






Genetic load and viability of a future restored northern white rhino population

Aryn P. Wilder¹  | Cynthia C. Steiner¹  | Sarah Hendricks^{1,2}  | Benjamin C. Haller³  | Chang Kim^{4,5} | Marisa L. Korody¹ | Oliver A. Ryder¹ 

¹Conservation Genetics, San Diego Zoo Wildlife Alliance, Escondido, California, USA

²Institute for Interdisciplinary Data Sciences, University of Idaho, Moscow, Idaho, USA

³Department of Computational Biology, Cornell University, Ithaca, New York, USA

⁴University of California, Santa Cruz Genomics Institute, Santa Cruz, California, USA

⁵Department of Neurological Surgery, University of California, San Francisco, California, USA

Correspondence

Aryn P. Wilder, Conservation Genetics, San Diego Zoo Wildlife Alliance, Escondido, California 92027, USA.
Email: awilder@sdzwa.org

Abstract

As biodiversity loss outpaces recovery, conservationists are increasingly turning to novel tools for preventing extinction, including cloning and in vitro gametogenesis of biobanked cells. However, restoration of populations can be hindered by low genetic diversity and deleterious genetic load. The persistence of the northern white rhino (*Ceratotherium simum cottoni*) now depends on the cryopreserved cells of 12 individuals. These banked genomes have higher genetic diversity than southern white rhinos (*C. s. simum*), a sister subspecies that successfully recovered from a severe bottleneck, but the potential impact of genetic load is unknown. We estimated how demographic history has shaped genome-wide genetic load in nine northern and 13 southern white rhinos. The bottleneck left southern white rhinos with more fixed and homozygous deleterious alleles and longer runs of homozygosity, whereas northern white rhinos retained more deleterious alleles masked in heterozygosity. To gauge the impact of genetic load on the fitness of a northern white rhino population restored from biobanked cells, we simulated recovery using fitness of southern white rhinos as a benchmark for a viable population. Unlike traditional restoration, cell-derived founders can be reintroduced in subsequent generations to boost lost genetic diversity and relieve inbreeding. In simulations with repeated reintroduction of founders into a restored population, the fitness cost of genetic load remained lower than that borne by southern white rhinos. Without reintroductions, rapid growth of the restored population (>20–30% per generation) would be needed to maintain comparable fitness. Our results suggest that inbreeding depression from genetic load is not necessarily a barrier to recovery of the northern white rhino and demonstrate how restoration from biobanked cells relieves some constraints of conventional restoration from a limited founder pool. Established conservation methods that protect healthy populations will remain paramount, but emerging technologies hold promise to bolster these tools to combat the extinction crisis.

KEYWORDS

fitness, genetic load, genetic restoration, in vitro conservation, runs of homozygosity, simulation

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd.



1 | INTRODUCTION

Human activities have led to a worldwide loss of biodiversity (Cardinale et al., 2012). An estimated 22% of mammal populations are at risk of extinction (<http://www.iucnredlist.org>). With the alarming pace of global biodiversity loss, conservationists must increasingly turn to nontraditional and last-ditch methods for saving species, including genetic restoration of populations from banked gametes or cryopreserved cell lines (Ben-Nun et al., 2011; Hildebrandt et al., 2021; Korody et al., 2021). Cloning, in vitro fertilization, and eventually in vitro gametogenesis from induced pluripotent stem cells (iPSCs) can restore lost genetic diversity into a population or even an extinct species, providing a second chance for recovery (Gross, 2021; Saragusty et al., 2016). But just as the viability of a recovering population hinges on habitat availability and the mitigation of threats (Crees et al., 2016), it may also depend on the genomic health of the population (Kardos et al., 2021, 2023). If cryopreserved individuals represent the genomic diversity of pre-collapse populations, the genomes of banked cells could harbor the same intrinsic constraints (e.g., low genetic diversity, maladaptation, inbreeding depression) that may have contributed to the demise of their original population. Thus, assessment of the genomic health of cryopreserved genomes and their capacity for restoring a viable population can help guide the use of cellular-based conservation techniques.

The fitness and viability of small populations are influenced by the distribution, frequency, and severity of deleterious genetic load (Bertorelle et al., 2022; Kardos et al., 2021). As genetic drift exceeds the strength of purifying selection, deleterious alleles can increase in frequency and become fixed, causing a “mutational meltdown” (Dussex et al., 2023; Lynch et al., 1995). Although smaller populations tend to accumulate mildly to moderately deleterious alleles at higher frequencies, large populations harbor more highly deleterious alleles segregating at low frequencies, termed inbreeding load or masked load (Bertorelle et al., 2022; Hedrick & García-Dorado, 2016). Masked or inbreeding load can make large populations more susceptible to inbreeding depression during severe population bottlenecks, as recessive, highly deleterious alleles are brought together in a homozygous state under inbreeding and genetic drift (Hedrick & García-Dorado, 2016). The quantification of genome-wide genetic load can provide a fitness-relevant assessment of genomic health, susceptibility to inbreeding depression and the capacity for recovery of endangered species (Robinson et al., 2022; Wilder et al., 2023).

The white rhinoceros (*Ceratotherium simum*) consists of two subspecies with very different levels of extinction risk (Emslie, 2020; Emslie & Knight, 2014). The southern white rhinoceros (SWR; *C. s. simum*), distributed across South Africa, is currently the most abundant rhinoceros in the world. Listed as Near Threatened by the International Union for the Conservation of Nature (IUCN), it is one of only two rhinos (along with the greater one-horned rhinoceros, *Rhinoceros unicornis*) not considered Critically Endangered. Once on the brink of extinction, swift conservation efforts enabled recovery of SWR from an estimated 50–200 individuals in the early

20th century (Figure 1a; Emslie, 2020; Rookmaaker, 2000). At its recent peak in 2014, the census population reached ~20,000 individuals, but a surge in poaching beginning a decade ago reduced the population to ~18,000 (Emslie, 2020; Emslie & Knight, 2014). The northern white rhinoceros (NWR; *C. s. cottoni*), once distributed over parts of Uganda, Chad, Sudan, the Central African Republic, and the Democratic Republic of Congo, has been driven effectively to extinction by poaching (Figure 1a), with the only remaining live individuals being two related, non-reproductive females.

The two white rhino subspecies have distinct recent population trajectories with a shared evolutionary history. Previous genomic analyses have suggested that the subspecies began diverging ~400–600 KYA, with intermittent periods of gene flow that ceased within the last 100 KYA (Moodley et al., 2020). After divergence, both subspecies had effective population sizes (N_e) in the thousands (Moodley et al., 2020; Tunstall et al., 2018). In SWR, the recent bottleneck was followed by a recovery to roughly half the pre-bottleneck N_e . NWR underwent a more recent bottleneck and has not recovered. Genomes from SWR individuals sampled post-bottleneck and recovery (born mostly mid-20th century; Figure 1a) have lower genetic diversity than NWR individuals born early in the NWR bottleneck (Tunstall et al., 2018). Both subspecies show evidence of genomic erosion: contemporary genomes had lower genetic diversity and higher inbreeding coefficients than those sampled in the late 19th century, before the most severe population declines (Sánchez-Barreiro et al., 2021). Yet genomic erosion has not limited recovery of the SWR population, which has grown more than 100-fold in the past 13 decades. NWR, by contrast, faces imminent extinction.

Cryopreserved cell lines from 12 individuals, eight of which are unrelated, offer hope for reviving a viable NWR population. NWR is the focus of research programs aimed at generating viable embryos using iPSC-derived gametes, assisted reproductive technologies (ART), and somatic cell nuclear transfer (SCNT) (Hayashi et al., 2022; Hildebrandt et al., 2021; Korody et al., 2021; Saragusty et al., 2016). Using ART with intracytoplasmic sperm injection of cryopreserved semen into oocytes, NWR embryos have already been created and cryopreserved for future transfer into a surrogate mother (Hildebrandt et al., 2018, 2021, 2023), demonstrating the feasibility of implementing these approaches in the very near future to begin restoring the NWR population.

The extent to which a future NWR population will be constrained by the limited number of potential founders (12 cell lines) is unknown. Small populations are susceptible to extinction from stochastic environmental events and Allee effects, a risk compounded by low fitness from inbreeding depression and loss of standing genetic variation (Gilpin, 1986). The genomes of NWR cell lines retain higher genetic diversity than individuals from the recovered SWR population (Tunstall et al., 2018), suggesting the future NWR founders have adequate adaptive potential. However, historically large populations with high genetic diversity may also harbor high masked load, where the fitness effects of deleterious alleles are masked in heterozygosity (Hedrick & García-Dorado, 2016). Inbreeding is expected to convert masked genetic load into homozygous load (Bertorelle et al., 2022), reducing

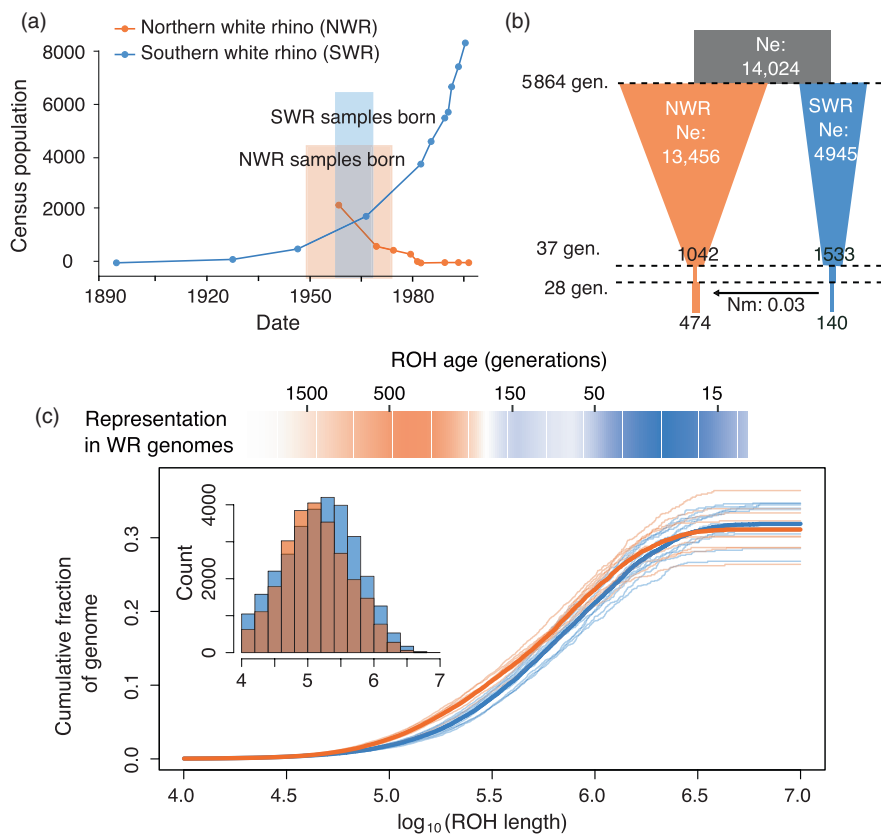


FIGURE 1 (a) Census population sizes of NWR and SWR over the past two centuries. When our study individuals were born, the SWR population was recovering from <200 individuals and the NWR population was still declining. (b) The demographic model that best fit the two-dimensional site frequency spectrum (2dSFS), showing population decline until 37 generations ago, after which both populations underwent more rapid size changes, with some evidence of very low levels of gene flow from SWR into NWR. (c) Cumulative fraction of genomes represented by ROH lengths (thin lines are individuals, thick lines are population means; inset shows raw counts across all 22 genomes). The relative abundance of ROH lengths is shown in the heatmap above, with more orange indicating higher abundance in NWR genomes and more blue indicating higher abundance in SWR. NWR genomes had smaller ROH likely stemming from older inbreeding (>500 generations ago), whereas SWR had larger ROH from recent inbreeding (as recently as ~15 generations ago).

fitness and potentially hindering recovery. In a population established by few founders, some level of inbreeding will likely be unavoidable. However, embryos produced via cloning or in vitro gametogenesis from iPSCs can be repeatedly reintroduced to the population over multiple generations. Unlike a single founding event, the reintroduction of founders could provide a source of genetic rescue to curb inbreeding and realized load (Wisely et al., 2015).

We conducted a comprehensive examination of genetic load to gauge the potential for future inbreeding depression and low fitness that might hinder recovery of a population founded by these cell lines. We first characterized the distributions of genome-wide deleterious variation and their relationships with the demographic histories of the two subspecies. We hypothesized the prolonged bottleneck in SWR led to an accumulation of mildly and moderately deleterious alleles and elevated homozygous genetic load, whereas the shorter bottleneck in NWR allowed the population to retain more highly deleterious variants masked in heterozygosity (Hedrick & García-Dorado, 2016). If confirmed, this high inbreeding load may reduce fitness as a restored NWR population becomes increasingly inbred and exposed to drift. Using SWR genomes as a benchmark for viable recovery potential, we

simulated genetic load and its fitness costs in NWR under two scenarios: one with a single founding event and the other with repeated reintroductions of founder genomes in subsequent generations. The comparison of independent populations with similar life-history characteristics, shared evolutionary history, and distinct recent demographic trajectories provides a unique framework for isolating the impact of past and future population fluctuations on genetic load, and a means for assessing the potential for genetic restoration and recovery of NWR through in vitro conservation technology.

2 | MATERIALS AND METHODS

We focused our analysis on publicly available genomes from contemporary samples (Sánchez-Barreiro et al., 2021; Tunstall et al., 2018) to provide a detailed and in-depth analysis of genetic load and demographic history. We excluded the historical samples of Sánchez-Barreiro et al. (2021) to avoid limitations from low read depth and DNA damage inherent to these samples. All necessary permits from Convention on International Trade in Endangered

Species (CITES), United States Fish and Wildlife Service (USFWS), United States Department of Agriculture (USDA), domestic and foreign customs, and local permits were acquired for legal possession and use of samples in this project.

2.1 | Read mapping and SNP calling

We analyzed whole genomes of nine NWR and 13 SWR individuals (Table S1). Eight of the nine NWR individuals are represented by fibroblast cell lines in San Diego Zoo Wildlife Alliance's Frozen Zoo and six of these have been successfully reprogrammed into iPSCs (Korody et al., 2021). We downloaded whole-genome sequence reads from these nine NWR born between 1952 and 1972, and four SWR born between 1958 and 1963, generated by Tunstall et al. (2018). Nine additional SWR samples were generated by Sánchez-Barreiro et al. (2021), two from wild individuals born 1968–1969, six captive individuals born 2008–2011, and one wild individual born in 2012.

TruSeq3 adapters were trimmed from paired-end reads using Trimmomatic v.0.3.2 (Bolger et al., 2014) under the following settings: seed mismatches=2, palindrome clip threshold=30, simple clip threshold=10, minimum adapter length=4, and keeping both reads. Trimmed reads were mapped to the SWR reference genome CerSimSim1.0 using bowtie2 under the very sensitive parameter settings (Langmead & Salzberg, 2012), with a maximum fragment length of 2 KB. Duplicate reads were removed using MarkDuplicates in Picard (Van der Auwera & O'Connor, 2020). Sequence read depth was higher in the more recently generated samples from Sánchez-Barreiro et al. (2021; 21.65×, range: 18.46–25.07×) than those from Tunstall et al. (2018; 14.32×, range: 11.01–15.54×; Table S1). Genotype calls from these moderate depths of coverage are expected to have high overall concordance (>98%) with true genotype calls, but are known to underestimate heterozygous genotypes from lower coverage samples (Kishikawa et al., 2019). To minimize biases in heterozygous call rates stemming from the different read depths, we randomly subset genome-wide reads from the higher coverage samples to 14.3× coverage (the average read depth across the Tunstall et al. (2018) samples) using Samtools (Danecek et al., 2021).

Variants were called in GATK v3.8 (Van der Auwera & O'Connor, 2020), first in all samples individually using HaplotypeCaller in gVCF mode, then across all samples combined using GenotypeGVCFs. We examined the distributions of genome-wide SNPs and indels to select thresholds for filtering variants and retained variants that passed the following criteria, for SNPs: $QD > 5$, $FS < 32$, $SOR < 3$, $MQRankSum > -2$, $MQ > 28$, $ReadPosRankSum > -2$, and for indels: $QD > 5$, $FS < 32$, $SOR < 3$, $ReadPosRankSum > -2$. After applying filters, we retained 7,653,428 variants.

2.2 | Polarizing ancestral and derived alleles

The ancestral and derived alleles were polarized at each variant using genomes from two black rhino subspecies, the eastern black

rhino (*Diceros bicornis michaelis*) and southern black rhino (*D. bicornis minor*). Reads used for the assembly of the eastern and southern black rhino reference genomes (SRR11430189 and SRR12010331, respectively) were downloaded and mapped to the SWR reference genome using the same methods described above. We then separately called genotypes in the two black rhino genomes at all positions that were polymorphic in white rhinos. We used GATK HaplotypeCaller to generate single sample gVCFs, followed by GenotypeGVCFs. We included sites that were non-variant in the black rhino genomes as well as variant ones. Raw variants were then filtered using the same methods as for white rhinos, by plotting the distributions of genome-wide SNPs and indels to select hard filter thresholds. For both black rhino genomes we retained genotypes that met the following criteria: $QD > 22$, $FS < 32$, $MQ > 32$, $MQRankSum > -5$, $ReadPosRankSum > -2$, $SOR < 3$. We defined the ancestral allele as the allele that was homozygous in both black rhino genomes, ignoring sites with more than one allele in the two black rhino genomes (i.e., heterozygous sites or sites with different alleles in the two black rhino genomes). We were able to polarize 6,267,082 variants with ancestral and derived alleles.

2.3 | Estimating heterozygosity and runs of homozygosity

We estimated runs of homozygosity (ROH) in white rhino genomes from the filtered variant set using the ROH function in bcftools v.1.8, which applies a hidden Markov model to biallelic sites to identify genomic stretches of the genome that show identical alleles on both chromosomes (Narasimhan et al., 2016). To estimate inbreeding coefficients (F_{ROH}), we summed the lengths of ROH >1 MB (with bcftools ROH quality scores >30) across the genome and divided by the genome length. We estimated heterozygosity in PLINK v1.9 (--het function) from a subset of high-quality variants, filtering out sites with >30% missing data after removing genotypes with $GQ < 20$ to minimize biases stemming from differences in sequencing batches (Purcell et al., 2007). We then divided the number of heterozygous sites per sample by the gapless length of the reference genome, adjusted for the proportion of called genotypes per sample.

2.4 | Demographic modeling

Previous analyses of the demographic history of white rhinos either fit models to a folded (i.e., non-polarized) two-dimensional site frequency spectrum (2dSFS; Tunstall et al., 2018), or used coalescence-based analysis of single genomes (Moodley et al., 2020). We built on these previous analyses, testing the fit of more complex demographic models to an unfolded 2dSFS using polarized ancestral and derived alleles. We tested demographic models with the GADMA program, which uses a genetic algorithm in an unsupervised, heuristic global search of demographic models (Noskova et al., 2020). GADMA implements the moments program

(Jouganous et al., 2017), which simulates the 2dSFS under different demographic scenarios and uses an optimization process to fit parameters to the data. We tested models with four time periods, where an ancestral population splits into two populations at time T1, followed by three time periods during which population sizes, growth dynamics (including instantaneous, exponential, and linear growth), and migration rates can change. The 2dSFS was generated from derived allele counts for each subspecies from the final set of SNPs with the script `vcf2sfs.py` (<https://github.com/marqueda/SFS-scripts>). To allow for model comparison, we block bootstrapped the data over 4 MB blocks using the `vcf2sfs.py` script, generating 100 bootstrapped 2dSFS. The SFS was projected to 16 NWR alleles and 24 SWR alleles to account for missing genotypes. We performed five independent runs of GADMA on block bootstrapped 2dSFS under default settings and selected the model that minimized the composite likelihood Akaike's information criterion (CLAIC) across all runs. We then ran local searches on block bootstrapped data to generate confidence intervals for parameters of the best model in the GADMA program. We rescaled the parameter estimates to demographic units assuming a mutation rate of 2.1×10^{-8} (Moodley et al., 2020).

To infer the timing of inbreeding events that generated ROH in the genomes of the two subspecies, we examined the distributions of ROH sizes. To estimate the relative representation of ROH lengths in each subspecies, we plotted the cumulative sum of lengths of ROH in each genome and averaged the curves across individuals of each subspecies (Figure 1c). We then estimated the mean slopes of the curves within ROH intervals, using 1 MB sliding windows with 10 KB step size, comparing the slopes of the two subspecies (van der Valk et al., 2019). A steeper slope in one subspecies indicates greater representation of the given ROH length within that subspecies. Assuming 50 MB identity-by-descent tracts are inherited per generation, we can infer the time to a common ancestor of a tract of a given length (Thompson, 2013); thus, to estimate the number of generations since the inbreeding event that results in ROH of a given size, we divided 50 MB by the ROH length.

2.5 | Genome annotation

We used the Comparative Annotation Toolkit (CAT; Fiddes et al., 2018) to annotate the rhinoceros alignment from progressive Cactus (Armstrong et al., 2020), using the NCBI horse (GCF_002863925.1) RefSeq annotation release 103 as reference. For input into the Augustus step in CAT, RNAseq data were generated from several archived tissues, including brain of a NWR female, testis of a NWR male, reproductive tissues of a SWR female, and fibroblast cell cultures from a NWR male. Libraries were prepared using the Illumina TruSeq stranded mRNA library prep kit and sequenced using paired-end 76 bp reads on an Illumina MiSeq. Reads that mapped to the genome (29.9 M reads totaling 2.3 Gb) using TopHat2 (Kim et al., 2013) were passed directly to the CAT

pipeline. All steps used for the final consensus finding and transcript classification in the CAT pipeline were run except the Augustus PB step, and the default parameters were used in Augustus. We used BUSCO v4.0.2 to evaluate the annotation coverage (Figures S1–S6). Overall, the reference annotation had high coverage of protein coding genes (Figure S1; Laurasiatheria complete BUSCO: 98.1%), with good assembly contiguity in genic regions (Figures S2, S3) and support of introns, exons, and splice site boundaries (Figures S4–S6). Because the indel error rate was high in the annotation, a large number of genes failed to map using transMap and were instead included as AugustusCGP transcripts. We omitted all AugustusCGP gene predictions with 0–2 introns in the final annotation.

2.6 | Estimating variant impacts and genetic load

To estimate genetic load, we used two approaches, inferring variant impacts based on (1) their effect on encoded proteins of protein coding genes, and (2) evolutionary conservation across mammalian taxa, with the assumption that mutations at sites that are more conserved across taxa are more likely to be deleterious. We examined 6.26 M variants that had ancestral/derived polarization of alleles and genotype calls in all 22 samples. We used default settings in the program `snpEff` (Cingolani, 2022) to predict the effects of variants in protein coding genes with our updated gene annotation. We estimated homozygous and heterozygous deleterious variants across all genes that are well conserved in mammals (i.e., genes in the BUSCO `mammalia_odb10` set). To estimate evolutionary conservation, we downloaded the GERP++ track scores generated from a 34-mammal alignment (<http://mendel.stanford.edu/SidowLab/downloads/gerp/index.html>; which does not include the white rhino genome). In this dataset, constraint intensity at each alignment position is quantified by a rejected substitutions (RS) score, which is the number of substitutions expected under neutrality minus the number of substitutions observed at the position. We defined evolutionarily conserved sites as those with $RS > 4$. We used the `liftOver` function in the R package `rtracklayer` to lift over coordinates of variant sites from the white rhino genome to the human genome (hg19), then from hg19 to the earlier assembly of the human genome (hg18) used in the 34 mammalian species alignment to assign RS scores to each variant position in the white rhino genome. To add a second metric of evolutionary conservation, we also downloaded phyloP scores from the UCSC Genome Browser computed from the 34 mammalian species alignment and similarly used `liftOver` to convert coordinates of variable sites in the white rhino genome to the hg19 reference. Results from GERP++ and phyloP were highly correlated, and thus we report only results from GERP++. We defined deleterious variants as those at evolutionarily conserved sites ($RS > 4$), and mutations at evolutionarily conserved sites predicted to have moderate impact on protein coding genes (e.g., missense mutations) by `snpEff`. We considered homozygous deleterious variants to be “homozygous load” and heterozygous deleterious variants to be “heterozygous load.”



Because long ROH from recent inbreeding may harbor high levels of deleterious variants (Szpiech et al., 2013), we examined putatively deleterious alleles in ROH. Within length bins ($0.1 \text{ MB} \leq \text{ROH} < 0.5 \text{ MB}$, $0.5 \text{ MB} \leq \text{ROH} < 1 \text{ MB}$, $1 \text{ MB} \leq \text{ROH} < 3 \text{ MB}$, and $\text{ROH} \geq 3 \text{ MB}$), we quantified the proportion of homozygous mutations at conserved sites within ROH, and the ratio of moderate-impact (missense) homozygous alleles at conserved ($\text{RS} > 4$) sites relative to low-impact homozygous alleles. We tested the relationship between ROH length and the proportion of deleterious alleles using linear regressions. We compared the proportion of homozygous sites with deleterious alleles within ROH relative to the proportion of homozygous sites with deleterious alleles in autozygous regions using paired *t*-tests. We also compared the proportion of homozygous sites with deleterious alleles within ROH relative to the proportion of heterozygous sites with deleterious alleles in autozygous regions using paired *t*-tests.

2.7 | Functional enrichment tests

We annotated the predicted coding sequences (CDS) from gene annotations in the white rhino genome with gene ontology (GO) terms from several public databases. First, we used TransDecoder v5.5.0 (<https://github.com/TransDecoder/TransDecoder>) to predict protein-coding regions in the white rhino transcripts based on the longest ORF. We searched for homologous sequences for both transcripts and predicted proteins in the SwissProt database (<ftp://ftp.ebi.ac.uk/pub/databases/uniprot/>) using BLASTX and BLASTP (NCBI-BLAST-2.9.0+), respectively (Altschul et al., 1990, 1997), retaining only top hits with an *e*-value $\leq 1 \times 10^{-5}$. We identified protein domains in white rhino transcripts by comparing to the Pfam-A database (<ftp://ftp.ebi.ac.uk/pub/databases/Pfam/>) using HMMER-3.2.1 (Finn et al., 2011). And finally, using homologous proteins in TRINOTATE-v3 (Bryant et al., 2017), we combined functional annotations from the following databases: (1) GO (Ashburner et al., 2000), (2) EggNOG (Powell et al., 2012) and (3) KEGG (Kanehisa et al., 2012). TransDecoder predicted 13,419 protein-coding genes within the 196,815 CDS transcripts. 12,658 protein-coding genes and 154,102 transcripts mapped to proteins in the SwissProt database, and 7508 were present in the Pfam-A database. Using the combined results of homology searches via SwissProt and Pfam, we assigned 22,749 unique GO terms across 17,662 transcripts.

We tested for enrichment of Biological Processes GO terms among genes with deleterious variation in the program ErmineJ v3.1.1 (Gillis et al., 2010). We used an over-representation analysis (ORA) to test for GO enrichment among genes with at least one deleterious mutation relative to GO terms from all genes. We also used the gene score resampling (GSR) method, where each gene was scored by the number of deleterious variants divided by the length of coding sequences for the gene, and GO terms are ranked by the scores of genes within the associated gene set. We restricted the

analysis to terms with 5–100 genes and used a mean class scoring method with 10,000 replicates.

2.8 | Forward genomic simulations

To explore how future inbreeding might impact a restored NWR population, we simulated genetic load and its fitness cost in a population founded by banked NWR cell lines using forward genomic simulations in the program SLiM v4.0.1 (Haller & Messer, 2023). We modeled neutral and deleterious mutations under different population growth and restoration scenarios, using SWR as a reference for a viable level of realized genetic load. We modeled only neutral and deleterious mutations to avoid assumptions about the distribution of positive fitness effects and the absolute fitness of individuals. Deleterious mutations at evolutionarily conserved sites are likely to impact vital organismal functions in all environments (Christmas et al., 2023), whereas beneficial mutations may be more context and environment-dependent, and their impacts are more likely to differ between taxa or in populations under human care. We started the simulations from eight NWR genomes with cryopreserved cell lines that can be used for genetic rescue. These individuals represent eight of the 12 individuals with cryopreserved cell lines. The four individuals not represented in our sequenced genomes were F1 or F2 offspring of three F0 individuals included in our dataset and thus would contribute little new diversity to the restored population. One sequenced genome in our dataset does not have cryopreserved cell lines and was excluded from the simulations.

We simulated two different types of genetic rescue strategies, one in which a restored population is founded by the eight cryopreserved cell lines in generation 0, and a second in which the restored population is supplemented with one randomly sampled cell line per generation. For both strategies, we simulated 10 non-overlapping generations of non-Wright-Fisher evolution under six population growth scenarios: 0%, 10%, 20%, 30%, 40%, and 50% growth per generation. We ran 50 replicates per scenario. Because we simulated only deleterious genetic variation across the genome, fitness values represented the cost of realized load in the genomes of individuals rather than their absolute fitness. The probability that an individual successfully reproduced was proportional to its fitness relative to others in the population. Specifically, each generation of parents was sampled with replacement weighted by their fitness, such that an individual with fitness 10% above the mean would have 10% higher than average probability of producing offspring. Population size and growth were predefined parameters rather than being emergent properties of the absolute fitness of individuals in the simulated population.

The simulations began in generation 0 with empirical data on deleterious variation in the genomes of the eight cryopreserved cell lines. We simulated all scaffolds in the genome $>1 \text{ MB}$ in length and assumed a per-base recombination rate of 1×10^{-8} (Dumont & Payseur, 2008) and independent assortment of scaffolds. We

reconstructed founder haplotypes from the observed genotypes of each individual at all mildly to severely deleterious variants ($RS \geq 2$), plus 10,000 randomly selected neutral variants ($RS < 2$). For heterozygous variants, the two alleles were randomly assigned to the two haplotypes. Variants were binned into four groups reflecting their likely severity: "neutral" ($-2 < RS < 2$), "mild" ($2 \leq RS < 4$), "moderate" ($4 \leq RS < 5.8$), and "severe" ($RS \geq 5.8$) and assigned selection and dominance coefficients to each bin based on estimates from the human genome (Henn et al., 2016); selection coefficients: $s_{\text{neutral}} = 0$, $s_{\text{mild}} = -0.0001$, $s_{\text{moderate}} = -0.001$, and $s_{\text{severe}} = -0.002$; dominance coefficients: $h_{\text{neutral}} = 0.5$, $h_{\text{mild}} = 0.293$, $h_{\text{moderate}} = 0.062$, and $h_{\text{severe}} = 0.033$. To assess the sensitivity of the simulations to the assigned selection coefficients, we also tested the simulations with the following selection coefficients: $s_{\text{neutral}} = 0$, $s_{\text{mild}} = -0.024$, $s_{\text{moderate}} = -0.032$, and $s_{\text{severe}} = -0.072$ (Peischl et al., 2018), but found that the two sets of selection coefficients produced broadly consistent results (Figure S7) and thus we report results only from the former set. Because we were mainly interested in the short-term fitness consequences of standing variation and not the long-term accumulation of genetic load, we set the mutation rate to 0. Simulated genomes of the entire population were output each generation. Fitness cost was estimated as the product of the fitness effects of all deleterious variants in the genomes, $(1 + s)$ for homozygous variants, and $(1 + h \times s)$ for heterozygous variants:

$$\prod_{i=1}^{L(\text{hom})} (1 + s_i) \prod_{j=1}^{L(\text{het})} (1 + h_j s_j)$$

Realized load was estimated by summing the selection coefficients of all homozygous derived mutations and summing the products of the dominance and selection coefficients across all heterozygous loci:

$$\sum_{i=1}^{L(\text{hom})} s_i + \sum_{j=1}^{L(\text{het})} h_j s_j$$

We estimated fitness and realized load of the 13 sequenced SWR genomes in the same way, using SWR as a benchmark for fitness in a related population that has successfully recovered from an extreme bottleneck.

One notable difference between our study and others is that our realized load estimates for the NWR and SWR genomes are many times higher than other studies because we included all putatively deleterious variants across the entire genome rather than restricting the analysis to the exome (e.g., Henn et al., 2016) or to ultraconserved regions (Speak et al., 2023). Estimating realized load from only the exome produces values comparable to those seen in studies of the exome or ultraconserved elements (mean realized load = 0.75 for NWR and 0.87 for SWR). Because fitness of individuals was modeled relative to the mean of the population, the absolute fitness values did not affect the outcomes of the models.

3 | RESULTS

3.1 | Demographic history and genomic variation

The model that best fit the observed 2dSFS for the two subspecies (Figure S8) was a scenario in which two populations split from an ancestral population with two recent periods of size changes in both populations (Figure 1b; Table S2). Under this model, an ancestral population of size $N_{\text{anc}} = 14,024$ split into two isolated populations of size $NWR1 = 13,456$ and $SWR1 = 4945$ at time $T1 = 5864$ generations ago. Population sizes decreased linearly until 37 generations ago to $NWR1.1 = 1042$ and $SWR1.1 = 1533$, after which both population sizes decreased instantaneously to $NWR2 = 192$ and $SWR2 = 349$. Over the next 28 generations, populations remained small ($NWR3 = 474$ and $SWR3 = 140$) with a very low level of asymmetrical gene flow ($N_m = 0.031$) from SWR into NWR. Assuming an eight-year generation time (Tunstall et al., 2018), the populations began diverging ~47,000 years ago, with more recent size changes beginning 296 years ago. Using 25.5 years per generation (Moodley et al., 2020), the populations began diverging ~150,000 years ago, with more recent size changes beginning 944 years ago.

Pairwise nucleotide diversity was greater in the NWR population ($\pi = 1.12 \times 10^{-3}$) than SWR ($\pi = 7.17 \times 10^{-4}$), and observed heterozygosity was higher in all NWR individuals (mean = 1.06×10^{-3}) compared to SWR individuals (mean = 7.41×10^{-4} ; $p = 2.27 \times 10^{-13}$; Figure S9), consistent with the larger historical N_e and with the findings of previous studies (Sánchez-Barreiro et al., 2021; Tunstall et al., 2018). The NWR population had more private alleles than the SWR population, both by count (2,791,323 in NWR versus 1,866,849 in SWR) and as a proportion of polymorphic sites (0.63 in NWR versus 0.54 in SWR). The difference may be explained in part by asymmetrical gene flow favoring migration of alleles from SWR into NWR (Figure 1b), but a longer population bottleneck likely also led to the loss of low-frequency, derived alleles from the SWR population.

There were more long ROH and $F_{\text{ROH} > 1\text{MB}}$ was higher in SWR (Figure 1c), reflecting inbreeding in the more recent past in SWR compared to NWR. To infer the timing of inbreeding events that contribute to ROH, we examined the distributions of ROH sizes in the two subspecies (van der Valk et al., 2019). SWR had greater representation of longer ROH (~2–3 MB). Assuming 50 MB identity-by-descent tracts are inherited per generation (Thompson, 2013), the more abundant 2–3 MB ROH in the SWR genome stemmed from inbreeding ~15 generations ago. By contrast, NWR had a greater representation of shorter ROH of ~100–200 KB. ROH of this size likely arose >190 generations ago (Figure 1c), potentially resulting from climate-associated population declines during the Younger Dryas (Moodley et al., 2018). Given variation in recombination rates across the genome and across species, there is likely much variation around these estimates, but the results generally suggest that ROH in the NWR genome reflect historical demographic events in the more distant past, whereas large ROH in the SWR genome are much younger, most likely arising during the recent bottleneck.

3.2 | Deleterious genetic load

Two approaches to estimate genetic load produced largely consistent results, pointing to higher homozygous and fixed load in SWR and higher heterozygous and segregating load in NWR. We first estimated load based on changes at evolutionarily conserved sites in the genome ($RS > 4$) assumed to have moderately to severely deleterious impacts on fitness. As expected for alleles under purifying selection, moderately to severely deleterious alleles ($RS > 4$) tended to segregate at lower frequency in both populations. In SWR, more deleterious alleles were fixed, both in total and as a proportion of all fixed derived alleles, compared to NWR ($\chi^2 = 4.99$, $p = 0.026$), suggesting stronger drift over the long term in SWR than in NWR (Figure 2a). By count, there were more deleterious variants segregating in the NWR population (40,034 sites with at least one deleterious allele in NWR compared to 31,821 in SWR). In individual NWR genomes, there were more deleterious alleles than in SWR genomes ($p = 0.034$), but more of these alleles were in a heterozygous state (Figure 2b). In SWR genomes, a larger proportion of homozygous sites had deleterious alleles compared to NWR (Figure 2c; $p = 2.82 \times 10^{-10}$), and sites with homozygous derived alleles had higher RS scores on average in SWR compared to NWR ($p = 1.63 \times 10^{-5}$). The proportion of heterozygous deleterious alleles was slightly but not significantly higher in SWR compared to NWR (Figure 2c; $p = 0.051$).

The results were largely similar when we examined variation by its predicted impact in protein coding genes. Here we defined deleterious alleles as derived alleles with moderate-impact (e.g., missense) or high-impact/loss of function (e.g., premature stop codons, frameshift mutations) effects on the protein that were also evolutionarily conserved ($RS > 4$). There were more moderate-impact,

evolutionarily conserved variants segregating in the NWR population than SWR (1083 vs. 931 variants, respectively). There were also more loss of function (LOF) variants segregating in the NWR population than SWR (114 vs. 74 variants, respectively). Individual SWR genomes carried more moderate-impact homozygous alleles by count ($p < 1 \times 10^{-6}$), and in proportion to homozygous derived alleles in coding sequences ($p < 1 \times 10^{-6}$; Figure 2b). High-impact, loss of function mutations showed the opposite pattern (Figure 2b). SWR genomes had fewer homozygous and heterozygous LOF alleles by count ($p = 4.71 \times 10^{-4}$ and $p = 4.55 \times 10^{-5}$ for homozygous and heterozygous, respectively) and in proportion to all coding variants ($p = 4.71 \times 10^{-4}$ and $p = 4.55 \times 10^{-5}$ for homozygous and heterozygous, respectively) compared to NWR, suggesting that some high-impact alleles may have been lost from the SWR population through drift or purging. Overall, the genetic load analyses suggest that SWR genomes harbor more homozygous, moderate-impact genetic load, whereas NWR genomes have lower realized load but higher masked load.

Because long ROH from recent inbreeding may harbor high levels of deleterious variants (Szpiech et al., 2013), we tested for enrichment of putatively deleterious alleles in ROH. Compared to shorter ROH, longer ROH had proportionally more homozygous deleterious variants in SWR (linear regression, $p = 6.25 \times 10^{-13}$ for the ratio of moderate-impact conserved to low-impact and $p = 2.65 \times 10^{-9}$ for the proportion conserved) and NWR ($p = 3.38 \times 10^{-5}$ for the ratio of moderate-impact conserved to low-impact and $p = 0.017$ for the proportion conserved; Figure 3). ROH > 1 MB had proportionally more homozygous deleterious variants compared to homozygous variants in non-autozygous regions of the genome (paired t-test, $p = 5.60 \times 10^{-5}$ for homozygous conserved and $p = 0.018$ for homozygous moderate-impact conserved mutations). By contrast,

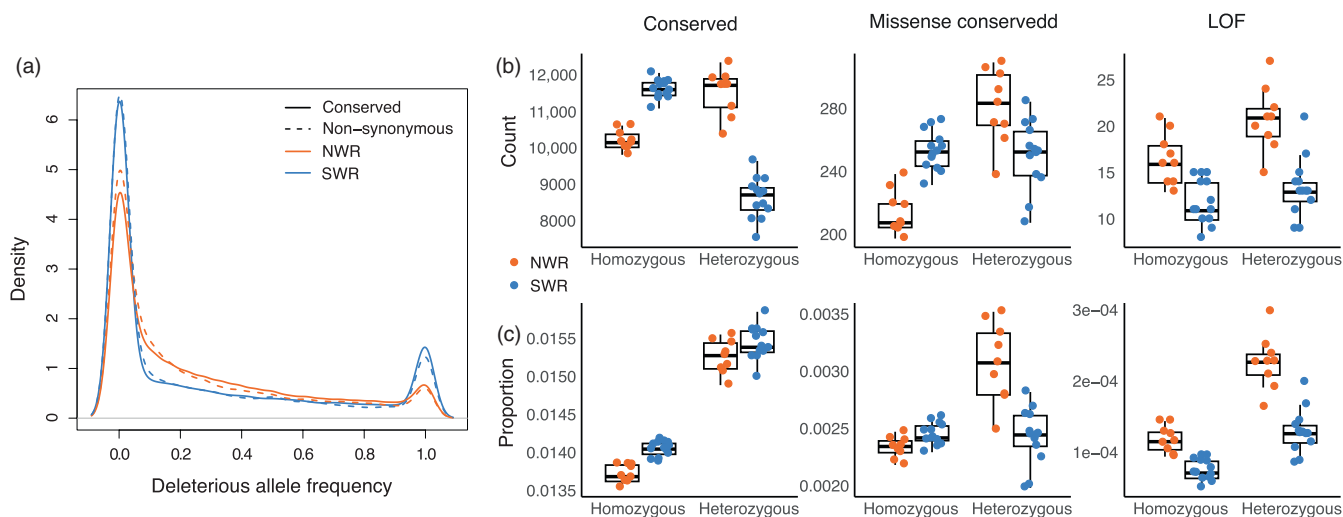


FIGURE 2 (a) Deleterious alleles tend to segregate at lower frequencies in both populations, but proportionally more conserved ($RS > 4$) and non-synonymous (missense and loss of function) deleterious alleles are fixed in SWR ($p = 0.025$) than NWR. (b) Compared to NWR, SWR genomes have more homozygous and fewer heterozygous deleterious alleles at conserved and missense conserved sites, but fewer homozygous and heterozygous loss of function (LOF) alleles by count. (c) SWR genomes have proportionally more homozygous deleterious alleles at conserved and missense conserved sites than SWR, but proportionally fewer homozygous and heterozygous LOF alleles than NWR.

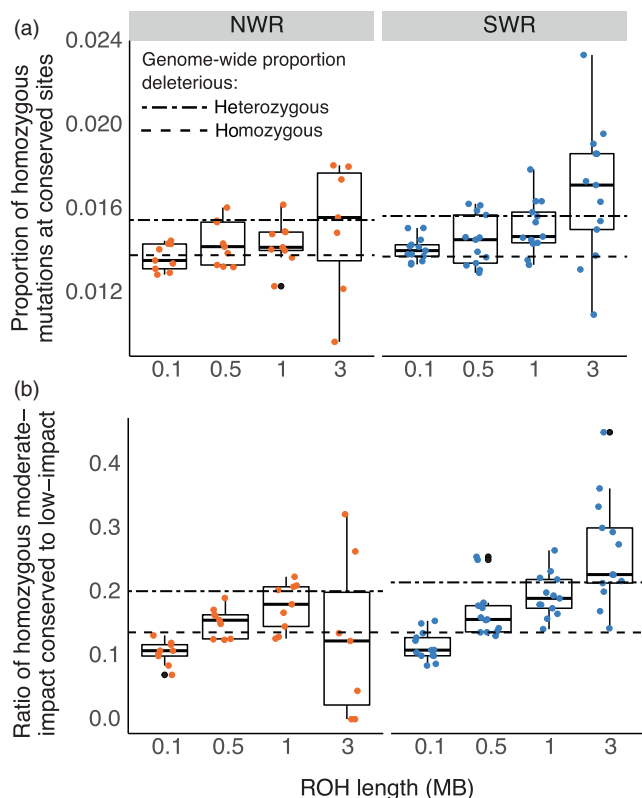


FIGURE 3 Longer ROH are enriched for homozygous deleterious mutations. The proportion of homozygous mutations at conserved sites (a) and the ratio of homozygous moderate-impact conserved ($RS > 4$) to homozygous low-impact sites (b) significantly increased with ROH length (bins: $0.1 \text{ MB} \leq \text{ROH} < 0.5 \text{ MB}$, $0.5 \text{ MB} \leq \text{ROH} < 1 \text{ MB}$, $1 \text{ MB} \leq \text{ROH} < 3 \text{ MB}$, and $\text{ROH} \geq 3 \text{ MB}$) in both white rhino subspecies. There were likely too few $\text{ROH} \geq 3 \text{ MB}$ ($n=7$ across all individuals) to accurately estimate load in large ROH for NWR. Dashed lines show the proportions of heterozygous sites with deleterious alleles and homozygous sites with deleterious alleles in non-autozygous regions of the genome.

heterozygous variants in non-autozygous regions had higher proportions of deleterious alleles compared to homozygous sites in ROH (paired *t*-test, $p=0.028$ for homozygous conserved and $p=1.53 \times 10^{-9}$ for homozygous moderate-impact conserved mutations). Together, these results imply that recent inbreeding may be responsible for a disproportionate share of homozygous genetic load, as heterozygous sites in non-autozygous regions of the genome are converted to homozygous states in ROH through inbreeding.

To assess the possible functional impacts of genetic load in these populations, we tested for functional enrichment among genes with deleterious variants. Among genes that had missense mutations at evolutionarily conserved sites ($RS > 4$), NWR had 42 enriched GO terms (Table S3). The most significantly enriched of these included heart development, cellular component morphogenesis, inner ear development, direction of mechanical stimulus involved in sensory perception of sound, epidermal cell differentiation, and cilium organization (all FDR-corrected $p < 0.01$). Interestingly, three of the top six terms were also enriched among genes with private, missense variants in SWR (heart development, cellular component

morphogenesis, and cilium organization). In SWR, the top five most significant GO terms (cilium or flagellum-dependent cell motility, cilium-dependent cell motility, heart development, microtubule-based movement, and biological adhesion; all FDR-corrected $p < 0.02$) were also significantly enriched among NWR (Table S4).

3.3 | Simulations of northern white rhino recovery

To estimate the potential fitness impacts of genetic load on a re-established NWR population, we simulated genetic load and its fitness cost. We started with eight cryopreserved NWR cell lines and simulated 10 generations with and without supplementation of founder genomes in subsequent generations (supplemented and unsupplemented scenarios, respectively). In all scenarios, there was an initial increase in heterozygosity and decrease in homozygosity in generation one, likely resulting from substructure within the group of NWR founders (Sánchez-Barreiro et al., 2021). Starting in generation two, homozygous load and realized load increased across all deleterious mutation types and fitness decreased, changing most rapidly under the low-growth scenarios (Figure 4; Figure S10).

In unsupplemented populations with little to no population growth (0–10% per generation), homozygous load and realized load typically exceeded that observed in the SWR population by generations 7–10 (Figure 4a). Under the higher-growth scenarios (40–50%), the increase in homozygous load typically leveled off around generation five and did not exceed that of SWR over the 10 simulated generations. The number of heterozygous deleterious sites decreased over time as they were largely converted to homozygous states. There was little change over time in the overall number of deleterious alleles per genome, as predicted from population genetic theory (Pinto et al., 2023). This is because, of all polymorphic sites that became fixed by generation 10, a substantial portion (~10–30%) were fixed for the deleterious allele, especially in the low-growth scenarios (Figure S11). Fitness declined in all unsupplemented populations, but most rapidly in the 0 and 10% growth scenarios, on average falling below the mean value for SWR by generations six to eight (Figure 4c). Fitness declines in the 40–50% growth scenarios typically leveled off before generation 10 and remained higher than the median fitness of SWR until the end of the simulations. The results were qualitatively similar using different selection coefficients (Figure S7).

In populations supplemented with founder genomes in subsequent generations, trends in fitness and genetic load over time were less dependent on population growth (Figure 4b,d; Figure S12). Numbers of homozygous and heterozygous sites were similar for populations with different growth rates (Figure 4b), and there was little separation of the fitness trajectories under different growth rates (Figure 4d). Increases in homozygous load and decreases in both heterozygous load and fitness tended to level off around generation four in all scenarios. Populations with 0% growth showed high stochasticity, likely because the introduction of a random founder genome each generation had a proportionally larger impact

et al., 2022; Hedrick & García-Dorado, 2016). The larger number of deleterious alleles in the NWR population are segregating at low frequencies and are largely masked in heterozygosity with, presumably, relatively little impact on fitness. By contrast, the SWR population has more fixed deleterious alleles, both by count and in proportion to the total number of fixed derived alleles. The large number of fixed alleles may reflect the impact of recent bottlenecks. Extreme bottlenecks may lead to the rapid fixation of many alleles across the genome (Dussex et al., 2023), including the conversion of masked to realized load (Bertorelle et al., 2022). The proportional overrepresentation of deleterious alleles among those fixed reflects longer-term demography. Enrichment of deleterious alleles suggests the impact of genetic drift from small N_e over the long term (Wilder et al., 2023), as strong genetic drift overwhelms purifying selection, leading to the accumulation and fixation of mildly and moderately deleterious alleles (Lynch et al., 1995).

Estimates of more recent N_e from the 2dSFS ($N_e=474$ and 140 for NWR and SWR, respectively) are likely influenced by the severe population bottlenecks in the past century. The longer duration of the bottleneck by the time the SWR individuals analyzed here were born (1968–2012) resulted in higher inbreeding coefficients, lower heterozygosity, more fixed deleterious alleles, and higher homozygous load compared to NWR, which was still declining when the banked individuals were born (1952–1975). The long tracts of ROH in the SWR genome reflect higher inbreeding in SWR over the past two centuries. ROH tracts stemming from this period may have an outsized impact on fitness, since long ROH are enriched for homozygous deleterious variants relative to shorter ROH and non-autozygous regions of the genome, which have been exposed to selection in homozygosity for longer periods. Because heterozygous sites in non-autozygous regions have proportionally more deleterious alleles, future inbreeding in a restored NWR population is likely to bring together regions enriched for heterozygous deleterious variants into homozygosity in long ROH.

For both subspecies, some of the most significantly enriched functions of genes with deleterious mutations were related to cilium- or flagellum-dependent cell motility, which may play a role in sperm motility and fertility. High rates of abnormal sperm have been found in several rhinoceros species (O'Brien & Roth, 2000; Roth et al., 2005; Stoops et al., 2010), including white rhinos (Hermes et al., 2005; Reid et al., 2009), and 81% of males in zoo settings do not sire offspring (Hermes et al., 2005). Abnormal sperm and low sperm motility are hallmarks of inbreeding depression across taxa (Fitzpatrick & Evans, 2009; Gage et al., 2006; Gomendio et al., 2000; Roldan et al., 1998; Wildt et al., 1983; Zajitschek et al., 2009). While it is possible that deleterious mutations in genes important for sperm viability may play a role in inbreeding depression, GO terms directly related to flagellated sperm motility nested within the broader cilium- or flagellum-dependent cell motility term were not significantly enriched, making it difficult to draw direct links between deleterious variation and this trait. The genomic mechanism underlying this commonly

observed phenotype of highly inbred populations warrants further investigation.

4.2 | Viability of a restored NWR population

Despite the comparatively high burden of homozygous deleterious alleles, the SWR population has mounted a remarkable recovery from dozens of individuals, suggesting viability of SWR was not severely limited by genetic load. Recovery of the NWR population will depend on the mitigation of threats, successful implementation of cellular conservation technologies, and on the standing genetic variation of cell lines from 12 NWR individuals. In a population established by few founders, some level of inbreeding will likely be unavoidable, leading to the conversion of masked heterozygous load to realized homozygous load and loss of fitness (Hedrick & García-Dorado, 2016). Assuming deleterious variants have similar impacts in SWR and NWR genomes, our findings suggest that a NWR population founded by cryopreserved cells from eight unrelated individuals would not be constrained by low fitness if founder genomes are repeatedly reintroduced in subsequent generations. By contrast, a population restored in a single founding event would require rapid population growth to maintain fitness similar to that of SWR. Populations unsupplemented by reintroduced founders showed loss of genetic diversity and increased homozygosity of all deleterious mutation types under every simulated scenario. Fitness from genetic load typically fell below the average for SWR, except when growth was $\geq 30\%$ per generation (equivalent to an increase from eight to >108 individuals over 10 generations). Given that the SWR population grew from <200 individuals in the early 1900s to 20,000 today (a per-generation growth rate of $\sim 33\text{--}45\%$), the goal may not be unrealistic. In captive settings, however, SWR populations have seen negative growth rates due to poor reproductive success of captive-born females (Swaigood et al., 2006). Low female fertility is thought to be caused by exposure to high phytoestrogen levels during development through the maternal diet and microbial metabolites (Tubbs et al., 2016; Williams et al., 2019). Diet modification can boost fertility and possibly improve population growth (Tubbs et al., 2017), but this issue, along with the potential for low sperm viability, highlights potential barriers to achieving the rapid growth needed to avoid high realized load. Restoration of populations from cryopreserved cell lines allows lost genetic variation to be repeatedly reintroduced as it is lost each generation. This allows for fixed and homozygous deleterious alleles to again be masked in heterozygosity to relieve inbreeding depression, escaping the requirement for rapid growth of the restored population to avoid severe inbreeding depression.

One possible alternative to the reintroduction of founders may be to use a genetic rescue strategy, hybridizing NWR and SWR to introduce genetic variation into a restored NWR population. At least one hybrid is known, though it never reproduced (Saragusty et al., 2016). However, the subspecies have been isolated far longer than standard guidelines for avoiding outbreeding depression

(Frankham et al., 2019), and loci that may cause genetic incompatibilities are difficult to predict from genomic data, making reintroduction of founders a potentially safer strategy.

Forward genomic simulations are increasingly used in conservation for understanding the constraints on population viability and recovery, enabling exploration of population dynamics and outcomes under a given set of assumptions. For example, genomic simulations have been used to assess the potential recovery of endangered vaquitas (Robinson et al., 2022), estimate purging in Channel Island foxes (Robinson et al., 2018), model genetic rescue through facilitated gene flow (Kyriazis et al., 2021), reveal adaptation in response to fisheries (Therkildsen et al., 2019), and predict the impact of habitat loss on genomic erosion (Pinto et al., 2023). Simulations have even been proposed as a potential source of data for training models to predict IUCN extinction risk categories and recovery potential of species (van Oosterhout et al., 2022a, 2022b). Here, we used forward-simulated populations from empirically derived genetic variation (Speak et al., 2023), a novel approach that is especially relevant when the founders of a future restored population are known.

Despite the advancement of genomic simulation tools and our ability to estimate the impact of genome-wide genetic variation in non-model species, there are many caveats to the interpretation of both. Quantifying the effect of mutations on individual fitness is notoriously complex and plagued by a general lack of data (Grueber & Sunnucks, 2022). To estimate fitness variation from genomic variants, we applied selection and dominance coefficients estimated from human genomic data and evolutionary constraints (Henn et al., 2016; Peischl et al., 2018). We used a simple evolutionary model and assumed that mutations have similar fitness effects in NWR and SWR. However, the fitness effects of mutations can differ between species, populations, or even individuals (Grossen et al., 2020; Mee & Yeaman, 2019; Zhang et al., 2020), and evolutionary constraint can reflect past selection on loci that no longer affect fitness (Huber et al., 2020). Given the moderate depth of coverage of the genome, heterozygous sites are likely underestimated (Kishikawa et al., 2019), which may impact the predicted fitness trajectory of the more heterozygous NWR genomes in particular. Beneficial mutations, which were not modeled in this study, also likely influence individual variation in fitness. Populations managed under human care and with ART may experience different selective pressures, some of which may lead to maladaptation (van Oosterhout et al., 2022a, 2022b; Williams & Hoffman, 2009). For these reasons, estimates of the fitness effects of genetic variants are imprecise, and management decisions based solely on presumably functional genetic variation of individuals can be misleading (Kardos et al., 2021; Ralls et al., 2020). As genomic data representing diverse taxa continue to grow, and especially when coupled with databases linking survival and lifetime reproduction (Grueber & Sunnucks, 2022), so too will our power to link genotype to genomic health, fitness, and population viability and eventually use genomics-informed strategies to manage realized load (Speak et al., 2023). These insights will be key for applying conservation strategies, including rapidly advancing in vitro conservation technologies, with greater precision.

Our simulations highlight the value of biobanks for restoring populations from even a small number of individuals, and demonstrate how cellular-based conservation strategies can be powerful resources for rescuing populations from the brink of extinction. However, these technologies are not panaceas, and on their own will not solve the extinction crisis and global loss of intraspecific genetic variation. Cellular-based restoration and de-extinction may cost many times more than traditional conservation strategies that protect natural populations, and those protections will still be needed after restoration. Rather, these established conservation methods can be augmented by new cellular-based tools (Phelan et al., 2021). Technological advances may remove some hurdles, but at present, application of the technologies envisioned in this study required a surrogate species for gestation and become feasible because viable cell lines from multiple individuals are available. Biobanking of viable somatic and germline cells today represents a significant opportunity for conserving biological diversity, building the capacity to help restore genetic variation to critically small or declining populations in the future.

ACKNOWLEDGEMENTS

We thank Love Dálen for early access to genomic data, Mark Diekhans for providing computational resources, Joel Armstrong and Ian Fiddes for help with the annotation, and Fátima Sánchez-Barreiro, Chris Kyriazis, Erik Funk, and Chris Tubbs for advice with interpretation. Three anonymous reviewers provided insightful feedback that improved the paper. Funding was through San Diego Zoo Wildlife Alliance.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

Raw sequence data for this study are available at NCBI: PRJNA394025; ENA: ERS6578958, ERS6578959, ERS6578960, ERS6578961, ERS6578962, ERS6578963, ERS6578964, ERS6578965, ERS6578966. Scripts are available on github at: https://github.com/apwilder/WhiteRhino_GeneticLoad.

ORCID

Aryn P. Wilder  <https://orcid.org/0000-0003-4970-6891>
 Cynthia C. Steiner  <https://orcid.org/0000-0002-2131-8072>
 Sarah Hendricks  <https://orcid.org/0000-0001-9571-9461>
 Benjamin C. Haller  <https://orcid.org/0000-0003-1874-8327>
 Oliver A. Ryder  <https://orcid.org/0000-0003-2427-763X>

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–3402.

- Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I. T., Novak, A. M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., Genereux, D., Johnson, J., Marinescu, V. D., Alföldi, J., Harris, R. S., Lindblad-Toh, K., Haussler, D., Karlsson, E., Jarvis, E. D., ... Paten, B. (2020). Progressive cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, 587(7833), 246–251.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. The gene ontology consortium. *Nature Genetics*, 25(1), 25–29.
- Beichman, A. C., Phung, T. N., & Lohmueller, K. E. (2017). Comparison of single genome and allele frequency data reveals discordant demographic histories. *G3*, 7(11), 3605–3620.
- Ben-Nun, I. F., Montague, S. C., Houck, M. L., Tran, H. T., Garitaonandia, I., Leonardo, T. R., Wang, Y.-C., Charter, S. J., Laurent, L. C., Ryder, O. A., & Loring, J. F. (2011). Induced pluripotent stem cells from highly endangered species. *Nature Methods*, 8(10), 829–831.
- Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., Morales, H. E., & van Oosterhout, C. (2022). Genetic load: Genomic estimates and applications in non-model animals. *Nature Reviews. Genetics*, 23(8), 492–503.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.
- Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T.-H., Davis, F. G., Bateman, J., Bryant, S., Zuzikowski, A. R., Tsai, S. L., Coyne, S., Ye, W. W., Freeman, R. M., Jr., Peshkin, L., Tabin, C. J., ... Whited, J. L. (2017). A tissue-mapped axolotl De novo transcriptome enables identification of limb regeneration factors. *Cell Reports*, 18(3), 762–776.
- Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., Mace, G. M., Tilman, D., Wardle, D. A., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S., & Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature*, 486(7401), 59–67.
- Christmas, M. J., Kaplow, I. M., Genereux, D. P., Dong, M. X., Hughes, G. M., Li, X., Sullivan, P. F., Hindle, A. G., Andrews, G., Armstrong, J. C., Bianchi, M., Breit, A. M., Diekhans, M., Fanter, C., Foley, N. M., Goodman, D. B., Goodman, L., Keough, K. C., Kirilenko, B., ... Karlsson, E. K. (2023). Evolutionary constraint and innovation across hundreds of placental mammals. *Science*, 380(6643), eabn3943.
- Cingolani, P. (2022). Variant annotation and functional prediction: SnpEff. *Methods in Molecular Biology*, 2493, 289–314.
- Crees, J. J., Collins, A. C., Stephenson, P. J., Meredith, H. M. R., Young, R. P., Howe, C., Price, M. R. S., & Turvey, S. T. (2016). A comparative approach to assess drivers of success in mammalian conservation recovery programs. *Conservation Biology: The Journal of the Society for Conservation Biology*, 30(4), 694–705.
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. <https://doi.org/10.1093/gigascience/giab008>
- Drake, N. A., Blench, R. M., Armitage, S. J., Bristow, C. S., & White, K. H. (2011). Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. *Proceedings of the National Academy of Sciences of the United States of America*, 108(2), 458–462.
- Dumont, B. L., & Payseur, B. A. (2008). Evolution of the genomic rate of recombination in mammals. *Evolution; International Journal of Organic Evolution*, 62(2), 276–294.
- Dussex, N., Morales, H. E., Grossen, C., Dalén, L., & van Oosterhout, C. (2023). Purging and accumulation of genetic load in conservation. *Trends in Ecology & Evolution*, 38(10), 961–969.
- Emslie, R. (2020). *Ceratotherium simum* [dataset]. In *IUCN Red List of Threatened Species*. IUCN. <https://doi.org/10.2305/iucn.uk.2020-1.rlts.t4185a45813880.en>
- Emslie, R. H., & Knight, M. H. (2014). Update on African rhino status and poaching trends from IUCN SSC African rhino specialist group (AfRSG). Report to CITES standing committee 65th meeting.
- Fiddes, I. T., Armstrong, J., Diekhans, M., Nachtweide, S., Kronenberg, Z. N., Underwood, J. G., Gordon, D., Earl, D., Keane, T., Eichler, E. E., Haussler, D., Stanke, M., & Paten, B. (2018). Comparative annotation toolkit (CAT)-simultaneous clade and personal genome annotation. *Genome Research*, 28(7), 1029–1038.
- Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, 39, W29–W37.
- Fitzpatrick, J. L., & Evans, J. P. (2009). Reduced heterozygosity impairs sperm quality in endangered mammals. *Biology Letters*, 5(3), 320–323.
- Frankham, R., Ballou, J. D., Ralls, K., Dudash, M. R., Fenster, C. B., & Sunnucks, P. (2019). *A practical guide for genetic management of fragmented animal and plant populations*. Oxford University Press.
- Gage, M. J. G., Surridge, A. K., Tomkins, J. L., Green, E., Wiskin, L., Bell, D. J., & Hewitt, G. M. (2006). Reduced heterozygosity depresses sperm quality in wild rabbits, *Oryctolagus cuniculus*. *Current Biology: CB*, 16(6), 612–617.
- Gasse, F. (2000). Hydrological changes in the African tropics since the last glacial maximum. *Quaternary Science Reviews*, 19(1–5), 189–211.
- Gillis, J., Mistry, M., & Pavlidis, P. (2010). Gene function analysis in complex data sets using ErmineJ. *Nature Protocols*, 5(6), 1148–1159.
- Gilpin, M. E. (1986). *Minimum viable populations: Processes of species extinction* (pp. 19–34). Conservation biology: the science of scarcity and diversity.
- Gomendio, M., Cassinello, J., & Roldan, E. R. (2000). A comparative study of ejaculate traits in three endangered ungulates with different levels of inbreeding: Fluctuating asymmetry as an indicator of reproductive and genetic stress. *Proceedings. Biological Sciences/the Royal Society*, 267(1446), 875–882.
- Gross, M. (2021). In vitro conservation. *Current Biology: CB*, 31(18), R1065–R1068.
- Grossen, C., Guillaume, F., Keller, L. F., & Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in alpine ibex. *Nature Communications*, 11(1), 1001.
- Grueber, C. E., & Sunnucks, P. (2022). Using genomics to fight extinction. *Science*, 376(6593), 574–575.
- Haller, B. C., & Messer, P. W. (2023). SLiM 4: Multispecies eco-evolutionary modeling. *The American Naturalist*, 201(5), E127–E139.
- Hayashi, M., Zywitza, V., Naitou, Y., Hamazaki, N., Goeritz, F., Hermes, R., Holtze, S., Lazzari, G., Galli, C., Stejskal, J., Diecke, S., Hildebrandt, T. B., & Hayashi, K. (2022). Robust induction of primordial germ cells of white rhinoceros on the brink of extinction. *Science Advances*, 8(49), eabp9683.
- Hedrick, P. W., & García-Dorado, A. (2016). Understanding inbreeding depression, purging, and genetic rescue. *Trends in Ecology & Evolution*, 31(12), 940–952.
- Henn, B. M., Botigué, L. R., Peischl, S., Dupanloup, I., Lipatov, M., Maples, B. K., Martin, A. R., Musharoff, S., Cann, H., Snyder, M. P., Excoffier, L., Kidd, J. M., & Bustamante, C. D. (2016). Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 113(4), E440–E449.
- Hermes, R., Hildebrandt, T. B., Blottner, S., Walzer, C., Silinski, S., Patton, M. L., Wibbelt, G., Schwarzenberger, F., & Göritz, F. (2005). Reproductive soundness of captive southern and northern white rhinoceroses (*Ceratotherium simum simum*, C.s. *cottoni*): Evaluation of male genital tract morphology and semen quality before and after cryopreservation. *Theriogenology*, 63(1), 219–238.

- Hildebrandt, T. B., Hermes, R., Colleoni, S., Diecke, S., Holtze, S., Renfree, M. B., Stejskal, J., Hayashi, K., Drukker, M., Loi, P., Göritz, F., Lazzari, G., & Galli, C. (2018). Embryos and embryonic stem cells from the white rhinoceros. *Nature Communications*, 9(1), 2589.
- Hildebrandt, T. B., Hermes, R., Goeritz, F., Appeltant, R., Colleoni, S., de Mori, B., Diecke, S., Drukker, M., Galli, C., Hayashi, K., Lazzari, G., Loi, P., Payne, J., Renfree, M., Seet, S., Stejskal, J., Swegen, A., Williams, S. A., Zainuddin, Z. Z., & Holtze, S. (2021). The ART of bringing extinction to a freeze – history and future of species conservation, exemplified by rhinos. *Theriogenology*, 169, 76–88.
- Hildebrandt, T. B., Holtze, S., Colleoni, S., Hermes, R., Stejskal, J., Lekool, I., Ndeereh, D., Omondi, P., Kariuki, L., Mijele, D., Mutisya, S., Ngulu, S., Diecke, S., Hayashi, K., Lazzari, G., de Mori, B., Biasetti, P., Quaggio, A., Galli, C., & Goeritz, F. (2023). In vitro fertilization program in white rhinoceros. *Reproduction*, 166(6), 383–399.
- Huber, C. D., Kim, B. Y., & Lohmueller, K. E. (2020). Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. *PLoS Genetics*, 16(5), e1008827.
- Jolly, D., Harrison, S. P., Damnati, B., & Bonnefille, R. (1998). Simulated climate and biomes of Africa during the late quaternary: Comparison with pollen and lake status data. *Quaternary Science Reviews*, 17, 629–657.
- Jouganous, J., Long, W., Ragsdale, A. P., & Gravel, S. (2017). Inferring the joint demographic history of multiple populations: Beyond the diffusion approximation. *Genetics*, 206(3), 1549–1567.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, 40(Database issue), D109–D114.
- Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M., Tallmon, D. A., & Funk, W. C. (2021). The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences of the United States of America*, 118(48), e2104642118. <https://doi.org/10.1073/pnas.2104642118>
- Kardos, M., Zhang, Y., Parsons, K. M., Kang, H., Xu, X., Liu, X., Matkin, C. O., Zhang, P., Ward, E. J., Hanson, M. B., Emmons, C., Ford, M. J., Fan, G., & Li, S. (2023). Inbreeding depression explains killer whale population dynamics. *Nature Ecology & Evolution*, 7(5), 675–686.
- Kim, D., Perte, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4), R36.
- Kishikawa, T., Momozawa, Y., Ozeki, T., Mushiroda, T., Inohara, H., Kamatani, Y., Kubo, M., & Okada, Y. (2019). Empirical evaluation of variant calling accuracy using ultra-deep whole-genome sequencing data. *Scientific Reports*, 9(1), 1–10.
- Korody, M. L., Ford, S. M., Nguyen, T. D., Pivaroff, C. G., Valiente-Alandi, I., Peterson, S. E., Ryder, O. A., & Loring, J. F. (2021). Rewinding extinction in the northern White rhinoceros: Genetically diverse induced pluripotent stem cell Bank for Genetic Rescue. *Stem Cells and Development*, 30(4), 177–189.
- Kyriazis, C. C., Wayne, R. K., & Lohmueller, K. E. (2021). Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evolution Letters*, 5(1), 33–47.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature Methods*, 9(4), 357–359.
- Lynch, M., Conery, J., & Burger, R. (1995). Mutation accumulation and the extinction of small populations. *The American Naturalist*, 146(4), 489–518.
- Mee, J. A., & Yeaman, S. (2019). Unpacking conditional neutrality: Genomic signatures of selection on conditionally beneficial and conditionally deleterious mutations. *The American Naturalist*, 194(4), 529–540.
- Moodley, Y., Russo, I.-R. M., Robovský, J., Dalton, D. L., Kotzé, A., Smith, S., Stejskal, J., Ryder, O. A., Hermes, R., Walzer, C., & Bruford, M. W. (2018). Contrasting evolutionary history, anthropogenic declines and genetic contact in the northern and southern white rhinoceros. *Proceedings. Biological Sciences/the Royal Society*, 285(1890), 20181567. <https://doi.org/10.1098/rspb.2018.1567>
- Moodley, Y., Westbury, M. V., Russo, I.-R. M., Gopalakrishnan, S., Rakotoarivelo, A., Olsen, R.-A., Prost, S., Tunstall, T., Ryder, O. A., Dalén, L., & Bruford, M. W. (2020). Interspecific gene flow and the evolution of specialization in black and White rhinoceros. *Molecular Biology and Evolution*, 37(11), 3105–3117.
- Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., & Durbin, R. (2016). BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*, 32(11), 1749–1751.
- Noskova, E., Ulyantsev, V., Koepfli, K.-P., O'Brien, S. J., & Dobrynin, P. (2020). GADMA: Genetic algorithm for inferring demographic history of multiple populations from allele frequency spectrum data. *GigaScience*, 9(3), gaa005.
- O'Brien, J. K., & Roth, T. L. (2000). Post-coital sperm recovery and cryopreservation in the Sumatran rhinoceros (*Dicerorhinus sumatrensis*) and application to gamete rescue in the African black rhinoceros (*Diceros bicornis*). *Journal of Reproduction and Fertility*, 118(2), 263–271.
- Peischl, S., Dupanloup, I., Foucal, A., Jomphe, M., Bruat, V., Grenier, J.-C., Gouy, A., Gilbert, K. J., Gbeha, E., Bosshard, L., Hip-Ki, E., Agbessi, M., Hodgkinson, A., Vézina, H., Awadalla, P., & Excoffier, L. (2018). Relaxed selection during a recent human expansion. *Genetics*, 208(2), 763–777.
- Phelan, R., Baumgartner, B., Brand, S., Brister, E., Burgiel, S. W., Alta Charo, R., Coche, I., Cofrancesco, A., Delborne, J. A., Edwards, O., Fisher, J. P., Gaywood, M., Gordon, D. R., Howald, G., Hunter, M. E., Kareiva, P., Mankad, A., Marvier, M., Moseby, K., ... Robbins, P. (2021). Intended consequences statement. *Conservation Science and Practice*, 3(4), e371.
- Pinto, A. V., Hansson, B., Patramanis, I., Morales, H. E., & van Oosterhout, C. (2023). The impact of habitat loss and population fragmentation on genomic erosion. *Conservation Genetics*, 25, 49–57. <https://doi.org/10.1007/s10592-023-01548-9>
- Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Arnold, R., Rattei, T., Letunic, I., Doerks, T., Jensen, L. J., von Mering, C., & Bork, P. (2012). eggNOG v3.0: Orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Research*, 40(Database issue), D284–D289.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), 559–575.
- Ralls, K., Sunnucks, P., Lacy, R. C., & Frankham, R. (2020). Genetic rescue: A critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful genetic variation. *Biological Conservation*, 251, 108784.
- Reid, C. E., Hermes, R., Blottner, S., Goeritz, F., Wibbelt, G., Walzer, C., Bryant, B. R., Portas, T. J., Streich, W. J., & Hildebrandt, T. B. (2009). Split-sample comparison of directional and liquid nitrogen vapour freezing method on post-thaw semen quality in white rhinoceroses (*Ceratotherium simum simum* and *Ceratotherium simum cottoni*). *Theriogenology*, 71(2), 275–291.
- Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E., & Wayne, R. K. (2018). Purging of strongly deleterious mutations explains Long-term persistence and absence of inbreeding depression in Island foxes. *Current Biology: CB*, 28(21), 3487–3494.e4.
- Robinson, J. A., Kyriazis, C. C., Nigenda-Morales, S. F., Beichman, A. C., Rojas-Bracho, L., Robertson, K. M., Fontaine, M. C., Wayne, R. K., Lohmueller, K. E., Taylor, B. L., & Morin, P. A. (2022). The critically endangered vaquita is not doomed to extinction by inbreeding depression. *Science*, 376(6593), 635–639.
- Roldan, E. R., Cassinello, J., Abaigar, T., & Gomendio, M. (1998). Inbreeding, fluctuating asymmetry, and ejaculate quality in an

- endangered ungulate. *Proceedings. Biological Sciences/the Royal Society*, 265(1392), 243–248.
- Rookmaaker, K. (2000). The alleged population reduction of the southern White rhinoceros (*Ceratotherium simum simum*) and the successful recovery. *Saugetierkundliche Mitteilungen*, 45(2), 55–70.
- Roth, T. L., Stoops, M. A., Atkinson, M. W., Blumer, E. S., Campbell, M. K., Cameron, K. N., Citino, S. B., & Maas, A. K. (2005). Semen collection in rhinoceroses (*Rhinoceros unicornis*, *Diceros bicornis*, *Ceratotherium simum*) by electroejaculation with a uniquely designed probe. *Journal of Zoo and Wildlife Medicine: Official Publication of the American Association of zoo Veterinarians*, 36(4), 617–627.
- Sánchez-Barreiro, F., Gopalakrishnan, S., Ramos-Madrugal, J., Westbury, M. V., de Manuel, M., Margaryan, A., Ciucani, M. M., Vieira, F. G., Patramanis, Y., Kalthoff, D. C., Timmons, Z., Sicheritz-Pontén, T., Dalén, L., Ryder, O. A., Zhang, G., Marqués-Bonet, T., Moodley, Y., & Gilbert, M. T. P. (2021). Historical population declines prompted significant genomic erosion in the northern and southern white rhinoceroses (*Ceratotherium simum*). *Molecular Ecology*, 30(23), 6355–6369.
- Saragusty, J., Diecke, S., Drukker, M., Durrant, B., Friedrich Ben-Nun, I., Galli, C., Göritz, F., Hayashi, K., Hermes, R., Holtze, S., Johnson, S., Lazzari, G., Loi, P., Loring, J. F., Okita, K., Renfree, M. B., Seet, S., Voracek, T., Stejskal, J., ... Hildebrandt, T. B. (2016). Rewinding the process of mammalian extinction. *Zoo Biology*, 35(4), 280–292.
- Skinner, C. B., & Poulsen, C. J. (2016). The role of fall season tropical plumes in enhancing Saharan rainfall during the African humid period. *Geophysical Research Letters*, 43(1), 349–358.
- Speak, S., Birley, T., Bortoluzzi, C., Clark, M., Percival-Alwyn, L., Morales, H., & Van Oosterhout, C. (2023). Genomics-informed captive breeding can reduce inbreeding depression and the genetic load in zoo populations. <https://doi.org/10.22541/au.169566325.52568264/v1>
- Stoops, M. A., Atkinson, M. W., Blumer, E. S., Campbell, M. K., & Roth, T. L. (2010). Semen cryopreservation in the Indian rhinoceros (*Rhinoceros unicornis*). *Theriogenology*, 73(8), 1104–1115.
- Swaigood, R. R., Dickman, D. M., & White, A. M. (2006). A captive population in crisis: Testing hypotheses for reproductive failure in captive-born southern white rhinoceros females. *Biological Conservation*, 129(4), 468–476.
- Szpiech, Z. A., Xu, J., Pemberton, T. J., Peng, W., Zöllner, S., Rosenberg, N. A., & Li, J. Z. (2013). Long runs of homozygosity are enriched for deleterious variation. *American Journal of Human Genetics*, 93(1), 90–102.
- Therkildsen, N. O., Wilder, A. P., Conover, D. O., Munch, S. B., Baumann, H., & Palumbi, S. R. (2019). Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. *Science*, 365(6452), 487–490.
- Thompson, E. A. (2013). Identity by descent: Variation in meiosis, across genomes, and in populations. *Genetics*, 194(2), 301–326.
- Tubbs, C. W., Durrant, B., & Milnes, M. (2017). Reconsidering the use of soy and alfalfa in southern white rhinoceros diets. *Pachyderm*, 58, 135–139.
- Tubbs, C. W., Moley, L. A., Ivy, J. A., Metrione, L. C., LaClaire, S., Felton, R. G., Durrant, B. S., & Milnes, M. R. (2016). Estrogenicity of captive southern white rhinoceros diets and their association with fertility. *General and Comparative Endocrinology*, 238, 32–38.
- Tunstall, T., Kock, R., Vahala, J., Diekhans, M., Fiddes, I., Armstrong, J., Paten, B., Ryder, O. A., & Steiner, C. C. (2018). Evaluating recovery potential of the northern white rhinoceros from cryopreserved somatic cells. *Genome Research*, 28(6), 780–788.
- Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the cloud: Using docker, GATK, and WDL in Terra*. O'Reilly Media.
- van der Valk, T., Díez-Del-Molino, D., Marques-Bonet, T., Guschanski, K., & Dalén, L. (2019). Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. *Current Biology: CB*, 29(1), 165–170.e6.
- van Oosterhout, C., Speak, S. A., Birley, T., Bortoluzzi, C., Percival-Alwyn, L., Urban, L. H., Groombridge, J. J., Segelbacher, G., & Morales, H. E. (2022a). Genomic erosion in the assessment of species extinction risk and recovery potential. *bioRxiv*. 2022.09.13.507768. <https://doi.org/10.1101/2022.09.13.507768>
- Wilder, A. P., Supple, M. A., Subramanian, A., Mudide, A., Swofford, R., Serres-Armero, A., Steiner, C., Koepfli, K.-P., Genereux, D. P., Karlsson, E. K., Lindblad-Toh, K., Marques-Bonet, T., Munoz Fuentes, V., Foley, K., Meyer, W. K., Ryder, O. A., & Shapiro, B. (2023). The contribution of historical processes to contemporary extinction risk in placental mammals. *Science*, 380(6643), eabn5856.
- Wildt, D. E., Bush, M., Howard, J. G., O'Brien, S. J., Meltzer, D., Van Dyk, A., Ebendes, H., & Brand, D. J. (1983). Unique seminal quality in the south African cheetah and a comparative evaluation in the domestic cat. *Biology of Reproduction*, 29(4), 1019–1025.
- Williams, C. L., Ybarra, A. R., Meredith, A. N., Durrant, B. S., & Tubbs, C. W. (2019). Gut microbiota and phytoestrogen-associated infertility in southern White rhinoceros. *mBio*, 10(2), e00311–e00319. <https://doi.org/10.1128/mBio.00311-19>
- Williams, S. E., & Hoffman, E. A. (2009). Minimizing genetic adaptation in captive breeding programs: A review. *Biological Conservation*, 142(11), 2388–2400.
- Wisely, S. M., Ryder, O. A., Santymire, R. M., Engelhardt, J. F., & Novak, B. J. (2015). A road map for 21st century genetic restoration: Gene Pool enrichment of the black-footed ferret. *The Journal of Heredity*, 106(5), 581–592.
- Zajitschek, S. R. K., Lindholm, A. K., Evans, J. P., & Brooks, R. C. (2009). Experimental evidence that high levels of inbreeding depress sperm competitiveness. *Journal of Evolutionary Biology*, 22(6), 1338–1345.
- Zhang, Y., Stern, A. J., & Nielsen, R. (2020). Evolution of the genetic architecture of local adaptations under genetic rescue is determined by mutational load and polygenicity. *bioRxiv*. 2020.11.09.374413. <https://doi.org/10.1101/2020.11.09.374413>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wilder, A. P., Steiner, C. C., Hendricks, S., Haller, B. C., Kim, C., Korody, M. L., & Ryder, O. A. (2024). Genetic load and viability of a future restored northern white rhino population. *Evolutionary Applications*, 17, e13683. <https://doi.org/10.1111/eva.13683>