Evaluating recovery potential of the northern white rhinoceros from cryopreserved somatic cells

Tate Tunstall,[1] Richard Kock,[2] Jiri Vahala,[3] Mark Diekhans,[4] Ian Fiddes,[4] Joel Armstrong,[4] Benedict Paten,[4] Oliver A. Ryder,[1,5] and Cynthia C. Steiner[1,5]

[1] San Diego Zoo Institute for Conservation Research, Escondido, California 92027, USA; [2] Royal Veterinary College, University of London, London NW1 0TU, United Kingdom; [3] Dvur Krlov Zoo, Dvr Krlov nad Labem 544 01, Czech Republic; [4] Jack Baskin School of Engineering, University California Santa Cruz, Santa Cruz, California 95064, USA

The critically endangered northern white rhinoceros is believed to be extinct in the wild, with the recent death of the last male leaving only two remaining individuals in captivity. Its extinction would appear inevitable, but the development of advanced cell and reproductive technologies such as cloning by nuclear transfer and the artificial production of gametes via stem cells differentiation offer a second chance for its survival. In this work, we analyzed genome-wide levels of genetic diversity, inbreeding, population history, and demography of the white rhinoceros sequenced from cryopreserved somatic cells, with the goal of informing how genetically valuable individuals could be used in future efforts toward the genetic rescue of the northern white rhinoceros. We present the first sequenced genomes of the northern white rhinoceros, which show relatively high levels of heterozygosity and an average genetic divergence of 0.1% compared with the southern subspecies. The two white rhinoceros subspecies appear to be closely related, with low genetic admixture and a divergent time <80,000 yr ago. Inbreeding, as measured by runs of homozygosity, appears slightly higher in the southern than the northern white rhinoceros. This work demonstrates the value of the northern white rhinoceros cryopreserved genetic material as a potential gene pool for saving this subspecies from extinction.

[Supplemental material is available for this article.]

The worldwide loss of wildlife has been described as the sixth mass extinction (Wake and Vredenburg 2008; Kolbert 2014), with 22% of mammals at risk of extinction (http://wwwiucnredlist.org). Rhinoceroses as a group are particularly affected, with three of the five extant species listed as critically endangered (Javan, Sumatran, and black rhinoceroses), one listed as vulnerable (greater one-horned rhinoceros), and only one, the white rhinoceros (*Ceratotherium simum*), listed as near threatened. Despite continuing threats, including poaching and habitat destruction, the southern population of white rhinoceros known as the southern white rhinoceros (SWR; *C. simum simum*) is currently the most abundant rhinoceros in the world, with about 20,000 individuals living primarily in South Africa (Emslie 2012). Thought to be nearly extinct at the beginning of the 20th century after being reduced to a single population of 20–50 individuals (Emslie 2012), subsequent conservation efforts have led to a dramatic recovery of this subspecies. However, poaching remains a serious threat, and white rhinoceroses are killed at a rate of about two per day, primarily to harvest their horns (Emslie and Knight 2014).

The northern population of white rhinoceros or northern white rhinoceros (NWR; *C. simum cottoni*) is listed as critically endangered and is believed to be extinct in the wild, with the recent death of the last male, Sudan, leaving only two remaining individuals in captivity (Emslie 2012). Extensive poaching and civil war throughout the 20th century has led to the near extinction of this subspecies, which occurred in the Central African Republic, Chad, the Democratic Republic of the Congo, South Sudan, Sudan, and Uganda (Fig. 1A; Rookmaaker and Antoine 2012). The last live wild NWR was seen in 2006, and extensive foot surveys have not discovered any remaining individuals (Emslie 2012).

The recovery of this critically endangered subspecies faces numerous challenges. The remaining two NWR females are not suitable for reproduction, and concerns still persist for the safety of these animals in the wild due to poaching. Saragusty et al. (2016) identified the steps required for a long-term program to "rewind" the extinction process and secure a viable population with sustainable levels of genetic diversity. These steps include estimating the genetic divergence of northern and southern white rhinoceros populations, validating their taxonomic status as subspecies, and assessing the genetic variation in the NWR through genome-wide comparisons with the SWR.

Traditional in situ and ex situ conservation efforts, such as the establishment of a captive breeding program and antipoaching measures, have not been effective in saving the NWR from the brink of extinction. The last NWR calf born in captivity was in 2000 (Saragusty et al. 2016), and efforts at captive breeding have not been successful due to limited reproduction in both the wild-captured founder animals as well as those born in captivity. The fate of the NWR would appear to be sealed, but the development of advanced cell and reproductive technologies, such as cloning by nuclear transfer and the artificial production of gametes via stem cells differentiation, offers a possible path forward (Nayernia et al. 2006; Hayashi and Saitou 2013; Easley et al. 2015; Hendriks et al. 2015; Saragusty et al. 2016). These reproductive technologies could provide new tools for the rescue of endangered, wild populations, especially those from where samples have
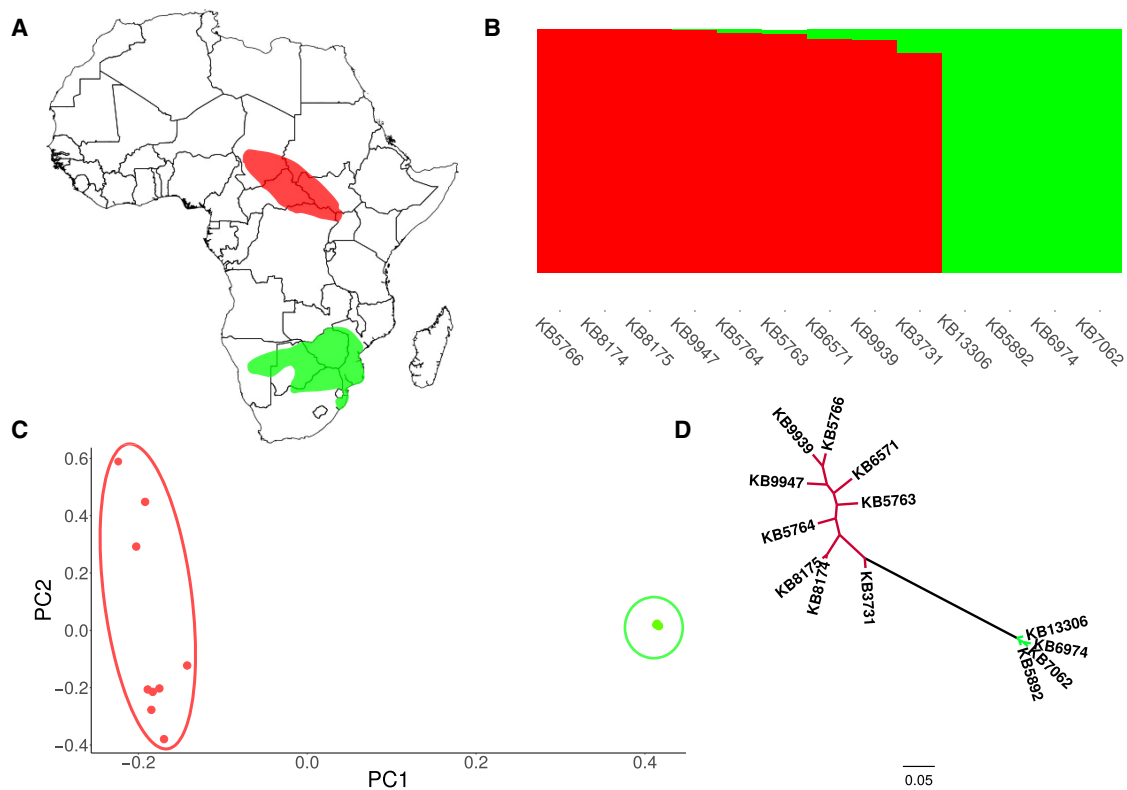
Figure 1. White rhinoceros distribution, population structure, and phylogeny. (*A*) Historical distributions of the northern (red) and southern white rhinoceros (green) in Africa according to work by Rookmaaker and Antoine (2012). (*B*) ADMIXTURE results of the nine northern and four southern white rhinoceroses. All individuals were grouped in two clusters (*K* = 2) colored by population. (*C*) PCA, with PC1 explaining 35% of the variance, and PC2, which explained 10%. (*D*) Maximum likelihood tree of the relationship between northern and southern white rhinoceroses using genome-wide SNPs.

been collected before large declines in genetic diversity. Populations with low genetic diversity can face increased susceptibility to disease (Tasmanian devils) (Morris et al. 2013) and reduced fertility (Felidae [Neubauer et al. 2004]; Iberian lynx [Ruiz-López et al. 2012]); therefore, banking of genetic material in the form of cells and gametes is crucial for future genetic rescue efforts.

Over the past 30 yr, the San Diego Zoo Frozen Zoo has cultured and banked 12 NWR fibroblast cell lines representing eight presumably unrelated founders. These cells correspond to the remaining living genetic material of the NWR, and as proposed by Saragusty et al. (2016), could be used for its genetic rescue. Of course, this approach would be a difficult undertaking, but success could greatly expand the genetic pool of this subspecies beyond the last two remaining individuals and help to develop advanced genetic and reproductive technologies that could also benefit other rhinoceros species in peril. Large-scale habitat destruction has not been the reason for the extirpation of the NWR (Saragusty et al. 2016), so presumably there is potential habitat available for its reintroduction. The greatest obstacles to the successful recovery of the NWR is the continuing threat of poaching in the wild and its small population size.

In this work, we present the complete genome sequences of four SWRs and nine NWRs derived from cryobanked material and provide information on genome-wide levels of genetic diversity and inbreeding among NWRs with the goal of assisting informed-based decisions on genetically valuable material to be used in future efforts of genetic rescue and assisted reproduction of this subspecies. By using genomic data, we inferred the recent population history and demography of the two white rhinoceros subspecies and identified potential signatures of selection in the NWR that may suggest local adaptation.

Results

We sequenced the genomes of nine NWR and four SWR individuals (Table 1). All sequences were generated using Illumina short-read sequencing, and were aligned to the reference SWR genome (cerSim1), which is assembled to the scaffold level with an N50 of 26 Mbp. After calling variants and filtering, about 9.4 million single-nucleotide polymorphisms (SNPs) were identified; 1,723,931 SNPs were polymorphic in both subspecies, 1,869,825 were fixed in both, and 2,511,658 and 4,065,345 were unique SNP variants in the SWR and the NWR, respectively (Supplemental Fig. S2). The average coverage depth of the NWR and SWR genomes sequenced ranged from 10× to 15×.

To examine population structure and admixture in the white rhinoceros, we used ADMIXTURE (Alexander et al. 2009) and EIGENSTRAT (Price et al. 2006) analyses on a marker set of approximately 144,000 SNPs that had been thinned for linkage disequilibrium. Results from ADMIXTURE suggest that the NWR and the SWR subspecies represent two distinct populations with little genetic admixture (Fig. 1B, 10-fold cross validation in Supplemental Fig. S3). These results may reflect historical gene flow between populations but most likely reflect the effect of low level of genetic differentiation between subspecies, homoplasy, or incomplete lineage sorting. Principal component analysis (PCA) also supported

Table 1. White rhinoceros samples used in this study

Studbook no.	Sample no.	Species name	Accession no.	Sex	House name	2n	Sample type	Birth type	Coverage
34	KB13306	*Ceratotherium simum simum*	137	M	KEHTLA	82	Cells	Wild born	13.06
156	KB5892	*Ceratotherium simum simum*	100264	F	MAVULA	82	Cells	Wild born	13.11
24	KB6974	*Ceratotherium simum simum*	61	F	JESSIE	82	Cells	Wild born	12.38
147	KB7062	*Ceratotherium simum simum*	100255	F	MACITE		Cells	Wild born	13.86
28	KB3731	*Ceratotherium simum cottoni*	100282	F	LUCY	82	Cells	Wild born	13.35
377	KB5763	*Ceratotherium simum cottoni*	59006	F	NESARI		Blood	Wild born	13.54
376	KB5764	*Ceratotherium simum cottoni*	689610	F	NADI	82	Cells	Wild born	13.12
372	KB5766	*Ceratotherium simum cottoni*	59001	M	SUDAN	81	Cells	Wild born	13.58
74	KB6571	*Ceratotherium simum cottoni*	100283	M	DINKA	82	Cells	Wild born	12.12
351	KB8174	*Ceratotherium simum cottoni*	59007	F	NASIMA/TWINK	82	Cells	Wild born	14.93
374	KB8175	*Ceratotherium simum cottoni*	689609	F	NOLA	82	Cells	Wild born	9.69
373	KB9939	*Ceratotherium simum cottoni*	689608	M	SAUT	82	Cells	Wild born	11.62
348	KB9947	*Ceratotherium simum cottoni*	690500	M	ANGALIFU	82	Cells	Wild born	12.93

Description includes the sample, accession and studbook numbers, sex and name of specimen, diploid chromosome number (2n), genetic material used for DNA extraction, birth type, and genomic coverage.

the two populations as distinct clusters (Fig. 1C). PC1 (35% of the variance) separates the rhinoceroses into two distinct populations, whereas PC2 (10% of the variance) evidences genetic variation in the NWR. By using the same SNP data, we generated a maximum likelihood tree using SNPhylo (Lee et al. 2014) that showed strong support for two distinct northern and southern clades (100%, in 100 bootstrapped samples) (Fig. 1D) but weaker support for any within-subspecies phylogenetic relationships. Similar phylogenetic results were obtained after comparing complete mitochondrial genomes from both subspecies (Supplemental Fig. S4).

As a measure of genetic variation, we estimated genome-wide heterozygosity calculated as the proportion of heterozygous sites in the genome of each individual. We found that the mean genome-wide heterozygosity in the NWR was slightly higher (0.0011) than that of the SWR samples studied (0.0009). Both subspecies had relatively high levels of genetic variation compared with other threatened species such as the Tasmanian devil or the Bornean orangutan (Fig. 2). We also found an average total genetic divergence of 0.001 (0.1%) between the two subspecies, which is similar to values calculated for other mammalian subspecies (see Supplemental Table S1).

To infer the divergence time and demographic history of the northern and southern subspecies, we used pairwise sequentially Markovian coalescent (PSMC) (Fig. 3A; Li and Durbin 2011). Both populations initially appear to share a similar demographic history, undergoing a relatively ancient decline in population size ~800,000 yr ago (kya), followed by population recovery start-

ing ~100 kya. The white rhinoceros then split into two populations ~80 kya with variable demographic trajectories. The NWR increases in effective population size, followed by a decline, while the SWR declines in effective population size, followed by a recovery to a larger size than the NWR. The NWR population seems to have reached a maximum effective population size of 15,000 individuals ~40 kya, while the SWR reached a maximum effective population size of 10,000 individuals ~50 kya.

To further explore demographic scenarios, we also tested a series of increasingly complex demographic models using ∂a∂i (Gutenkunst et al. 2009). In each model, the two populations split either into a fraction of the ancestral population size or into two distinct populations and then grew or shrank to the current population sizes. The most well supported model proposes a split of the NWR and SWR populations into a fraction of the ancestral population, followed by a change in effective population size to the recent numbers (see Supplemental Table S3). We assumed a mutation rate of 2.5×10^{-8}, which has been estimated for humans (Nachman and Crowell 2000), and a generation time of 8 yr (Hillman-Smith et al. 1986). This model estimated a divergence time of ~10–20 kya for the two white rhinoceros populations, which is lower than the 80 kya estimate from the PSMC analysis. Our model showed that both populations split from an ancestral effective population of 16,000 individuals, followed by growth in both the NWR and SWR. Current effective population size estimates correspond to 1300 and 2800 for the NWR and SWR, respectively (Fig. 3C).

To determine inbreeding in white rhinoceroses, we calculated runs of homozygosity (ROH) by scanning 1-Mbp sliding windows throughout the genomes. Our analysis showed similar levels of autozygosity (F_{roh}) in the two white rhinoceros subspecies, with estimates of F_{roh} slightly higher in the southern subspecies overall (Fig. 4A). Among the NWRs examined, all individuals except one (KB8174) showed the lower levels of inbreeding, with between 2% and 3% of their genome being autozygous; in contrast, three of the four SWR showed autozygosity >3% (KB6974, KB5892, and KB13306). Additionally, both the NWR and the SWR had a median ROH length of 1.6 Mbp, but the SWR had a larger maximum ROH length of 32 Mbp versus 23 Mbp in the NWR (Fig. 4B; Supplemental Fig. S5), which suggests that the differences in autozygosity may be due to recent inbreeding in the SWR. These estimators, F_{roh} and the low mean size of ROH, in addition to the number of unique SNPs per individual (see Supplemental Table S2), allowed us to identify NWR cell lines particularly
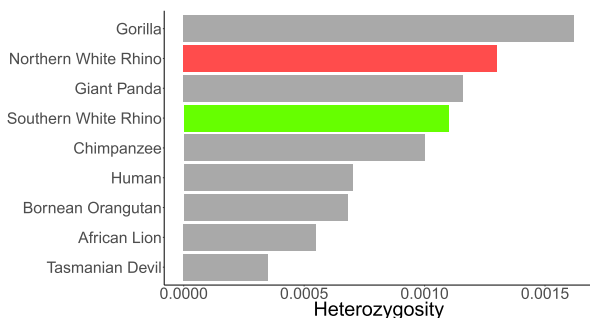
Figure 2. Estimates of genome-wide diversity in the northern (red) and southern (green) white rhinoceros relative to other mammalian genomes. Heterozygosity in each species is based on work by Dobrynin et al. (2015).
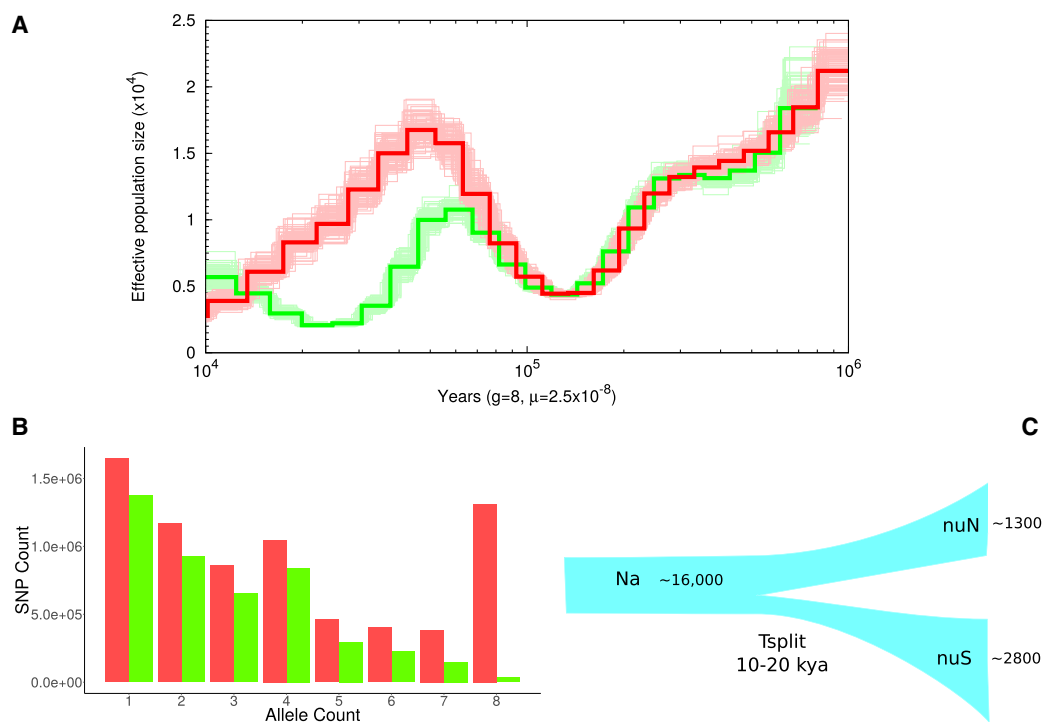
Figure 3. Demographic history and divergence time of the white rhinoceros. (*A*) PSMC plot for one northern (KB8174; red) and one southern (KB7062; green) white rhinoceros genomes with highest coverage (15× and 14×, respectively), assuming a generation time (g) of 8 yr, a mutation rate (μ) of 2.5 × 10^{-8}, and a false-negative rate of 6% due to low coverage. Lighter lines represent 100 bootstrap samples. (*B*) Site-frequency spectra of the four northern (KB3731, KB5763, KB6571, KB8174) and four southern white rhinoceroses used for the ∂a∂i analysis, showing the number of SNPs with the respective allele count in each subspecies. (*C*) The best-supported ∂a∂i model showing divergence of the two rhinoceros populations with low rates of migration and population growth. Ancestral effective population size (Na), current effective population sizes for the northern (nuN) and southern (nuS) white rhinoceros, and the split of both populations (Tsplit) are shown.

valuable for any attempt of genetic rescue. These samples include individuals KB3731, KB9939, KB6571, and KB5764, which have low levels of autozygosity, a low mean size of ROH, and a high number of unique SNPs.

To identify regions in the NWR genome potentially under selection, we calculated Tajima's D in sliding windows of 50 kbp and identified regions where the calculated values fell within the 1% quantiles as potentially under selection (Supplemental Fig. S6). We identified all coding SNPs that occurred in these regions according to the generated white rhinoceros genome annotation. This resulted in about 240 SNPs in 28 genes potentially identified under selection (Supplemental Table S4). Among those, we found 100 SNPs that were fixed differences in either subspecies. None of the SNPs were fixed in both subspecies, suggesting soft selective sweeps. We then identified the gene ontology categories for each gene using Ensembl, finding the olfactory receptor genes as the most common category, with 11 genes involved in sensory perception of smell potentially under selection (Table 2).

Discussion

Whole-genome sequencing has the potential to assist conservation and management efforts by providing information on the taxonomic status and demographic history of populations and by estimating genome-wide levels of genetic variation associated with local adaptation and inbreeding (Allendorf et al. 2010; Steiner et al. 2013). Our work presents the first complete NWR genomes, representing the current gene pool of this subspecies.

These genomes provided estimates of genome-wide levels of genetic diversity and inbreeding that may inform decisions on the genetic value of cells to be used in genetic rescue and assisted reproduction efforts. We also examined the recent population history and demography of these two white rhinoceros subspecies and identified potential regions of selection in the NWR that may suggest local adaptation and divergent evolution from the SWR. Because of the comparative approach we employed using two closely related populations, this work may be the first to use genome-wide analysis as an indicator of the recovery potential of an endangered species, which is relevant for evaluating extinction risk and conservation recovery strategies (Frankham et al. 2010).

Information on the taxonomic relationship, population structure, and divergence time between the NWR and SWR is relevant for designing conservation strategies that will assist rescuing the NWR. Recent genetic evidence using complete mitochondrial genomes (Harley et al. 2016) suggests the NWR and SWR represent subspecies of white rhinoceros, while others have used morphological differences to support both populations as distinct species (Groves et al. 2010). Our genome-wide analyses found modest levels of genetic divergence between the NWR and SWR, even though these subspecies show distinct genetic structure. The level of genomic divergence between the NWR and SWR appears consistent with other mammalian subspecies such as chimpanzees (0.0019) and gorillas (0.0016) (Prado-Martinez et al. 2013).

Estimates from demographic analyses suggest that these two subspecies diverged between 10 and 80 kya, with little or no recent gene flow. Population divergence times estimates differ depending
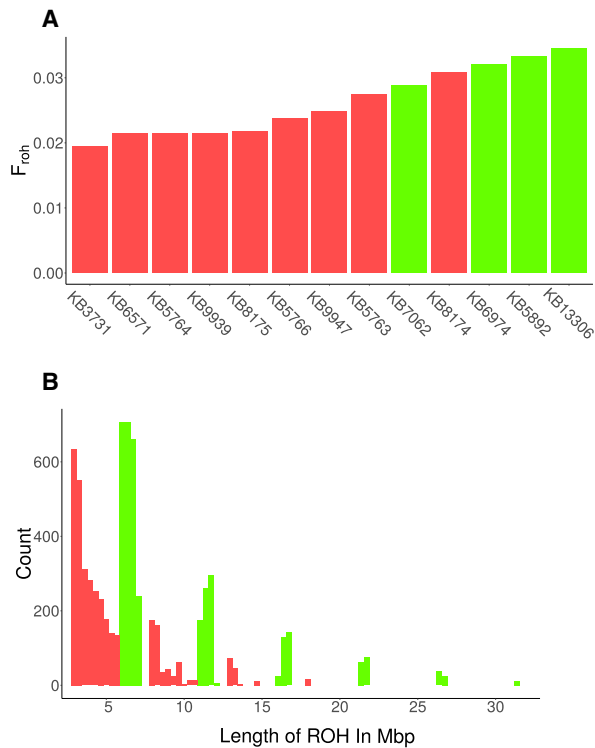
Figure 4. Inbreeding in the white rhinoceros. (*A*) Estimated percentage of genome-wide autozygosity (F_{roh}) for nine northern (red) and four southern (green) white rhinoceroses. (*B*) Length distribution of ROH in each NWR and SWR (individual order from *left* to *right*: KB8174, KB5763, KB5766, KB133068, KB8175, KB3731, KB5764, KB6571, KB9939, KB13306, KB6974, KB5892, KB7062), grouped by ROH lengths from 5 to 30 Mbp.

bryo implantation. The genetic diversity represented in the NWR preserved cell lines at the San Diego Zoo Frozen Zoo is comparable to that of the SWRs, with higher levels of genome-wide heterozygosity and slightly lower levels of autozygosity in the NWR, consistent with the fact that the SWR has undergone a population bottleneck in the last century. The SWR population is thought to have been as low as 20–50 individuals at the beginning of the 20th century, and it has relatively recently recovered within the last 50 yr (Emslie 2012). These years of low population size likely led to the increased inbreeding and reduced genetic diversity seen in this study. In contrast, the decline of the NWR population happened relatively recently, with the population thought to be around 2360 in 1960 (Emslie and Brooks 1999) but has rapidly declined since then due to intense poaching (Emslie 2012).

The NWR samples analyzed in this study likely represent individuals from generations immediately after the recent population decline, before drift and inbreeding could reduce genetic diversity. The SWR samples likely represent the recent expansion of the population, as shown by the increased levels of inbreeding found in these individuals. The SWR samples show long tracts of homozygosity compared with that of the NWR, consistent with recent inbreeding. However, overall levels of autozygosity in both subspecies are low compared with other inbred species, such as the mountain gorilla (Xue et al. 2015) and Scandinavian wolves (Kardos et al. 2018), suggesting that inbreeding occurred recently and over relatively few generations. Genetic estimates of ROH will be important to any future rescue program for the NWR, as whole-genome sequencing provides a better estimate of inbreeding than pedigree analysis (Kardos et al. 2018).

To identify potentially adaptive trends between white rhinoceros subspecies, we identified a set of 28 genes as likely under selection in the NWR. Among the genes, several are associated with olfaction or smell perception, which have been found to quickly evolve during speciation in other mammalian species, including primates and pandas (Moyle 2005; Zhao et al. 2013; Paudel et al. 2015). It is possible that these regions have been under recent selection or soft selective sweep in the NWR and represent genomic areas of potential adaptive variation (Messer and Petrov 2013). These results are preliminary, but rapidly evolving olfactory receptors may be linked to any differences in ecology and behavior. While Groves et al. (2010) concluded that reported behavioral and ecological observations do not provide a clear distinction between the two subspecies, our results point to regions of the genome where further investigation is warranted.

We believe that the current NWR genetic material banked at the San Diego Zoo Frozen Zoo in the form of cell lines is promising for future genetic rescue efforts in saving this subspecies from extinction. This approach would involve the use of methods such as the artificial production of gametes by directed differentiation of pluripotent stem cells (PSCs) in vitro, or combined with maturation in vivo into germline stem/progenitor cells (Nayernia et al. 2006; Hayashi and Saitou 2013; Easley et al. 2015; Hendriks et al. 2015), and cloning by nuclear transfer of cryopreserved material (Saragusty et al. 2016). The possibility of generating artificial gametes from PSCs in mice has been demonstrated with fertile offspring born from gametes generated this way (Nayernia et al. 2006; Hayashi et al. 2011; Zhu et al. 2012; Hayashi and Saitou 2013), and further experiments indicate that gametes could be generated this way in other species (Aflatoonian et al. 2009; Eguizabal et al. 2011; Panula et al. 2011).

Interspecies somatic cell nuclear transfer can also be used to clone endangered species, which has been demonstrated, for

on the method used (Zhou and Teo 2015); therefore, the variation in divergent estimates may be due to the fact that ∂a∂i infers split time from a population's allele frequency spectrum, while PSMC makes inferences based on the local density of heterozygotes across the genome. The more recent history of the southern white rhinoceros, such as the large bottleneck in the last century (Saragusty et al. 2016), could not be detected by using PSMC, presumably because there are too few recombination events in the genome to detect these recent changes in population size (Li and Durbin 2011). It should also be noted that PSMC can confound population structure with changes in effective population size and is less accurate in estimating those changes in the recent or very ancient past (Li and Durbin 2011).

The north/south split in population structure shown between the NWR and SWR is similar to the pattern observed among many other African ungulates (Lorenzen et al. 2012), thought to be driven by pluvial cycles. Lorenzen et al. (2012) found that for those species whose ranges extended into southern Africa, all but buffalo were structured into subspecies north and south of the equatorial forest belt. This suggests that a vicariance event such as the expansion of tropical forests across Central Africa during glacial cycles could have driven the common north/south divergence seen in many taxa.

The relatively close genetic relationship of the two white rhinoceros subspecies is promising for the potential success of reproductive attempts in rescuing the NWR by using the SWR as model population for physiological studies and surrogate females for em-

Table 2. Olfactory receptor coding SNPs identified as outliers by the Tajima's *D* analysis in the northern white rhinoceros

Gene description	Ensembl ID	SNP count	HGNC symbol	Gene Ontology ID
Olfactory receptor family 4 subfamily E member 2	ENSCAFG00000005679	2	OR4E2	GO:0004871, GO:0004888, GO:0004930, GO:0004984, GO:0005886, GO:0007165, GO:0007186, GO:0007608, GO:0016020, GO:0016021, GO:0050896, GO:0050907, GO:0050911
Olfactory receptor family 52 subfamily N member 2	ENSCAFG00000006321	2	OR52N2	GO:0004871, GO:0004930, GO:0004984, GO:0005886, GO:0007165, GO:0007186, GO:0007608, GO:0016020, GO:0016021, GO:0050896, GO:0050911
Olfactory receptor family 10 subfamily V member 1	ENSCAFG00000007561	1	OR10V1	GO:0004871, GO:0004930, GO:0004984, GO:0005549, GO:0005886, GO:0007165, GO:0007186, GO:0007608, GO:0016020, GO:0016021, GO:0050896, GO:0050911
Olfactory receptor family 5 subfamily G member 3	ENSCAFG00000007988	8	OR5G3	
Olfactory receptor family 2 subfamily W member 1	ENSCAFG00000012133	17	OR2W1	GO:0004871, GO:0004930, GO:0004984, GO:0005886, GO:0007165, GO:0007186, GO:0007608, GO:0016020, GO:0016021, GO:0050896, GO:0050911
Olfactory receptor family 8 subfamily K member 1	ENSCAFG00000028823	64	OR8K1	GO:0004871, GO:0004930, GO:0004984, GO:0005549, GO:0005886, GO:0007165, GO:0007186, GO:0007608, GO:0016020, GO:0016021, GO:0050896, GO:0050911
Olfactory receptor family 5 subfamily M member 3	ENSCAFG00000032727	8	OR5M3	GO:0004871, GO:0004930, GO:0004984, GO:0005549, GO:0005886, GO:0007165, GO:0007186, GO:0007608, GO:0016020, GO:0016021, GO:0050896, GO:0050911
Olfactory receptor family 56 subfamily A member 1	ENSBTAG00000000368	2	OR56A1	GO:0001591, GO:0001963, GO:0004871, GO:0004930, GO:0004984, GO:0005886, GO:0005887, GO:0007165, GO:0007186, GO:0007194, GO:0007195, GO:0007608, GO:0009636, GO:0014059, GO:0016020, GO:0016021, GO:0030672, GO:0035240, GO:0042493, GO:0043266, GO:0048148, GO:0048149, GO:0050896, GO:0050911, GO:0051481, GO:0051967, GO:0060158, GO:1901386
Olfactory receptor family 2 subfamily S member 2	ENSBTAG00000032670	47	OR2S2	GO:0004871, GO:0004930, GO:0004984, GO:0005886, GO:0007165, GO:0007186, GO:0007608, GO:0016020, GO:0016021, GO:0050896, GO:0050911
Olfactory receptor family 2 subfamily D member 3	ENSBTAG00000038518	37	OR2D3	GO:0004871, GO:0004930, GO:0004984, GO:0005886, GO:0007165, GO:0007186, GO:0007608, GO:0016020, GO:0016021, GO:0050896, GO:0050911
Olfactory receptor family 56 subfamily B member 4	ENSBTAG00000047176	11	OR56B4	GO:0004871, GO:0004930, GO:0004984, GO:0005886, GO:0007165, GO:0007186, GO:0007608, GO:0016020, GO:0016021, GO:0050896, GO:0050911

Gene description, Ensembl identifier, number of SNPs, gene symbol and Gene Ontology (GO) identifier are indicated.

example, in the wild ox, *Bos gaurus* (Lanza et al. 2000), and an extinct wild goat subspecies, *Capra pyrenaica pyrenaica* (Folch et al. 2009). Somatic nuclear transfer has also been suggested as a way to save other endangered wildlife populations in which the genetic diversity is thought to be insufficient for survival, such as the black-footed ferret (Wisely et al. 2015).

Other approaches have also been considered for genetic rescue of the NWR, including inter-crossing SWR with NWR from the wild (Emslie 2012) or from cryopreserved gametes (Saragusty et al. 2016). However, the single known hybrid between northern and southern white rhinoceroses was Nasi, born 1977 and died in 2007. This individual reached adulthood but never reproduced and was in relatively poor health (Groves et al. 2010). Nothing in our findings would invalidate the interbreeding of NWR and SWR; however, given the death of the one remaining male NWR and the limited number of NWR cryopreserved gametes, the amount of NWR genetic variation that could be maintained through interbreeding is likely limited. While some might argue against efforts to genetically rescue the NWR subspecies given the close evolutionary relationship with the SWR, this makes using the SWR as a surrogate more likely to succeed, and the benefits and lessons of any genetic rescue effort could apply to other rhinoceros species, as well as to mammalian species with similar life history and conservation concerns.

The NWR is extinct in the wild, and the two remaining female rhinoceroses cannot reproduce naturally. Previous investigators have suggested that the best hope to save this subspecies rests in the use of a genetic rescue approach taking advantage of the genetic lineages preserved in the San Diego Zoo Frozen Zoo (Saragusty et al. 2016). Here we have shown the genetic diversity represented by these preserved cell lines is comparable with that of the SWR samples used in this study. At the beginning of the last century, the SWR teetered on the brink of extinction, but through dedicated conservation efforts was able to recover and is now the largest population of rhinoceros in the world. The potential is there for the NWR to make a similar recovery. The newly sequenced genomes could allow a genetic rescue program to maximize retention of genetic diversity and minimize inbreeding. While technological and logistical hurdles certainly remain, this work shows that time may not have run out on the NWR.

Methods

Samples

We selected a total of 13 wild-born white rhinoceros samples, four SWR and nine NWR, none of which were related according to the known pedigree (Supplemental Fig. S1). Eight NWR samples were

derived from cell lines preserved in the San Diego Zoo Frozen Zoo, and one sample corresponds to a zoo-based individual for which no viable frozen cells exist and only DNA were available. Of the NWR samples, seven individuals were from Sudan and two from Uganda (Christman 2011). All white rhinoceros cell lines were cultured, harvested, and chromosome banded following the techniques described by Houck et al. (1995). One individual had a diploid ($2n$) chromosome number of 81, a known variant in white rhinoceroses of the most common $2n = 82$ karyotype (Houck et al. 1994). Utilization of samples was compliant with applicable regulatory procedures for CITES and the US Endangered Species Act. DNA was extracted using the DNeasy cell line kits (Qiagen) according to the manufacturer's instructions.

Sequencing

Nine NWR and four SWR were sequenced by Cofactor Genomics using Illumina paired-end sequencing to a coverage of 10×–15×. Briefly, Genomic DNA libraries were constructed by shearing genomic DNA to the desired size using the Covaris S2 (Covaris). Following shearing, DNA was end-repaired and A-tailed to prepare for adaptor ligation. Indexed adaptors were ligated to sample DNA, and the adaptor-ligated DNA was then size-selected on a 2% SizeSelect E-Gel (Invitrogen) and amplified by PCR. Library quality was assessed by measuring nanomolar concentration and the fragment size in base pairs. Cluster generation and the following sequencing were performed according to the cluster generation manual and sequencing manual from Illumina (https://supportillumina.com/content/dam/illumina-support/documents/documentation/system_documentation/cluster_station/Cluster Station_UserGuide_15018818_D.pdf). Base calls were generated using CASAVA 1.8.2 (Illumina), and the resulting demultiplexed sequence reads were filtered for low quality.

All samples were aligned to the SWR reference genome (cerSim1; accession ID: GCA 0002831551.1) using SAMtools mpileup (Li et al. 2009), and variants were called using the multiallelic calling mode in BCFtools (bcftools m). Variants were filtered using the following criteria:

- variants within 3 bp of an indel;
- clusters of indels separated by ≤10 bp, allowing only one to pass;
- quality score <10;
- Mann-Whitney $U < 0.1$ and quality <15; and
- ancestral count <2 and quality <15.

In order to identify scaffolds in the rhinoceros genome corresponding to the X Chromosome, we first attempted to BLAST all scaffolds against the horse X Chromosome. However, a large number of scaffolds contained sequences highly similar to the horse X (see Supplemental Table S5) and were subsequently included in the analysis.

Population structure and phylogeny

We used the software ADMIXTURE (Alexander et al. 2009) to detect population structure and levels of admixture between the northern and southern subspecies. We used a set of approximately 144,000 SNPs that were thinned for potential linkage disequilibrium using the software PLINK v1.9 (Gaunt et al. 2007; Chang et al. 2015). Specifically, we used a 50-SNP sliding window, advanced 10 SNPs at a time, and removed any SNPs with and R^2 value of >0.1. We used this data set to perform 10-fold cross validation using K values ranging from one to five. The same set of markers were used to perform PCA using EIGENSTRAT (Price et al. 2006), which corrects for population stratification. These markers were further used to

generate a maximum likelihood tree using SNPhylo (Lee et al. 2014) and 100 nonparametric bootstrap samples.

Demographic history and divergence time

PSMC (Li and Durbin 2011) was used to estimate historical population size and divergence of the NWR and SWR. Here, we used a generation time of 8 yr and mutation rate of 2.5×10^{-8} (substitutions/site/yr) (Hillman-Smith et al. 1986). In order to account for low genome coverage, we down-sampled our 15× genome to 10× and found a nearly identical loss of heterogeneity to that of the horse genome. We therefore used a false-negative rate of 0.06 for our 15× rhinoceros genomes used in the analysis, as previously described (Orlando et al. 2013).

We also estimate demographic history using ∂a∂i by testing a series of four increasingly complex models (Gutenkunst et al. 2009). We selected the four NWR with the highest coverage and, along with the four SWR, calculated the site-frequency spectrum (Fig. 3B). The simplest model consisted of a split into two populations, with no growth or migration between populations, and the complex model consisted of a split into two populations, with change in size from past to present and migration between populations. A mutation rate of 2.5×10^{-8} was also used for inferring scaled population parameters. Likelihood and parameter estimates for each model can be found in the Supplemental Materials.

Levels of inbreeding within individuals in the northern and southern subspecies were evaluated by analyzing genome-wide ROH. To identify ROH, we calculated genome-wide heterozygosity in overlapping regions of 1 Mbp with 200-kbp sliding windows. We plotted the density of heterozygosity in each window, which showed regions of low heterozygosity in both the SWR and NWR. By using previous estimates from the literature (Prado-Martinez et al. 2013) and visualizing the data using density plots, we used a threshold value of 0.0004, or 400 per 1 Mbp, to determine the window that could be considered a ROH. We then calculated the percentage of the genome that could be considered autozygous (F_{roh}). The length distribution of ROH was calculated using R (R Core Team 2016). The length is defined as the number of consecutive, overlapping 1-Mbp ROH, determined as described above.

Gene annotation

RNA-seq data were generated from brain, testis, oviduct, and fibroblast cells using paired-end reads in the Illumina MiSeq platform. We lifted the Ensembl 85 (Yates et al. 2016) release annotations for both cow and dog to the rhinoceros assembly using a progressive Cactus alignment (Paten et al. 2011). We then fed the protein coding transcripts as hints to AUGUSTUS (Stanke et al. 2006) along with RNA-seq data on a per-transcript basis. The output of this underwent a consensus finding tool with the original mapped transcripts that performs filtering and determines if we have a high-quality ortholog in any of our input sets (Fiddes et al. 2018).

Selection

We calculated Watterson's estimate of θ and θπ for nonoverlapping windows of 50 kb in our nine NWR genomes and calculated Tajima's D for each window. We then selected the regions in the outlying 1% quantile using R (R Core Team 2016), representing regions potentially under balancing or positive selection. We used the annotation available for the white rhinoceros on the UCSC genome browser (cerSim1) to identify all genes occurring in these outlier regions. We used biomaRt (Durinck et al. 2005, 2009) to identify the HGNC code and Gene Ontology for all genes that contained SNPs that fall within transcribed regions. Specific

enrichment for a particular class of genes was identified using PANTHER (Mi et al. 2013).

Data access

The data generated from this study have been submitted to the NCBI BioProject (https://wwwncbi.nlm.nih.gov/bioproject/394025) under accession number PRJNA394025.

Acknowledgments

References

Aflatoonian B, Ruban L, Jones M, Aflatoonian R, Fazeli A, Moore H. 2009. *In vitro* post-meiotic germ cell development from human embryonic stem cells. *Hum Reprod* **24:** 3150–3159.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19:** 1655–1664.

Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation genetics. *Nat Rev Genet* **11:** 697–709.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation plink: rising to the challenge of larger and richer datasets. *Gigascience* **4:** 7.

Christman J. 2011. *International studbook for the white rhinoceros Ceratotherium simum (Burchell 1817)*, Vol. 12 (status 31.12.2011), pp. 1–139. Disney's Animal Kingdom, Bay Lake, FL. wwwrhinoresourcecenter.com.

Dobrynin P, Liu S, Tamazian G, Xiong Z, Yurchenko AA, Krasheninnikova K, Kliver S, Schmidt-Küntzel A, Koepfli K-P, Johnson W, et al. 2015. Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome Biol* **16:** 277.

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. 2005. Biomart and bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21:** 3439–3440.

Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRT. *Nat Protoc* **4:** 1184–1191.

Easley CA, Simerly CR, Schatten G. 2015. Gamete derivation from embryonic stem cells, induced pluripotent stem cells or somatic cell nuclear transfer-derived embryonic stem cells: state of the art. *Reprod Fertil Dev* **27:** 89–92.

Eguizabal C, Montserrat N, Vassena R, Barragan M, Garreta E, Garcia-Quevedo L, Vidal F, Giorgetti A, Veiga A, Belmonte J, et al. 2011. Complete meiosis from human induced pluripotent stem cells. *Stem Cells* **29:** 1186–1195.

Emslie R. 2012. *Ceratotherium simum*. The IUCN Red List of Threatened Species 2012: e.t4185a16980466. http://dxdoi.org/10.2305/IUCN.UK.2012.RLTS.T4185A16980466.en.

Emslie RH, Brooks M. 1999. *African rhino: status survey and conservation action plan*. IUCN/SSC African Rhino Specialist Group. IUCN, Gland, Switzerland and Cambridge, UK.

Emslie RH, Knight MH. 2014. Update on African rhino status and trends: from IUCN SSC African Rhino Specialist Group (AfRSG). *Report to CITES Standing Committee 65th meeting*: pp. 1–6. wwwrhinoresourcecenter.com.

Fiddes IT, Armstrong J, Diekhans M, Nachtweide S, Kronenberg ZN, Underwood JG, Gordon D, Earl D, Keane T, Eichler EE, et al. 2018. Comparative Annotation Toolkit (CAT) - simultaneous clade and personal genome annotation. bioRxiv doi: 10.1101/231118.

Folch J, Cocero M, Chesné P, Alabart J, Domínguez V, Cognié Y, Roche A, Fernández-Arias A, Martí J, Sánchez P, et al. 2009. First birth of an animal from an extinct subspecies (*Capra pyrenaica pyrenaica*) by cloning. *Theriogenology* **71:** 1026–1034.

Frankham R, Ballou JD, Briscoe DA. 2010. *Introduction to conservation genetics*. Cambridge University Press, Cambridge, UK.

Gaunt TR, Rodríguez S, Day IN. 2007. Cubic exact solutions for the estimation of pairwise haplotype frequencies: implications for link-age disequilibrium analyses and a web tool 'cubex'. *BMC Bioinformatics* **8:** 428.

Groves CP, Fernando P, Robovský J. 2010. The sixth rhino: a taxonomic reassessment of the critically endangered northern white rhinoceros. *PLoS One* **5:** e9703.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* **5:** 1–11.

Harley EH, de Waal M, Murray S, O'Ryan C. 2016. Comparison of whole mitochondrial genome sequences of northern and southern white rhinoceroses (*Ceratotherium simum*): the conservation consequences of species definitions. *Conserv Genet* **17:** 1285–1291.

Hayashi K, Saitou M. 2013. Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells. *Nat Protoc* **8:** 1513–1524.

Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M. 2011. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146:** 519–532.

Hendriks S, Dancet EA, van Pelt AM, Hamer G, Repping S. 2015. Artificial gametes: a systematic review of biological progress towards clinical application. *Hum Reprod Update* **21:** 285–296.

Hillman-Smith AKK, Owen-Smith N, Anderson JL, Hall-Martin AJ, Selaladi JP. 1986. Age estimation of the white rhinoceros (*Ceratotherium simum*). *J Zool* **210:** 355–377.

Houck ML, Ryder OA, Vhala J, Kock RA, Oosterhuis JE. 1994. Diploid chromosome number and chromosomal variation in the white rhinoceros (*Ceratotherium simum*). *J Hered* **85:** 30–34.

Houck M, Ryder O, Kumamoto A, Benirschke K. 1995. Cytogenetics of the rhinocerotidae. *Verh ber Erkrg Zootiere* **37:** 25–32.

Kardos M, Åkesson M, Fountain T, Flagstad O, Liberg O, Olason P, Sand H, Wabakken P, Wikenros C, Ellegren H, et al. 2018. Genomic consequences of intensive inbreeding in an isolated wolf population. *Nat Ecol Evol* **2:** 124–131.

Kolbert E. 2014. *The sixth extinction: an unnatural history*. Henry Holt and Co., New York.

Lanza RP, Cibelli JB, Diaz F, Moraes CT, Farin PW, Farin CE, Hammer CJ, West MD, Damiani P. 2000. Cloning of an endangered species (*Bos gaurus*) using interspecies nuclear transfer. *Cloning* **2:** 79–90.

Lee T-H, Guo H, Wang X, Kim C, Paterson AH. 2014. Snphylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* **15:** 162.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* **475:** 493–496.

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

Lorenzen E, Heller R, Siegismund HR. 2012. Comparative phylogeography of African savannah ungulates. *Mol Ecol* **21:** 3656–3670.

Messer PW, Petrov DA. 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol Evol* **28:** 659–669.

Mi H, Muruganujan A, Casagrande JT, Thomas PD. 2013. Large-scale gene function analysis with the panther classification system. *Nat Protoc* **8:** 1551–1566.

Morris K, Austin JJ, Belov K. 2013. Low major histocompatibility complex diversity in the Tasmanian devil predates European settlement and may explain susceptibility to disease epidemics. *Biol Lett* **9:** 20120900.

Moyle LC. 2005. Assessing the origin of species in the genomic era. *Genome Biol* **6:** 217.

Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156:** 297–304.

Nayernia K, Lee JH, Drusenheimer N, Nolte J, Wulf G, Dressel R, Gromoll J, Engel W. 2006. Derivation of male germ cells from bone marrow stem cells. *Lab Invest* **86:** 654–663.

Neubauer K, Jewgenow K, Blottner S, Wildt DE, Pukazhenthi BS. 2004. Quantity rather than quality in teratospermic males: a histomorphometric and flow cytometric evaluation of spermatogenesis in the domestic cat (*Felis catus*). *Biol Reprod* **71:** 1517–1524.

Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, et al. 2013. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499:** 74–78.

Panula S, Medrano JV, Kee K, Bergström R, Nguyen HN, Byers B, Wilson KD, Wu JC, Simon C, Hovatta O, et al. 2011. Human germ cell differentiation from fetal- and adult-derived induced pluripotent stem cells. *Hum Mol Genet* **20:** 752–762.

Paten B, Earl D, Nguyen N, Diekhans M, Zerbino D, Haussler D. 2011. Cactus: algorithms for genome multiple sequence alignment. *Genome Res* **21:** 1512–1528.

Paudel Y, Madsen O, Megens H-J, Frantz LA, Bosse M, Crooijmans RP, Groenen MA. 2015. Copy number variation in the speciation of pigs: a possible prominent role for olfactory receptors. *BMC Genomics* **16:** 330.

Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, et al. 2013. Great ape genetic diversity and population history. *Nature* **499:** 471–475.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38:** 904–909.

R Core Team. 2016. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://wwwR-project.org/.

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

Ruiz-López MJ, Ganan N, Godoy JA, Del Olmo A, Garde J, Espeso G, Vargas A, Martinez F, Roldán ER, Gomendio M, et al. 2012. Heterozygosity-fitness correlations and inbreeding depression in two critically endangered mammals. *Conserv Biol* **26:** 1121–1129.

Saragusty J, Diecke S, Drukker M, Durrant B, Friedrich Ben-Nun I, Galli C, Göritz F, Hayashi K, Hermes R, Holtze S, et al. 2016. Rewinding the process of mammalian extinction. *Zoo Biol* **35:** 280–292.

Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7:** 62.

Steiner CC, Putnam AS, Hoeck PE, Ryder OA. 2013. Conservation genomics of threatened animal species. *Annu Rev Anim Biosci* **1:** 261–281.

Wake DB, Vredenburg VT. 2008. Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proc Natl Acad Sci* **105** (Supplement 1): 11466–11473.

Wisely SM, Ryder OA, Santymire RM, Engelhardt JF, Novak BJ. 2015. A road map for 21st century genetic restoration: gene pool enrichment of the black-footed ferret. *J Hered* **106:** 581–592.

Xue Y, Prado-Martinez J, Sudmant PH, Narasimhan V, Ayub Q, Szpak M, Frandsen P, Chen Y, Yngvadottir B, Cooper DN, et al. 2015. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348:** 242–245.

Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L, et al. 2016. Ensembl 2016. *Nucleic Acids Res* **44:** D710.

Zhao S, Zheng P, Dong S, Zhan X, Wu Q, Guo X, Hu Y, He W, Zhang S, Fan W, et al. 2013. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat Genet* **45:** 67–71.

Zhou J, Teo Y-Y. 2015. Estimating time to the most recent common ancestor (TMRCA): comparison and application of eight methods. *Eur J Hum Genet* **24:** 1195.

Zhu Y, Hu H-L, Li P, Yang S, Zhang W, Ding H, Tian R-H, Ning Y, Zhang L-L, Guo X-Z, et al. 2012. Generation of male germ cells from induced pluripotent stem cells (iPS cells): an *in vitro* and *in vivo* study. *Asian J Androl* **14:** 574.