Conservation management strategy impacts inbreeding and mutation load in scimitar-horned oryx

Emily Humble[1], Martin A Stoffel[2], Kara Dicks[3], Alex D Ball[3], Rebecca M Gooley[4,5], Justin Chuven[6,7], Ricardo Pusey[6], Mohammed Al Remeithi[6], Klaus-Peter Koepfli[4,5], Budhan Pukazhenthi[5], Helen Senn[3], Rob Ogden[1]

[1]Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, EH25 9RG, Edinburgh, UK

[2]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3FL, UK

[3]RZSS WildGenes, Conservation Department, Royal Zoological Society of Scotland, Edinburgh, EH12 6TS, UK

[4]Smithsonian-Mason School of Conservation, George Mason University, Front Royal, VA 22630 USA

[5]Smithsonian's National Zoo and Conservation Biology Institute, Center for Species Survival, Front Royal, Virginia 22630 and Washington, D.C. 20008 USA

[6]Terrestrial & Marine Biodiversity Sector, Environment Agency – Abu Dhabi, United Arab Emirates

[7]US Fish and Wildlife Service, Colorado, USA

Corresponding Author:

Emily Humble, Royal (Dick) School of Veterinary Studies and the Roslin Institute
University of Edinburgh, EH25 9RG, UK
Phone: 0131 506 210, Email: emily.humble@ed.ac.uk

Keywords:

Ex-situ populations, runs of homozygosity (ROH), deleterious mutations, reintroduction, effective population size

Competing interests:

The authors declare no competing interests.

Abstract

In an age of habitat loss and overexploitation, small populations, both captive and wild, are increasingly facing the effects of isolation and inbreeding. Genetic management has therefore become a vital tool for ensuring population viability. However, little is known about how the type and intensity of intervention shape the genomic landscape of inbreeding and mutation load. We address this using whole genome sequence data of scimitar-horned oryx (*Oryx dammah*), an iconic antelope that has been subject to contrasting management strategies since it was declared extinct in the wild. We show that unmanaged populations are enriched for long runs of homozygosity (ROH) and have significantly higher inbreeding coefficients than managed populations. Additionally, despite the total number of deleterious alleles being similar across management strategies, the burden of homozygous deleterious genotypes was consistently higher in unmanaged groups. These findings emphasise the risks associated with deleterious mutations through multiple generations of inbreeding. As wildlife management strategies continue to diversify, our study reinforces the importance of maintaining genome-wide variation in vulnerable populations and has direct implications for one of the largest reintroduction attempts in the world.

Significance statement

Conservation genetic management is becoming increasingly important for safeguarding and restoring wildlife populations. Understanding how the intensity of intervention influences genomic components of fitness is therefore essential for supporting species viability. We investigate the impact of contrasting management strategies on the genomic landscape of inbreeding and mutation load in captive populations of scimitar-horned oryx. We reveal how several decades of management have prevented the formation of long runs of homozygosity and masked the expression of deleterious mutations. Our findings highlight the dynamics between inbreeding, mutation load and population size and have direct implications for future management of threatened species.

Main text

Introduction

Captive populations have become an essential insurance against extinctions in the wild (1). However, due to inbreeding and drift, they are intrinsically vulnerable to reduced genetic variation and the expression of partially recessive deleterious mutations (2–6). It is therefore of paramount importance that appropriate plans are in place to safeguard their potential as source populations. *Ex-situ* management strategies fall along a continuum from high-intensity pedigree-based breeding (7), to low-intensity pedigree-free group management (8, 9), to a complete absence of breeding intervention whatsoever. Empirical evidence on how these approaches influence the combined landscape of inbreeding and deleterious variation is limited (10, 11). As wildlife management strategies begin to diversify (12–15), there is a pressing need to leverage current genomic techniques to understand how genetic components of fitness are impacted by conservation intervention.

Alongside this, recent debate on the significance of neutral genetic variation in conservation biology has raised practical considerations for sourcing populations for restorations (16–20). For example, an increasing number of studies are uncovering genomic evidence for purging in the wild (21–26), some of which have used this to challenge the small population paradigm of conservation biology (27–29). Furthermore, simulation-based studies on the interaction between effective population size, genetic variation and extinction risk have called for more emphasis on functional genomic variation in genetic rescue attempts (18, 19). These observations go against decades of empirical and theoretical work in favour of maximising genetic variation to enhance population viability (30–33) including recent studies highlighting the complex dynamics of deleterious mutation frequencies in small populations (34–38). Founder selection for translocations rests on a complex set of considerations, with genetics making up only one component (39). In most cases, conservation practitioners will favour a unifying strategy to minimise risk and maximise return (40–42). In light of this, empirical data on the patterns of inbreeding and deleterious mutations in species undergoing active conservation management are urgently required.

Ex-situ populations of scimitar-horned oryx provide an excellent opportunity to evaluate the genomic consequences of management in the context of a global reintroduction. This iconic antelope was once widespread across North Africa, yet during the 20th century, hunting and land-use competition led to their rapid population decline and eventual extinction from the wild (43). Prior to disappearing, captive populations had already been established from what is

thought to be less than 100 animals originating from Chad in the 1960s (43). In the following years, the *ex-situ* population has grown to reach approximately 15,000 individuals (44). Around 1,000 of these are held within coordinated breeding programmes, but the vast majority are held in collections in places like Texas and the Arabian Peninsula where little to no genetic management takes place. Crucially, the scimitar-horned oryx is now being reintroduced back into its former range and *ex-situ* populations with varying management strategies have been used to source individuals for release. Here, we use runs of homozygosity (ROH) and predicted deleterious mutations to evaluate the impacts of captive-breeding practices on inbreeding and mutation load in scimitar-horned oryx, and discuss the implications for its ongoing management.

Results

We generated whole-genome sequence data for 49 scimitar-horned oryx from four *ex-situ* populations. Two of these, the EAZA *Ex situ* Programmes (EEP, n = 8) and the USA (n = 17), represent captive populations where genetic management practices are in place. The EEP population comprised individuals from zoological institutions across Europe. The USA population comprised individuals from both privately owned ranches and institutions within the AZA Species Survival Plan ® (SSP). The remaining populations from the Environment Agency – Abu Dhabi originate from two genetically unmanaged collections in the United Arab Emirates (EAD A: n = 9 and EAD B: n = 15). Census sizes for the EEP and SSP population are approximately 619 and 223, respectively, while those for EAD A and EAD B are approximately 3,000 and 70. For further details on population origins, management strategies and sampling approach, please refer to the Supplementary Material.

High coverage sequencing (~15X) was performed for 20 of the individuals and the remaining 29 were sequenced at a lower depth (6–8X, Table S1). Sequencing reads were mapped to the scimitar-horned oryx reference genome (45) and to account for coverage biases, SNPs and genotype likelihoods were called after downsampling high coverage individuals (see Methods for details). Analysis of population structure using NGSadmix and PCAngsd detected differentiation between the four sampling groups (Figures S1–3). Individual admixture proportions highlighted two major ancestral source populations (Figures S1A), with further hierarchical structure being resolved up to values of K=4 (Figures S1B and S2), corresponding to the four *ex-situ* groups. PCA distinguished EEP and USA populations as discrete clusters along PC2 and PC3, while EAD A and EAD B clustered separately along PC1 (Figure S4).

Levels of inbreeding across management strategies

To investigate how genomic patterns of inbreeding vary with management strategy, we examined the ROH landscape across individuals (Figure 1). The average number and total length of ROH was 247 (min = 65, max = 638) and 2.0 Mb (0.5–22.0 Mb) respectively, which on average spanned 20% of the autosomal genome (min = 0.03, max = 0.55, Figure 1A and Figure S4). Oryx from managed populations had significantly lower inbreeding coefficients (F_{ROH}) than oryx from unmanaged populations (ß = -0.19, 95% CI = -0.24–-0.14, $P = 6.43^{e-9}$, Figure 1A). This pattern was driven by both the number and length of ROH, the former being almost three times higher in the most inbred population than in the least inbred population (Figure 1B and Figure S5).

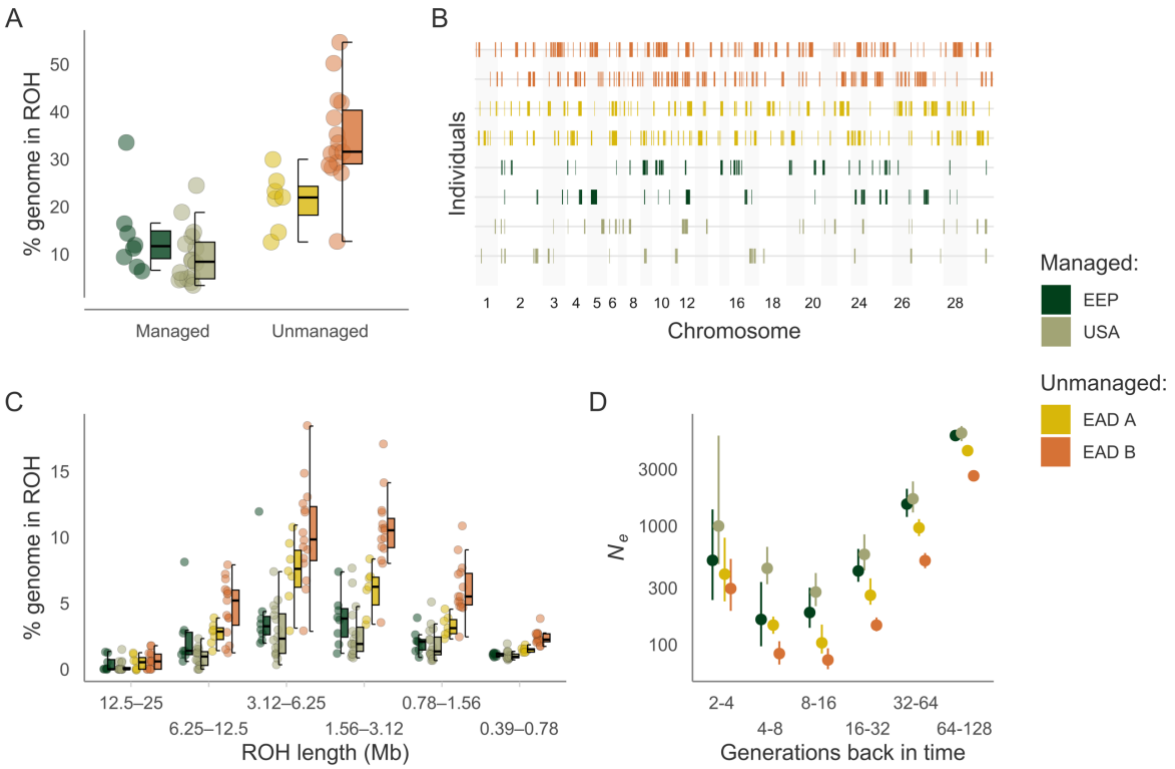

Figure 1. Runs of homozygosity (ROH) landscape across contrasting management strategies of scimitar-horned oryx. (A) Distribution of F_{ROH} across scimitar-horned oryx management strategies. Values were multiplied by 100 to reflect the percentage of the autosomal genome in ROH. Centre lines of boxplots reflect the median, bounds of the boxes extend from the first to the third quartile and upper and lower whiskers reflect variability outside the interquartile range. **(B)** ROH in the two individuals with intermediate inbreeding coefficients F_{ROH} from each population. **(C)** Distribution of ROH within different length classes. Data points represent the percentage of ROH of a given length within an individual's autosomal genome. **(D)** Effective population size estimates inferred from the mean F_{ROH} in a population for a given time-period (see Materials and Methods for details). Error bars represent 95% bootstrap confidence intervals.

ROH length distribution and recent demography

We also observed variation in the fraction of the genome in ROH of different length classes. There was a steep decrease in the genome fraction in ROH above lengths of around 6.25 Mb (Figure 1C). ROH longer than this made up a relatively small fraction of the genome, reaching a minimum average frequency of 0.4% between 12.5–25 Mb. ROH between 3.12–6.25 Mb had the highest frequency, making up on average 6.2% of an individual's genome. Although this pattern was observed in each population, there was variation in absolute proportions across individuals. For example, length class 3.12–6.25 Mb made up on average only 3% of the genome in the least inbred population, USA, while it comprised on average 10% in the

most inbred population, EAD B (Figure 1C). Interestingly, long ROH >12.5 Mb, which are likely the result of recent shared parental ancestry, were identified in less than 30% of individuals from managed populations, yet were present in over 60% of individuals from unmanaged populations.

As ROH lengths decrease, their underlying haplotypes are expected to have originated from ancestors further back in time (46). The genome fraction in ROH of different length classes can therefore provide insights into past changes in effective populations size (N_e) (47, 48). In line with this, our estimates of N_e based on individual inbreeding coefficients were inversely proportional to the fraction of the genome in ROH of different length classes (Figure 1D). N_e declines to reach a minimum of around 150 individuals between 8–16 generations ago, after which it shows a steady increase towards the present day. Managed populations had higher N_e estimates across all time-periods than unmanaged populations (mean N_e: USA = 1,672, EEP = 1,429 versus EAD A = 1,028, EAD B = 625, see also Figure S6). These patterns were reflected in estimates of mean pairwise nucleotide diversity which were also higher in managed (USA = 0.46 x 10^5, EEP = 0.44 x 10^5) than unmanaged populations (EAD A = 0.42 x 10^5, EAD B = 0.27 x 10^5).

Mutation load landscape across management strategies

We next investigated how mutation load varies across management strategies using metrics based on putative deleterious variants identified using annotation-based methods. As the overall patterns were qualitatively similar across two variant effect prediction software (Figure S7), results using annotations from SnpEff are presented here. We first estimated two components of mutation load; heterozygous and homozygous mutation load for both weakly (missense) and highly (loss of function) deleterious mutations. The heterozygous mutation load was calculated as the absolute number of missense and loss of function heterozygotes per individual. The homozygous mutation load was calculated as the absolute number of derived missense and loss of function homozygotes per individual. Based on the assumption that most deleterious mutations are (at least) partially recessive (2, 49–51) and therefore partially hidden from selection when in a heterozygous state, we considered the homozygous mutation load to be most informative of the fitness cost due to inbreeding in the current population, and the heterozygous mutation load to reflect the potential for inbreeding to reduce fitness in future generations.

Heterozygous mutation load for both missense and loss of function (LoF) mutations was consistently higher in managed than unmanaged populations (Missense: ß = 1079, 95% CI = 769–1390, P = 1.12^{e-8}, LoF: ß = 21.7, 95% CI = 12.7–30.6, P = 1.51^{e-5}, Figure 2A–B). As

expected, this pattern inversely tracked overall inbreeding levels, where individuals with lower inbreeding coefficients had a larger number of heterozygotes at missense and loss of function sites (Figure S8). In direct contrast, the homozygous mutation load for both missense and LoF mutations was lower in managed than in unmanaged populations (Missense: ß = -608, 95% CI = -779–-437, $P = 6.41^{e-9}$, LoF: ß = -13.6, 95% CI = -18.3–-8.86, $P = 6.99^{e-7}$, Figure 2C–D). Of the homozygous missense genotypes, 1,763 (11.5%), 6,790 (17.1%), 236 (0.81%) and 874 (5.43%) were due to alleles fixed in EAD A, EAD B, USA and EEP, respectively. Of the homozygous LoF genotypes, 173 (37.5%), 313 (28.4%), 46 (5.07%) and 95 (20.1%) were due to alleles fixed in EAD A, EAD B, USA and EEP, respectively.

To account for a scenario where deleterious mutations are approximately additive, we also calculated the total number of derived deleterious alleles per individual (52). Here, no significant difference in the total of number of LoF or missense alleles was observed between managed and unmanaged populations (Missense: ß = 136, 95% CI = -83.7–357, $P = 0.22$, LoF: ß = 5.53, 95% CI = -1.48–12.5, $P = 0.12$, Figure 2E–F). In addition, we used the measure *Rxy* to determine whether there was an excess of putative deleterious mutations in one management strategy over another. *Rxy* compares the relative frequency of derived alleles within a given impact category and is standardised over a set of intergenic SNPs, making it robust to population-specific biases. Unmanaged populations displayed a marginal excess of missense mutations compared to managed populations while no difference in the frequency of highly deleterious loss of function mutations could be detected between the management groups (Figure S9).

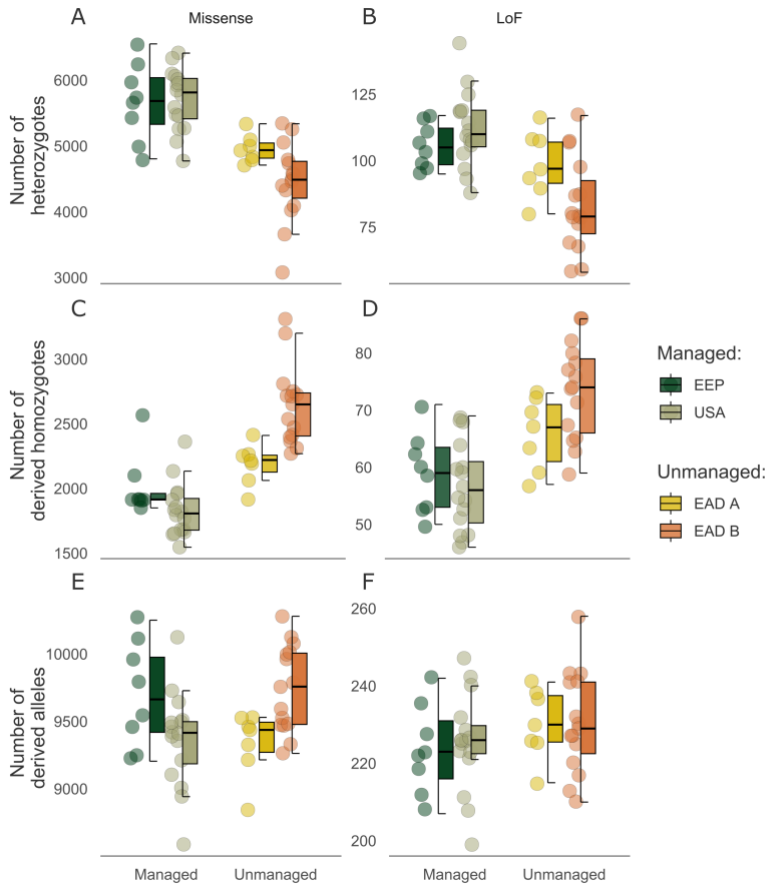

Figure 2. Deleterious load landscape across contrasting management strategies of scimitar-horned oryx based on SNPeff annotations. Distribution of the number of heterozygotes per individual (heterozygous mutation load) for missense **(A)** and loss of function mutations **(B)** across management strategies. Distribution of the number of derived homozygotes per individual (homozygous mutation load) for missense **(C)** and loss of function **(D)** mutations across management strategies. Distribution of the total number of derived alleles per individual for missense **(E)** and loss of function **(F)** mutations across management strategies.

Discussion

The scimitar-horned oryx was declared Extinct in the Wild in 2000, yet the species has persisted *ex-situ*. Understanding how management shapes the genomic landscape of inbreeding and mutation load is essential for supporting species viability. We used whole genome resequencing data to characterise runs of homozygosity and deleterious mutations in scimitar-horned oryx populations undergoing contrasting management strategies. Our study highlights the dynamics between inbreeding, mutation load and population size and has broad-reaching implications for practical conservation management.

We first demonstrated how signatures of recent population history can be identified in the genomes of present-day animals. Across *ex-situ* oryx populations, both managed and unmanaged, we observed a peak in the genome fraction in ROH between 3.12–6.25 Mb. Although it is not possible to precisely estimate the time to the most recent common ancestor (MRCA) when ROH are inferred using physical positions (53), the expected coalescent time for ROH of this size is approximately 8–16 generations ago (46). This shift in the genome fraction in ROH indicates a smaller population size around this time-period which we could reconstruct with our measures of N_e. Interestingly, assuming a generation time of around seven years (44), this directly corresponds to the mid-20th century when oryx were close to extinction in the wild and when *ex-situ* populations were founded (15, 43). We observed highly comparable patterns of N_e with the software GONe, together highlighting the power of ROH and LD-based methods for inferring the strength and timing of recent bottlenecks and for placing contemporary patterns of nucleotide diversity into a historical context.

The overall pattern of ROH lengths was qualitatively similar across populations yet the absolute proportion of the genome in ROH was considerably lower in managed than unmanaged populations for all length classes. Long ROH are indicative of recent shared ancestry of the underlying haplotypes because recombination has had little opportunity to break them up (54–57). The relative absence of long ROH therefore strongly indicates that close inbreeding is uncommon in managed populations, which work to mitigate this process. Furthermore, the smaller proportion of short ROH suggest managed populations also have lower levels of background relatedness (56, 58). Historic data on the origins of the unmanaged populations are lacking (15), yet it is not unreasonable to expect a higher level of relatedness among founder individuals compared to those of breeding programmes. Overall, these findings reveal the genomic effects of multiple generations of inbreeding, while on the other hand demonstrate how 30–40 years of *ex-situ* management has been successful at maximising the genome-wide diversity of captive populations.

We next shed light on the relationship between inbreeding, diversity and deleterious mutations by exploring how mutation load compares across management strategies. At an individual level, we show that animals from collections employing genetic management practices have a higher heterozygous mutation load for both missense and loss of function mutations than animals from unmanaged populations. Indeed, with lower levels of inbreeding – as observed in unmanaged SHO populations – a greater number of heterozygotes are expected. Furthermore, theory and simulations predict that large populations will have higher frequencies of segregating deleterious mutations (20, 59, 60). This is in part due to being masked from the effects of purifying selection in populations with larger N_e, but also by genetic

drift driving deleterious mutations to fixation in small populations. In line with this, we show that managed populations of oryx have higher nucleotide diversity and effective population sizes than unmanaged collections.

The presence of segregating deleterious alleles within insurance populations may be considered a concern for conservation management. Indeed, there has been recent debate surrounding the risks associated with sourcing individuals for restoration from large, genetically diverse populations, given the higher expected levels of strongly deleterious variation (18, 19). However, these concerns are primarily relevant for populations predicted to remain small and isolated with high levels of inbreeding. Restoration programs that follow established IUCN/SSC guidelines (39) will seek to source individuals from genetically differentiated populations, release large numbers of animals over extended time frames and maximise initial population growth rate. When *ex-situ* populations and reintroductions are managed in such a way, recessive deleterious mutations are more likely to remain partially masked as heterozygotes (16, 61–63), as we demonstrate is the case in the *ex-situ* population of oryx. Furthermore, high genetic diversity in reintroduced populations should theoretically enrich adaptive potential due to an increase in additive genetic variance (20, 64–66). The scimitar-horned oryx reintroduction has followed best practice guidelines having so far released over 250 animals over a five-year time-period, and in eight separate release batches. Consequently, the released population has now reached close to 400 individuals, with over 150 calves born in the wild. Follow-up monitoring of release herds will provide a rare opportunity to validate these efforts within the context of a large-scale reintroduction effort.

In addition to the heterozygous mutation load, we also considered how both the homozygous mutation load and the total number of derived deleterious alleles varied across populations. Several recent studies have demonstrated reductions in the relative number of highly deleterious mutations in small versus large (21–23, 26, 37) and in modern versus historical populations (24, 25), and attributed these differences to the effects of purifying selection. We observed no difference in the total number of derived loss of function alleles between unmanaged and managed populations, nor any deficit in the *Rxy* analysis. Instead, we find an increase in the homozygous mutation load for loss of function mutations in unmanaged collections. This suggests that while the total number of alleles has remained unchanged, inbreeding has increased the homozygosity of strongly deleterious variants. Consequently, unmanaged populations of oryx may carry a higher fitness cost associated with a greater burden of homozygous genotypes. In line with this, there is some anecdotal evidence of high disease prevalence and low juvenile survival in individuals originating from EAD B.

Similarly, missense mutations displayed no difference in total number between managed and unmanaged populations. However, higher values were observed in EAD B and EEP, the two populations thought to have experienced the strongest bottlenecks (see Supplementary Material). Missense mutations are more likely to be weakly deleterious with higher starting allele frequencies than loss of function mutations, and therefore are predicted to be more susceptible to accumulation by drift when population size is reduced (67–70). Critically however, the proportion of missense alleles that could be attributed to homozygous genotypes was higher in the unmanaged populations. If deleterious mutations are assumed to be approximately additive, the total number of derived alleles will provide a better estimate of genetic load and by extension be more informative of fitness (52, 71). On the other hand, if deleterious mutations are closer to recessive (2, 49–51), the homozygous mutation load will be more instructive. Indeed, inbreeding depression is ubiquitous in wild mammal populations (29, 33, 72–76), indicating that deleterious alleles have the most fitness impact when homozygous. This is exemplified by empirical work on a small, isolated population of Soay sheep with an estimated N_e of 194 that is thought to have remained in the low hundreds for thousands of years (77, 78). Despite this, long-term fitness and genomic data have revealed strong inbreeding depression caused by the expression of many weakly deleterious mutations (35, 72). With this in mind, even under a mutation load burden, *ex-situ* management to maximise genetic diversity will likely mask the effects of many deleterious alleles. Nevertheless, until we can validate molecular measures of load against empirical inbreeding depression, we must exercise caution when inferring fitness effects. Moreover, it is important to consider that even in the presence of reliable estimates of inbreeding depression, understanding the consequences on population viability is not trivial. This is because the demographic consequences of natural selection are dependent on the ecological context (33, 79–81) and therefore cannot be predicted with molecular data alone.

Ex-situ breeding and species reintroduction planning are ultimately exercises in risk management, with genetics making up only one component of a multifaceted set of considerations (39). Overall, our study provides empirical support for the value of genetic management not only for minimising inbreeding, but also for producing populations with enhanced genetic diversity for adaptation to changing environmental conditions and release back into the wild (7, 63, 82–85). Indeed, as part of the World Herd approach (86), mixing of animals from multiple collections is now a key part of the scimitar-horned oryx reintroduction strategy. While such actions can largely be informed using traditional measures of genetic variation, our study demonstrates how the application of whole genome sequencing in the context of *ex-situ* management has the power to resolve previously unknown aspects of variation. We recognise that it is impractical to consider comprehensive genomic approaches

for the genetic management of every species (87). Rather, we suggest the application of studies such as this to guide conservation breeding strategies across diverse taxa and highlight the need for further work to link molecular predictions of inbreeding depression with empirical demographic data.

Materials and methods

Sampling and sequencing

Blood (in EDTA) and tissue (in 100% ethanol) samples were collected for whole genome resequencing from 49 scimitar-horned oryx representing four *ex-situ* populations: the EEP (n = 8), USA (n = 17), EAD A (n = 9) and EAD B (n = 15). The EEP and USA are captive collections undergoing genetic management practices, while EAD A and EAD B represent collections in the United Arab Emirates with no genetic management in place (see Supplementary Methods). Total genomic DNA was extracted between one and five times per sample using the DNeasy Blood and Tissue Kit (Qiagen, Cat. No. 69504). Elutions were pooled and concentrated in an Eppendorf Concentrator Plus at 45°C and 1400 rpm until roughly 50 µl remained. Library construction was carried out using the Illumina TruSeq Nano High Throughput library preparation kit (Illumina CA, UKA). Twenty samples from across all four populations were 150 bp paired-end sequenced on an Illumina HiSeq X Ten platform at a target depth of coverage of 15X. The remaining 29 samples from three of the populations were 150 bp paired-end sequenced on an Illumina NovaSeq 6000 instrument at a target depth of coverage of 7X (Table S1).

Read processing and alignment

Sequence reads were assessed for quality using FastQC v0.11.7 (88) and trimmed for adaptor content using cutadapt v1.16 (89). Reads were then mapped to the scimitar-horned oryx reference genome assembly (*Oryx dammah* assembly version 1.1, Genbank accession number GCF_014754425.2) using BWA MEM v0.7.17 (90) with default parameters. Unmapped reads were removed from the alignment files using SAMtools v1.9 (91). Alignments were then sorted, read groups added and duplicates removed using Picard Tools v2.18.16. This resulted in a set of 49 filtered alignment files, one for each resequenced individual. To account for coverage variation in our data (92), we used SAMtools to downsample our 20 high coverage alignment files to approximately 6X, which was the average depth of coverage of our low coverage samples. All subsequent analyses were carried out on the set of alignments with comparable coverage.

Variant calling and filtering

Haplotype Caller and GenotypeGVCFs in GATK v3.8 (93) were used for joint genotyping across all samples. The resulting SNP data were filtered for biallelic sites using BCFtools v1.9 (94). To obtain a high-quality set of variants we then used VCFtools (95) to remove loci with a quality score less than 30, a mean depth of coverage less than 5 or greater than 20, a genotyping rate less than 95% and a minor allele count less than 1. We removed SNPs originating from the X chromosome or any of the unplaced scaffolds within the assembly. One individual with a high relatedness score was dropped from subsequent analysis (Figure S10, see Supplementary Methods for details). The resulting SNP dataset contained over 10 million polymorphic sites with a genotyping rate of 98%.

Population structure

We characterised population structure using genotype likelihood-based approaches in NGSadmix (96) and PCAngsd (97). Genotype likelihoods were first estimated from bam files in ANGSD (98) using the GATK model (-GL 2), inferring major and minor alleles (-doMajorMinor 1) and outputting only polymorphic sites (-SNP_pval 1e-6) with data in at least 60% of individuals (-minInd 30). We restricted this analysis to the 28 chromosome-length autosomes and included only regions with Phred quality and mapping scores over 30. Admixture proportions for the individuals in our dataset were calculated using NGSadmix. We performed admixture runs for ancestry clusters ranging from K=1–6, with ten runs for each K. The runs with the highest likelihood were plotted. The optimal K was identified based on the maximum value of the mean estimated *ln* probability of the data (99) and the Delta K method (100). Two individuals with intermediate admixture proportions between EAD A and EAD B were dropped from further analysis (Figure S2, see Supplementary Methods for details). We then performed a principal components analysis (PCA) using PCAngsd with the default parameters. Eigenvectors were computed from the covariance matrix using R.

ROH calling and individual inbreeding coefficients

We used the filtered SNP genotypes to estimate inbreeding as the proportion of the genome in runs of homozygosoty (F_{ROH}). ROH were called with a minimum length of 500 kb and a minimum of 50 SNPs using the --homozyg function in PLINK v1.9 (101) and the following parameters: --homozyg-window-snp 50 --homozyg-snp 50 --homozyg-kb 500 --homozyg-gap 1000 --homozyg-density 50 --homozyg-window-missing 5 --homozyg-window-het 3. We then calculated individual inbreeding coefficients F_{ROH} as the sum of the detected ROH lengths for each individual over the total autosomal assembly length (2.44 Gb). To ensure that our inbreeding estimates were not confounded by coverage, we compared F_{ROH} with the mean sequencing depth for each individual (Figure S11). Next, we ran linear models to explore the effect of management on inbreeding coefficients with F_{ROH} as the response variables and

management strategy as the predictor variable. We also calculated F_{ROH} based on ROH inferred using bcftools roh and the following parameters: --AF-dflt 0.16 (average minor allele frequency), -G 30 and -M 1.2 (cattle recombination rate, Mouresan *et al.* 2019). We observed a near-perfect correlation (r = 0.99) with our PLINK-based estimates (Figure S12).

ROH length distribution and recent demography

To assess recent changes in oryx population size, we characterised the fraction of the genome in ROH of seven different length classes (≥25, 12.5–25, 6.25–12.5, 3.12–6.25, 1.56–3.12, 0.78–1.56 and 0.39–0.78 Mb). Length classes (L) were calculated using the formula $L = 100/(2 \times g)$ (46), and reflect the expected lengths of ROH when the underlying haplotypes have most recent common ancestors <2, 2–4, 4–8, 8–16, 16–32, 32–64 and 64–128 generations (g) ago respectively. These generations were chosen to capture the time-period during which the wild population of oryx went extinct and captive populations were established. As there is no linkage map for the oryx, we assumed a genome-wide, homogenous, mean recombination rate of 1 cM/Mb and therefore used physical map lengths as opposed to genetic map lengths. For each time-period and for each individual, F_{ROH} was calculated as the sum of the ROH expected to coalesce in that time-period over the total autosomal assembly length (2.44 Gb). We then used measures of F_{ROH} to infer recent changes in effective population size across each population. For each time-period described above (t), we first calculated the average F_{ROH} for each population. We then estimated N_e given the following expression where $F_{ROH,t}$ corresponds to the average population inbreeding coefficient for a given time-period and t is the maximum number of generations in that time-period:

$$F_{ROH,t} = 1 - (1 - \frac{1}{2N_e})^t$$

To calculate 95% bootstrap confidence intervals around our estimates, we resampled individuals within each population with replacement 100 times, and recalculated N_e. In addition to this identity-by-descent (IBD) based approach, we implemented an LD-based method for estimating N_e using the software GONe (103). This is because under intensive inbreeding in recent generations, there are fewer ROH with long coalescent times making it more difficult to reliably estimate N_e in deeper history (see Supplementary Material for further information). As expected, historical estimates of N_e were larger using the LD-based approach, however the general pattern of population size change was the same across both methods (Figure S6).

Nucleotide diversity

Nucleotide diversity estimates were calculated for each population using ANGSD. We first estimated the unfolded site-frequency spectrum (SFS) using the -doSaf and -realSFS

commands while restricting the analysis to the 28 chromosome-length autosomes and regions with Phred quality and mapping scores over 30. Per-site pairwise nucleotide diversity estimates were then calculated using the -thetaStat command.

Identification of deleterious mutations

As most deleterious mutations are likely to be derived alleles, we first polarised our SNP genotypes as ancestral or derived using the blue wildebeest (*Connochaetes taurinus*), topi (*Damaliscus lunatus*) and hartebeest (*Alcelaphus buselaphus*) as outgroup species. Short read sequencing data from wildebeest (SRR6902709), topi (SRR6913384) and hartebeest (SRR6922939 and SRR6922940) were downloaded from NCBI and mapped to the scimitar-horned oryx reference genome using BWA MEM with the default parameters. The alignments were then merged using SAMtools. A consensus was generated by selecting the most common base from the alignment using the doFasta 2 and doCounts 1 options in ANGSD. We then used PLINK v2.0 to polarise the oryx SNPs in our VCF based on the alleles in the consensus. First, we removed SNPs from our VCF whose positions were not present in the consensus sequence. Second, we removed SNPs where the ancestral allele in the consensus matched neither allele in the VCF file. Finally, we rotated alleles so that the reference allele in our VCF matched the ancestral allele in the consensus.

To identify deleterious mutations, we predicted the functional effects of the polarised SNP variants using both SnpEff v5.0 (104) and the Variant Effect Predictor (VEP) v99.2 (105). These methods compare a set of variants to an annotation database and predict the consequence of the derived alleles on genes, transcripts and proteins. Both were run using the NCBI RefSeq scimitar-horned oryx genome annotation downloaded from: https://ftp.ncbi.nlm.nih.gov/genomes/all/annotation_releases/59534/100/GCF_014754425.2_SCBI_Odam_1.1/. For each approach, sites with warnings were removed from the VCF file and SNPs were categorised as loss of function or missense according to the classifications provided in Table S2. 643 loss of function and 34,387 missense sites were identified with SnpEff and 760 loss of function and 42,440 missense were identified by the VEP. For each dataset, we also extracted a random subset of 100,000 intergenic SNPs. For each set of SNPs, genotypes were extracted for all individuals using a combination of VCFtools and PLINK.

Mutation load landscape across management strategies

To assess how the mutation load varies across populations we used multiple measures. First, we approximated two components of mutation load; heterozygous and homozygous mutation load. The heterozygous mutation load was measured as the total number of heterozygotes

per individual for both loss of function and missense variants. The homozygous mutation load was measured as the total number of derived homozygotes per individual for both loss of function and missense variants. Next, we calculated the total number of derived alleles per individual for both missense and loss of function variants. To ensure comparisons across groups were not confounded by patterns of neutral variation, we also calculated the total number of derived alleles at a random subset of 100,000 intergenic SNPs. No significant difference between management strategies was observed (ß = 139, 95% CI = -217–495, *P* = 0.44, Figure S13). Finally, we used the *Rxy* statistic to estimate the relative frequency of loss of function and missense mutations in one population over another (106). Derived allele frequencies were calculated based on individuals from managed and unmanaged populations separately. A random subset of 100,000 intergenic SNPs was used to standardise our estimates and account for population-specific biases. To calculate 95% bootstrap confidence intervals around our estimates, we randomly resampled SNPs with replacement 100 times, and recalculated *Rxy*. To explore the effect of management on genetic load, we ran linear models with homozygous mutation load, heterozygous mutation load or the total number of derived alleles as the response variable and management strategy as the predictor variable.

Data availability

EEP samples are archived at the EAZA Biobank https://www.eaza.net/conservation/research/eaza-biobank. Whole genome resequencing data have been deposited to the European Nucleotide Archive under study accession number PRJEB37295. Analysis code is available at https://github.com/elhumble/SHO_roh_load_2022 and https://github.com/elhumble/SHO_reseq_2022.

Author contributions

EH, RO, HS, ADB, and KD conceived and designed the study. JC, RP and MAR contributed materials and funding. BP provided samples from SSP populations. EH analysed the data with input from MAS and RMG. EH wrote the manuscript. All authors commented on and helped improve the final manuscript.

Acknowledgements

We thank the EAD, EAZA and AZA institutions, along with private ranch owners in the USA, that provided samples for this study. We thank Jennifer Kaden for initial processing of samples and DNA extraction, and Edinburgh Genomics for carrying out the whole genome sequencing. SaharaConservation provided materials and wider project support. We also acknowledge Katerina Guschanski for helpful discussions and Tania Gilbert at Marwell Wildlife for information on the international studbook of scimitar-horned oryx. For the purpose of open

access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

References

1. P. J. K. McGowan, K. Traylor-Holzer, K. Leus, IUCN guidelines for determining when and how ex situ management should be used in species conservation. *Conservation Letters* **10**, 361–366 (2017).

2. M. Kimura, T. Maruyama, J. F. Crow, The mutation load in small populations. *Genetics* **48**, 1303–1312 (1963).

3. R. C. Lacy, Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology* **1**, 143–158 (1987).

4. R. Lande, Risk of population extinction from fixation of new deleterious mutations. *Evolution* **48**, 1460–1469 (1994).

5. M. Lynch, J. Conery, R. Burger, Mutation accumulation and the extinction of small populations. *The American Naturalist* **146**, 489–518 (1995).

6. D. Spielman, B. W. Brook, R. Frankham, Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15261–15264 (2004).

7. J. Ballou, R. Lacy, “Identifying genetically important individuals for management of genetic diversity in pedigreed populations” in *Analytical Methods and Strategies in Small Population Conservation*, (1995).

8. R. M. Gooley, C. J. Hogg, K. Belov, C. E. Grueber, The effects of group versus intensive housing on the retention of genetic diversity in insurance populations. *BMC Zool* **3**, 2 (2018).

9. D. Wildt, *et al.*, Breeding centers, private ranches, and genomics for creating sustainable wildlife populations. *BioScience* **69**, 928–943 (2019).

10. R. M. Gooley, *et al.*, Comparison of genomic diversity and structure of sable antelope (*Hippotragus niger*) in zoos, conservation centers, and private ranches in North America. *Evolutionary Applications* **13**, 2143–2154 (2020).

11. M. J. Lott, B. R. Wright, L. F. Kemp, R. N. Johnson, C. J. Hogg, Genetic management of captive and reintroduced bilby populations. *The Journal of Wildlife Management* **84**, 20–32 (2020).

12. D. J. Pritchard, J. E. Fa, S. Oldfield, S. R. Harrop, Bring the captive closer to the wild: redefining the role of *ex situ* conservation. *Oryx* **46**, 18–23 (2012).

13. K. H. Redford, D. B. Jensen, J. J. Breheny, Integrating the captive and the wild. *Science* **338**, 1157–1158 (2012).

14. K. R. Schwartz, E. C. M. Parsons, L. Rockwood, T. C. Wood, Integrating in-situ and ex-situ data management processes for biodiversity conservation. *Frontiers in Ecology and Evolution* **5** (2017).

15. R. Ogden, *et al.*, Benefits and pitfalls of captive conservation genetic management: Evaluating diversity in scimitar-horned oryx to support reintroduction planning. *Biological Conservation* **241**, 108244 (2020).

16. K. Ralls, P. Sunnucks, R. C. Lacy, R. Frankham, 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 (2020).

17. J. A. DeWoody, A. M. Harder, S. Mathur, J. R. Willoughby, The long-standing significance of genetic diversity in conservation. *Molecular Ecology* **30**, 4147–4154 (2021).

18. C. C. Kyriazis, R. K. Wayne, K. E. Lohmueller, Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evolution Letters* **5**, 33–47 (2021).

19. J. C. Teixeira, C. D. Huber, The inflated significance of neutral genetic diversity in conservation genetics. *Proceedings of the National Academy of Sciences* **118**, e2015096118 (2021).

20. M. Kardos, *et al.*, The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences* **118**, e2104642118 (2021).

21. Y. Xue, *et al.*, Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348**, 242–245 (2015).

22. J. A. Robinson, C. Brown, B. Y. Kim, K. E. Lohmueller, R. K. Wayne, Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. *Curr Biol* **28**, 3487–3494 (2018).

23. J. A. Robinson, *et al.*, Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Science Advances* **5**, eaau0757 (2019).

24. T. van der Valk, D. Díez-del-Molino, T. Marques-Bonet, K. Guschanski, L. Dalén, Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. *Current Biology* **29**, 165-170.e6 (2019).

25. J. von Seth, *et al.*, Genomic insights into the conservation status of the world's last remaining Sumatran rhinoceros populations. *Nat Commun* **12**, 2393 (2021).

26. N. Dussex, *et al.*, Population genomics of the critically endangered kākāpō. *Cell Genomics* **1**, 100002 (2021).

27. G. Caughley, Directions in conservation biology. *Journal of Animal Ecology* **63**, 215–244 (1994).

28. P. W. Hedrick, R. C. Lacy, F. W. Allendorf, M. E. Soulé, Directions in conservation biology: comments on Caughley. *Conservation Biology* **10**, 1312–1320 (1996).

29. L. Keller, Inbreeding effects in wild populations. *Trends in Ecology & Evolution* **17**, 230–241 (2002).

30. R. Lande, S. Shannon, The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution* **50**, 434–437 (1996).

31. I. Saccheri, *et al.*, Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**, 491–494 (1998).

32. D. H. Reed, R. Frankham, Correlation between fitness and genetic diversity. *Conservation Biology* **17**, 230–237 (2003).

33. C. Bozzuto, I. Biebach, S. Muff, A. R. Ives, L. F. Keller, Inbreeding reduces long-term growth of Alpine ibex populations. *Nat Ecol Evol* **3**, 1359–1364 (2019).

34. C. Grossen, F. Guillaume, L. F. Keller, D. Croll, Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nature Communications* **11**, 1–12 (2020).

35. M. A. Stoffel, S. E. Johnston, J. G. Pilkington, J. M. Pemberton, Mutation load decreases with haplotype age in wild Soay sheep. *Evolution Letters* **5**, 187–195 (2021).

36. S. Mathur, J. A. DeWoody, Genetic load has potential in large populations but is realized in small inbred populations. *Evol Appl* **14**, 1540–1557 (2021).

37. A. Khan, *et al.*, Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. *Proceedings of the National Academy of Sciences* **118**, e2023018118 (2021).

38. D. Kleinman-Ruiz, *et al.*, Purging of deleterious burden in the endangered Iberian lynx. *Proceedings of the National Academy of Sciences* **119**, e2110614119 (2022).

39. IUCN/SSC, “Guidelines for reintroductions and other conservation translocations. Version 1.0” (2013).

40. S. Edmands, Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology* **16**, 463–475 (2007).

41. A. R. Weeks, *et al.*, Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications* **4**, 709–725 (2011).

42. C. J. Hogg, *et al.*, Preserving the demographic and genetic integrity of a single source population during multiple translocations. *Biological Conservation* **241**, 108318 (2020).

43. T. Woodfine, T. Gilbert, “The Fall and Rise of the Scimitar-Horned Oryx” in *Antelope Conservation*, (John Wiley & Sons, Ltd, 2016), pp. 280–296.

44. T. Gilbert, *International studbook for the scimitar-horned oryx* Oryx dammah, Seventeenth edition (Marwell Wildlife, 2022).

45. E. Humble, *et al.*, Chromosomal-level genome assembly of the scimitar-horned oryx: Insights into diversity and demography of a species extinct in the wild. *Molecular Ecology Resources* **20**, 1668–1681 (2020).

46. E. A. Thompson, Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics* **194**, 301–326 (2013).

47. I. M. MacLeod, D. M. Larkin, H. A. Lewin, B. J. Hayes, M. E. Goddard, Inferring demography from runs of homozygosity in whole-genome sequence, with correction for sequence errors. *Molecular Biology and Evolution* **30**, 2209–2223 (2013).

48. S. R. Browning, B. L. Browning, Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *The American Journal of Human Genetics* **97**, 404–418 (2015).

49. B. Charlesworth, D. Charlesworth, The genetic basis of inbreeding depression. *Genetics Research* **74**, 329–340 (1999).

50. T. Mukai, S. I. Chigusa, L. E. Mettler, J. F. Crow, Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**, 335–355 (1972).

51. S. Peischl, L. Excoffier, Expansion load: recessive mutations and the role of standing genetic variation. *Molecular Ecology* **24**, 2084–2094 (2015).

52. Y. B. Simons, G. Sella, The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. *Current Opinion in Genetics & Development* **41**, 150–158 (2016).

53. M. Kardos, A. Qvarnström, H. Ellegren, Inferring individual inbreeding and demographic history from segments of identity by descent in *Ficedula* flycatcher genome sequences. *Genetics* **205**, 1319–1334 (2017).

54. R. McQuillan, *et al.*, Runs of homozygosity in European populations. *The American Journal of Human Genetics* **83**, 359–372 (2008).

55. M. Kirin, *et al.*, Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* **5**, e13996 (2010).

56. T. J. Pemberton, *et al.*, Genomic patterns of homozygosity in worldwide human populations. *The American Journal of Human Genetics* **91**, 275–292 (2012).

57. M. Kardos, G. Luikart, F. W. Allendorf, Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity* **115**, 63–72 (2015).

58. H. Ringbauer, J. Novembre, M. Steinrücken, Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nat Commun* **12**, 5425 (2021).

59. D. Charlesworth, B. Charlesworth, Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology & Systematics* **18**, 237–268 (1987).

60. M. C. Whitlock, Selection, load and inbreeding depression in a large metapopulation. *Genetics* **160**, 1191–1202 (2002).

61. L. N. Tracy, G. P. Wallis, M. G. Efford, I. G. Jamieson, Preserving genetic diversity in threatened species reintroductions: how many individuals should be released? *Animal Conservation* **14**, 439–446 (2011).

62. A. R. Whiteley, S. W. Fitzpatrick, W. C. Funk, D. A. Tallmon, Genetic rescue to the rescue. *Trends in Ecology & Evolution* **30**, 42–49 (2015).

63. R. Frankham, Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology* **24**, 2610–2618 (2015).

64. P. de Villemereuil, *et al.*, Little adaptive potential in a threatened passerine bird. *Current Biology* **29**, 889-894.e3 (2019).

65. R. Frankham, *et al.*, Do population size bottlenecks reduce evolutionary potential? *Animal Conservation* **2**, 255–260 (1999).

66. Y. Willi, J. Van Buskirk, A. A. Hoffmann, Limits to the adaptive potential of small populations. *Annual Review of Ecology, Evolution, and Systematics* **37**, 433–458 (2006).

67. J. Wang, W. G. Hill, Effect of selection against deleterious mutations on the decline in heterozygosity at neutral loci in closely inbreeding populations. *Genetics* **153**, 1475–1489 (1999).

68. A. García-Dorado, Understanding and predicting the fitness decline of shrunk populations: inbreeding, purging, mutation, and standard selection. *Genetics* **190**, 1461–1476 (2012).

69. B. Charlesworth, Mutational load, inbreeding depression and heterosis in subdivided populations. *Molecular Ecology* **27**, 4991–5003 (2018).

70. R. C. Lacy, J. D. Ballou, Effectiveness of selection in reducing the genetic load in populations of *Peromyscus polionotus* during generations of inbreeding. *Evolution* **52**, 900–909 (1998).

71. Y. B. Simons, M. C. Turchin, J. K. Pritchard, G. Sella, The deleterious mutation load is insensitive to recent population history. *Nat Genet* **46**, 220–224 (2014).

72. M. A. Stoffel, S. E. Johnston, J. G. Pilkington, J. M. Pemberton, Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. *Nat Commun* **12**, 2972 (2021).

73. J. Huisman, L. E. B. Kruuk, P. A. Ellis, T. Clutton-Brock, J. M. Pemberton, Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 3585–3590 (2016).

74. P. Crnokrak, D. A. Roff, Inbreeding depression in the wild. *Heredity* **83**, 260–270 (1999).

75. P. W. Hedrick, A. Garcia-Dorado, Understanding inbreeding depression, purging, and genetic rescue. *Trends in Ecology & Evolution* **31**, 940–952 (2016).

76. K. Ralls, J. D. Ballou, A. Templeton, Estimates of lethal equivalents and the cost of inbreeding in mammals. *Conservation Biology* **2**, 185–193 (1988).

77. T. H. Clutton-Brock, J. M. Pemberton, Eds., *Soay sheep: dynamics and selection in an island population* (Cambridge University Press, 2004).

78. J. W. Kijas, *et al.*, Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology* **10**, e1001258 (2012).

79. D. A. Bell, R. P. Kovach, Z. L. Robinson, A. R. Whiteley, T. E. Reed, The ecological causes and consequences of hard and soft selection. *Ecology Letters* **24**, 1505–1521 (2021).

80. E. I. Svensson, T. Connallon, How frequency-dependent selection affects population fitness, maladaptation and evolutionary rescue. *Evolutionary Applications* **12**, 1243–1258 (2019).

81. E. Day, H. Kokko, Relaxed selection when you least expect it: why declining bird populations might fail to respond to phenological mismatches. *Oikos* **124**, 62–68 (2015).

82. M. E. Montgomery, *et al.*, Minimizing kinship in captive breeding programs. *Zoo Biology* **16**, 377–389 (1997).

83. M. Pickup, D. L. Field, D. M. Rowell, A. G. Young, Source population characteristics affect heterosis following genetic rescue of fragmented plant populations. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20122058 (2013).

84. R. Frankham, *et al.*, *Genetic management of fragmented animal and plant populations* (Oxford University Press, 2017) https:/doi.org/10.1093/oso/9780198783398.001.0001 (March 21, 2022).

85. M. Ørsted, A. A. Hoffmann, E. Sverrisdóttir, K. L. Nielsen, T. N. Kristensen, Genomic variation predicts adaptive evolutionary responses better than population bottleneck history. *PLOS Genetics* **15**, e1008205 (2019).

86. K. Mertes, *et al.*, Management background and release conditions structure post-release movements in reintroduced ungulates. *Front. Ecol. Evol.* **7** (2019).

87. M. Kardos, A. B. A. Shafer, The peril of gene-targeted conservation. *Trends in Ecology & Evolution* **33**, 827–839 (2018).

88. S. Andrews, FastQC: A quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).

89. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

90. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* (2013).

91. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

92. R. N. Lou, N. O. Therkildsen, Batch effects in population genomic studies with low-coverage whole genome sequencing data: Causes, detection and mitigation. *Molecular Ecology Resources* **22**, 1678–1692 (2022).

93. G. A. Van der Auwera, *et al.*, From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11.10.1-11.10.33 (2013).

94. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

95. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

96. L. Skotte, T. S. Korneliussen, A. Albrechtsen, Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).

97. J. Meisner, A. Albrechtsen, Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731 (2018).

98. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).

99. J. K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).

100. G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**, 2611–2620 (2005).

101. S. Purcell, *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).

102. E. F. Mouresan, *et al.*, Mapping recombination rate on the autosomal chromosomes based on the persistency of linkage disequilibrium phase among autochthonous beef cattle populations in spain. *Frontiers in Genetics* **10** (2019).

103. E. Santiago, *et al.*, Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. *Molecular Biology and Evolution* **37**, 3642–3653 (2020).

104. P. Cingolani, *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).

105. W. McLaren, *et al.*, The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122 (2016).

106. R. Do, *et al.*, No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat Genet* **47**, 126–131 (2015).