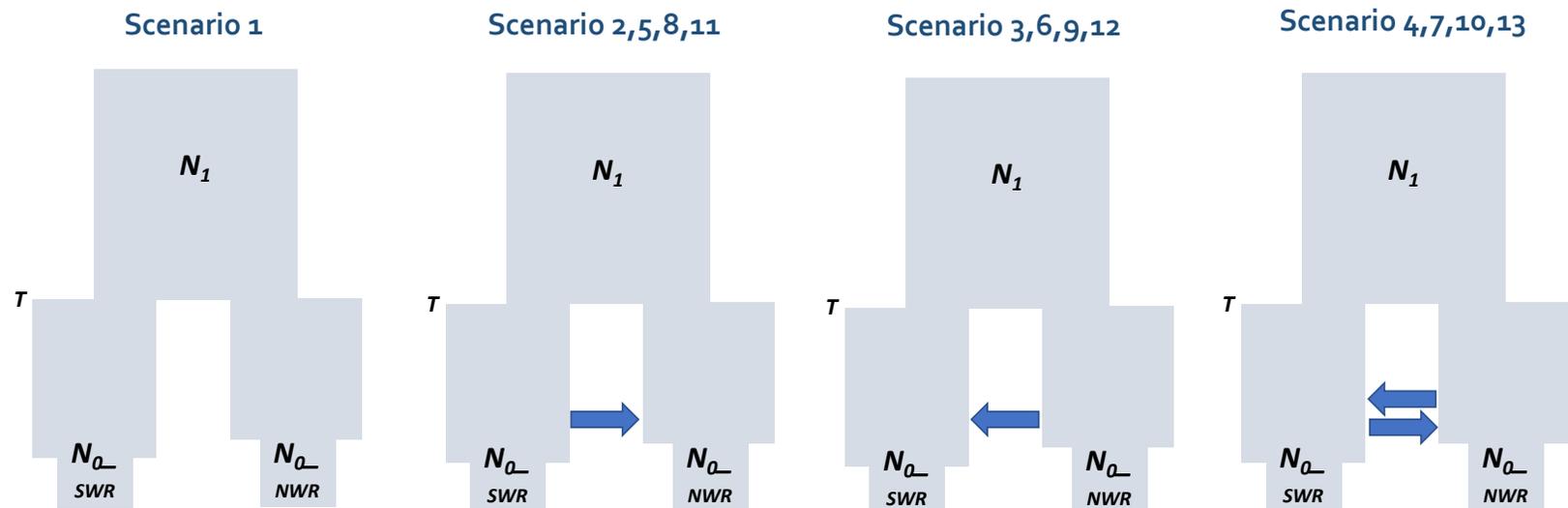
Figure S4. Genetic structure of the white rhinoceros showing clear separation into two populations. Structuring of multilocus nuclear microsatellite profiles (*k* = 1 to *k* = 7).

STRUCTURE plots (K1-7) for white rhinoceros microsatellite data
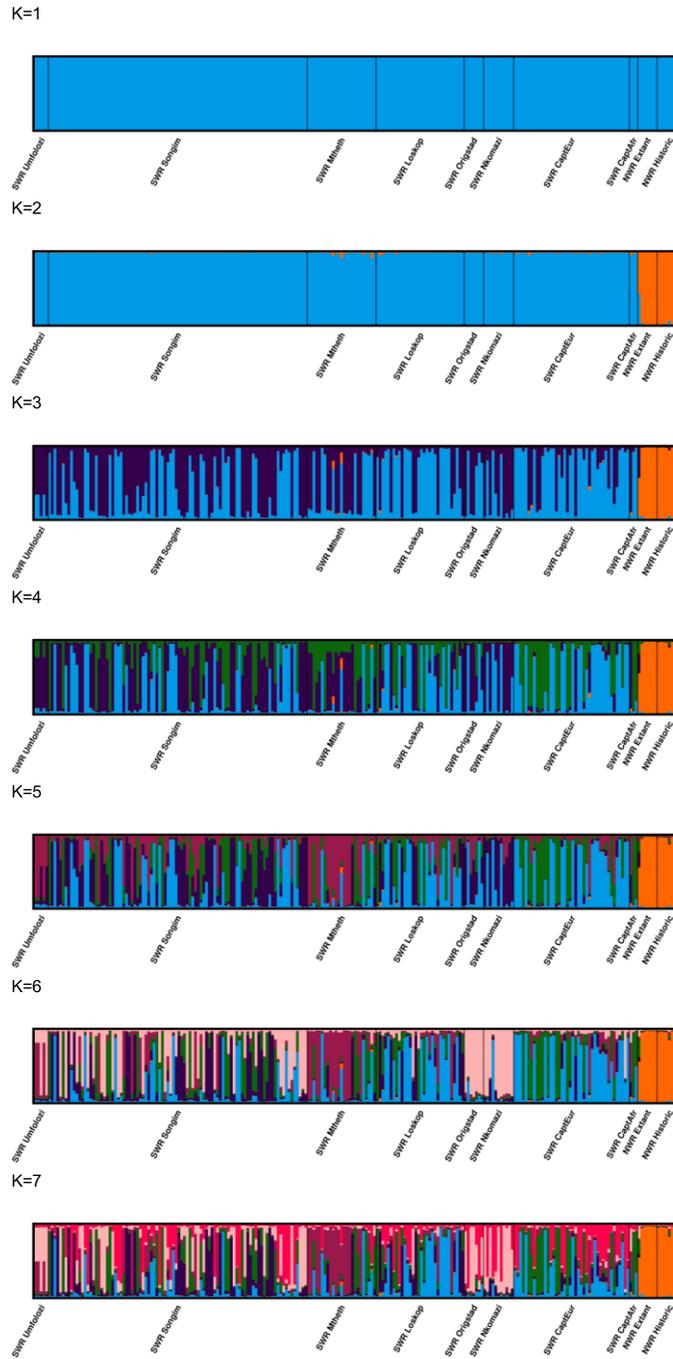
Figure S5. Changes in prehistoric effective population size (N) for southern and northern white rhinoceros populations using a Bayesian coalescent approach. The posterior distributions of current (N_0) and ancestral (N_1) effective population sizes are given for three independent MSVar runs for the southern white rhinoceros (SWR; A) and the northern white rhinoceros (NWR; C) and show a population decline in both cases. Timing of demographic declines for SWR and NWR are indicated in Figures B and D, respectively. Coloured solid lines indicate the posterior distributions whereas the dotted lines represent the prior distributions. For each population, lighter colours depict the posterior distribution of N_1, whereas darker shades show the posterior distribution of N_0. All parameters are given in log10 scale, for median values and the 90% highest probability density (HPD) limits of the posterior distributions see ESM1, Table S9.
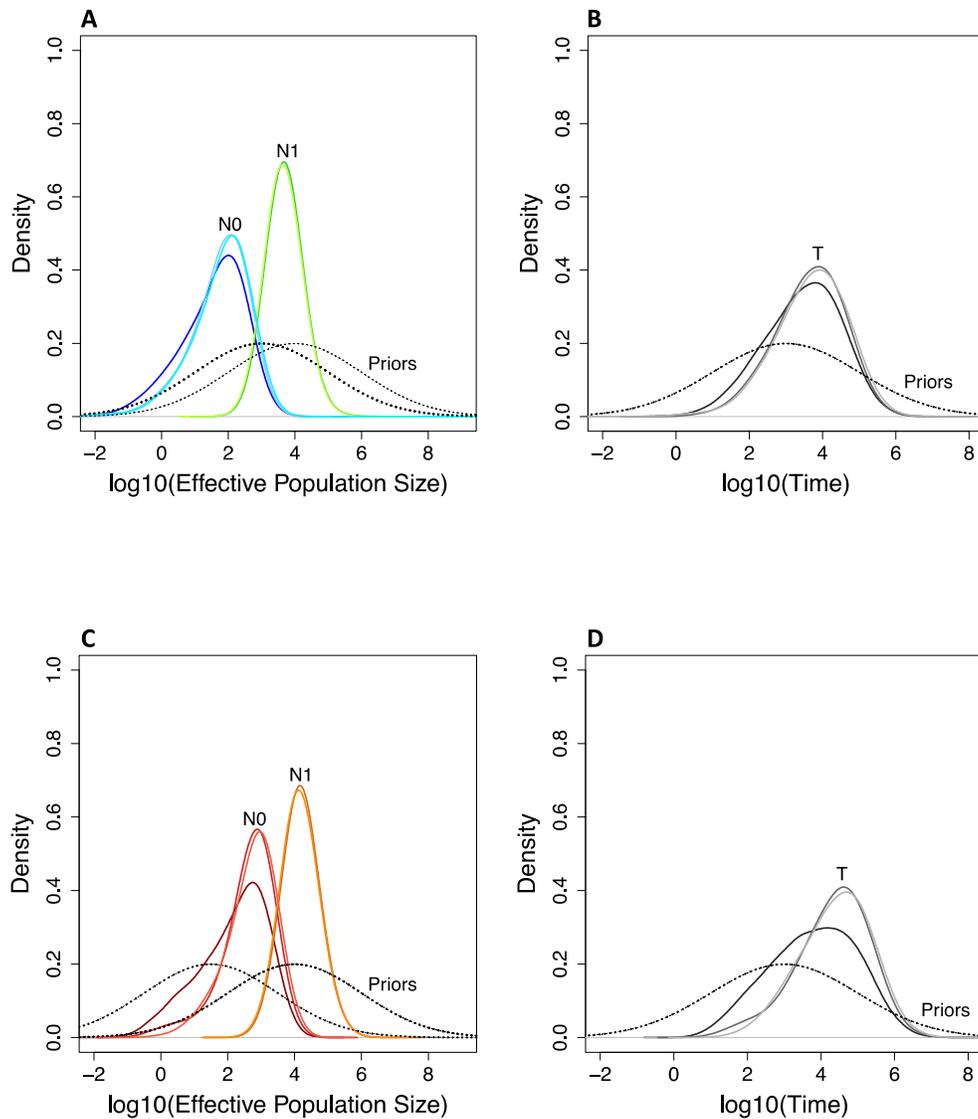
Figure S6. The most likely scenarios for human-induced demographic changes in southern and northern white rhinoceros tested using an approximate Bayesian framework. N_1, ancestral effective population size; N_0, current effective population size, T, time since population size change. Inset figures on the left and right indicate the posterior distributions of the estimated parameters for the SWR and the NWR populations, respectively. MD = marginal density of the posterior distributions.

Southern White Rhinoceros

N_1= 54,705
(HPD90: 9,570-394,548)

T = 264 years
(HPD90: 138-394)

N_0 = 21
(HPD90: 12-29)

MD = 2.04
P-value = 1

Northern White Rhinoceros

N_1 = 97,825
(HPD90: 13,564-792,445)

T = 1,370 years
(HPD90: 518-1,869)

N_0 = 137
(HPD90: 52-305)

MD = 1.04
P-value = 1

N_ANC

N_CUR

Decline
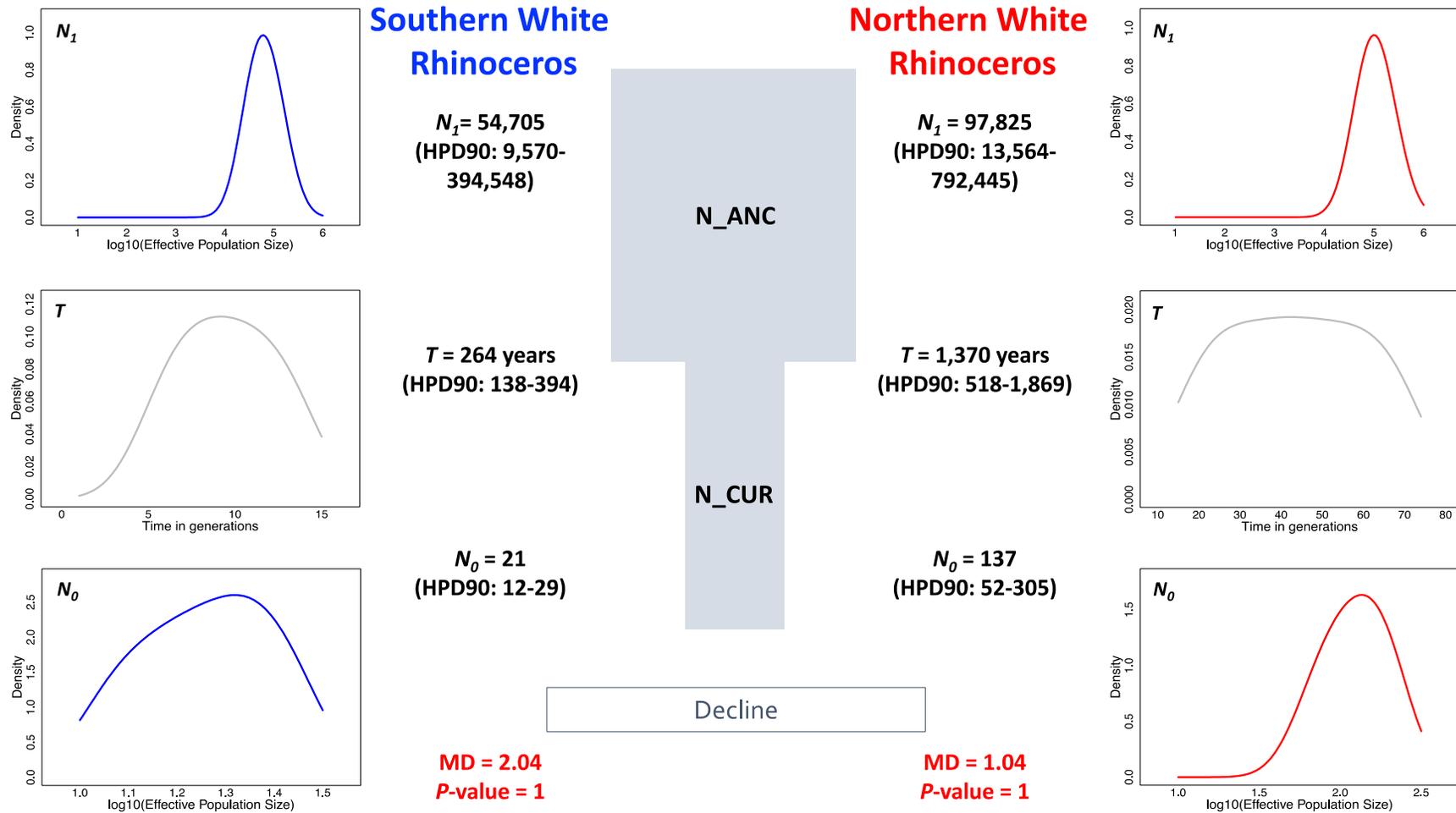
Supplementary References

Robovský, J., Anděra, M. and Benda, P., 2010. Revised catalogue of ceratomorph ungulates in the collection of the National Museum Prague and several other collections in the Czech Republic (Perissodactyla: Rhinocerotidae, Tapiridae). *Lynx, Series Nova* 41(1).

Rohland, N., Siedel, H., & Hofreiter, M. (2010). A rapid column-based ancient DNA extraction method for increased sample throughput. *Molecular ecology resources*, *10*(4), 677-683.

Moodley, Y., Russo, I.M., Dalton, D.L., Kotzé, A., Muya, S., Haubensak, P., Bálint, B., Munimanda, G.K., Deimel, C., Stezer, A., Dicks, K., Herzig-Straschil, B., Kalthoff, D.C., Siegismund, H.R., Robovský, J., O'Donoghue, P. and Bruford, M.W., 2017. Extinctions, genetic erosion and conservation options for the black rhinoceros (*Diceros bicornis*). *Sci. Rep.* 7, p.41417.

Beaumont, M., 1999. Detecting population expansion and decline using microsatellites. *Genetics* 153 (4), 2013–2029.

Storz, J. and Beaumont, M., 2002. Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evol.* 56, 154-166.

Peery, M.Z., Kirby, R., Reid, B.N., Stoelting, R., Doucet-Bër, E., Robinson, S., Va squez-Carrillo, C., Pauli, J.N. and Palsbøll, P.J., 2012. Reliability of genetic bottleneck tests for detecting recent population declines. *Mol. Ecol.* 21(14), 3403-3418.

Bertschinger, H.J., 1994. Reproduction in black and white rhinos: a review. Proceedings of a symposium on rhino as game ranch animals. Onderstepoort, September.

Gelman, A. and Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7(4), 457–472.

Brooks, S.P. and Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. *J. of Comput. Graph. Stat.* 7(4), 434-455.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin D. B., 2004. Bayesian data analysis, 2nd edition. London, Chapman & Hall.

Smith, B.J., 2007. Boa: an R package for MCMC output convergence assessment and posterior inference. *J. of Stat. Softw.* 21(11), 1–37.

Wegmann, D. and Excoffier, L., 2010. Bayesian inference of the demographic history of chimpanzees. *Mol. Biol. Evol.* 27(6), 1425–1435.

Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C. and Pagel, M., 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proc. Natl. Acad. Sci.* 112(43), 13296-13301.