

Genomics-informed captive breeding can reduce inbreeding depression and the genetic load in zoo populations

Samuel A. Speak^{1,2,3}  | Thomas Birley¹ | Chiara Bortoluzzi^{4,5} | Matthew D. Clark^{2,6} | Lawrence Percival-Alwyn⁷ | Hernán E. Morales⁸  | Cock van Oosterhout¹ 

¹School of Environmental Sciences,
University of East Anglia, Norwich, UK

²Natural History Museum, London, UK

³North of England Zoological Society,
Chester Zoo, Chester, UK

⁴University of Cambridge, Cambridge, UK

⁵Wellcome Sanger Institute, Hinxton, UK

⁶Earlham Institute, Norwich, UK

⁷NIAB, Cambridge, UK

⁸Globe Institute, University of
Copenhagen, Copenhagen, Denmark

Correspondence

Samuel A. Speak and Cock van
Oosterhout, University of East Anglia,
Norwich Research Park, Norwich, NR4
7TJ, UK.

Email: s.speak@uea.ac.uk and
c.van-oosterhout@uea.ac.uk

Funding information

Royal Society, Grant/Award Number:
ICA/R1/201194; University of East
Anglia, Grant/Award Number: RTSG;
100162318RA1; Wellcome Trust, Grant/
Award Number: WT207492; Natural
Environment Research Council, Grant/
Award Number: T209447; European
Molecular Biology Organization, Grant/
Award Number: grant 1111-2018;
H2020 Marie Skłodowska-Curie Actions,
Grant/Award Number: grant 840519;
Biotechnology and Biological Sciences
Research Council, Grant/Award Number:
BB/M011216/1, BB/S019669/1, BBS/
E/T/000PR5885, BBS/E/T/000PR6193,
BBS/E/T/000PR9816 and BBS/E/
T/000PR9818

Handling Editor: Luke Browne

Abstract

Zoo populations of threatened species are a valuable resource for the restoration of wild populations. However, their small effective population size poses a risk to long-term viability, especially in species with high genetic load. Recent bioinformatic developments can identify harmful genetic variants in genome data. Here, we advance this approach, analysing the genetic load in the threatened pink pigeon (*Nesoenas mayeri*). We lifted the mutation-impact scores that had been calculated for the chicken (*Gallus gallus*) to estimate the genetic load in six pink pigeons. Additionally, we perform in silico crossings to predict the genetic load and realized load of potential offspring. We thus identify the optimal mate pairs that are theoretically expected to produce offspring with the least inbreeding depression. We use computer simulations to show how genomics-informed conservation can reduce the genetic load whilst reducing the loss of genome-wide diversity. Genomics-informed management is likely to become instrumental in maintaining the long-term viability of zoo populations.

KEYWORDS

CADD, captive populations, genetic load, genomics-informed conservation, inbreeding depression, *Nesoenas mayeri*

1 | INTRODUCTION

More than 28% of the 157,190 species on the Red List of the International Union for Conservation of Nature (IUCN) are threatened with extinction (IUCN, 2023). A relatively small subset of these species is kept as “insurance populations” in zoos (Gilbert et al., 2017). However, given their often-small effective population size, the long-term viability of captive-bred populations is not guaranteed, and many show signs of inbreeding depression (Boakes et al., 2007). Mutations introduce new genetic variants into the genome, many of which are harmful (see figure 1 in de Jong et al. (2024) for an illustration of the impact of mutations according to the leading models of molecular evolution). These harmful genetic variants can reduce fitness, and this potential reduction in fitness is known as the genetic load (Bertorelle et al., 2022). High genetic load can compromise population viability and recovery potential of species, especially if they experience a recent population size decline (Jackson et al., 2022; Sachdeva et al., 2022). In declining populations, drift and inbreeding increase the frequency of homozygous harmful variants, which results in inbreeding depression. It can take many generations before these harmful genetic variants become homozygous, a phenomenon known as the ‘drift debt’ (Pinto et al., 2023). Consequently, the long-term viability of many zoo populations could be at risk, despite individuals and populations thriving now. The analysis of the genetic load enables us to assess this indirect threat of inbreeding on present and future population viability.

In the past 50 years, conservation geneticists have focused on maintaining genetic variation (DeWoody et al., 2021; García-Dorado & Caballero, 2021; Kardos et al., 2021) as genome-wide diversity generally correlates positively with fitness and adaptive potential (Charlesworth, 2009; Harrisson et al., 2014; Mathur et al., 2023, but see Wood et al., 2016). Recently, the Group on Earth Observations Biodiversity Observation Network (GEO BON) developed Essential Biodiversity Variables (EBVs) to assess spatiotemporal variation in biodiversity, and proposed four genetic EBVs: genetic diversity, genetic differentiation, inbreeding and effective population size (N_e) (Hoban et al., 2022). Notably, risks posed by genetic load are generally not considered a conservation priority (van Oosterhout, 2020), which may be an oversight. Inbreeding depression in declining populations is a function of both the rate of inbreeding and the genetic load of recessive deleterious mutations that is present at heterozygous loci. This part of the genetic load is known as the inbreeding load or masked load (Bertorelle et al., 2022). Inbreeding exposes the harmful effects of these mutations by increasing homozygosity, converting the masked load into a realized load. Recent advances in genomics and bioinformatics now allow us to study the size and composition of the genetic load without necessarily exposing the deleterious effects of mutations or harming the fitness of individuals.

Leveraging the extensive genomic research on human and model animals enables us to estimate the potential fitness

impact of mutations in species of conservation concern (Bertorelle et al., 2022). The fitness impact of deleterious alleles can be estimated by the Combined Annotation-Dependent Depletion (CADD) framework (Rentzsch et al., 2019). Initially developed in humans (Kircher et al., 2014), the CADD framework has been successfully applied to other model organisms, including mouse (Groß et al., 2018), pig (Groß, Derks, et al., 2020) and chicken (Groß, Bortoluzzi, et al., 2020). CADD scores rank genetic variants such as single-nucleotide polymorphisms (SNPs) and insertions and deletions (indels) throughout the genome. In the human genome, the CADD scores are ranked for every possible mutation (~8.6 billion single-nucleotide variants). The highest scoring 10% of variants (i.e. mutations) are allocated CADD scores of 10 and higher, the highest 1% of mutations receive a score over 20, the highest 0.1% a score over 30, etc. (Rentzsch et al., 2019). This analysis integrates surrounding sequence context, gene model annotation, evolutionary constraints (e.g. GERP scores), epigenetic measurements and functional predictions into CADD scores. The CADD framework was employed to investigate conserved elements in the chicken Combined Annotation-Dependent Depletion (chCADD) (Groß, Bortoluzzi, et al., 2020), which has helped identify regions within the chicken genome associated with known genetic disorders reported in the Online Mendelian Inheritance in Animals (OMIA). Therefore, by identifying deleterious alleles, the CADD framework can estimate the genetic load within an individual's genome.

Presently, we cannot translate the impact scores of mutations such as CADD scores into fitness effects. Nevertheless, we can calculate CADD scores for all putative deleterious mutations present in an individual's genome and compare this proxy of the genetic load between individuals. Similarly, we can estimate the proportion of genetic load expressed as realized load and the proportion whose fitness effects remain masked as an inbreeding load or masked load (Bertorelle et al., 2022). The realized load comprises the genetic load that reduces fitness when the harmful effect of the mutations comes to light. Inbreeding increases the realized load because more deleterious mutations become fully expressed as homozygous. By minimizing realized load, conservation managers can reduce the severity of inbreeding depression. This could be particularly useful in manipulating breeding pairs of captive populations in which individuals are related, to improve the fitness of offspring.

A considerable amount of genetic variation codes for polygenic or quantitative traits. Mutations that affect the value of a quantitative trait (e.g. body size) can be harmful or beneficial depending on whether it brings the trait value closer to the optimum. In contrast, unconditionally deleterious mutations are harmful irrespective of genetic background or environmental conditions. Mutations in ultraconserved elements (UCEs) are likely to be unconditionally deleterious (Silla et al., 2014), thereby contributing substantially to the genetic load. UCEs were originally defined as areas of the genome of 200bp which were 100% phylogenetically conserved across diverged taxa (Bejerano et al., 2004). In the present study,

we use the more recent definition of a UCE, that is, a region that shows 80% or more conservation across a nucleotide sequence of 100bp (Faircloth et al., 2012). Their high level of sequence conservation is thought to be maintained by strong purifying selection (Lee & Venkatesh, 2013). Some polymorphisms in UCEs are associated with genetic diseases or phenotypic traits (Habic et al., 2019), whilst others are linked to enhancers in early development in both mammals (Visel et al., 2008) and flies (Warnefors et al., 2016). Given their high level of phylogenetic conservation, we can build on the knowledge of model organisms and use a comparative genomic approach to obtain a proxy for the genetic load. Studying UCEs in reference genomes allows for between-species comparisons of the proxies of genetic load, realized load and masked load. Additionally, analysis of genetic load at UCEs shows promise for captive breeding and conservation management of zoo populations.

Here, we conduct a proof-of-concept study to demonstrate the utility of genomics-informed breeding in the conservation management of captive populations. We quantify the genetic load of six pink pigeon individuals using chCADD scores assigned to single-nucleotide variants in the UCEs derived from the chicken genome. We show that genetic load components can be estimated using CADD scores calculated on a phylogenetic closely related species and cross-mapped to the annotation of the pink pigeon, our focal species. We also calculate realized load and genetic load of potential future offspring of all possible crosses. Finally, we employ computer simulations to demonstrate the potential of genomics-informed conservation, showing how it can help reduce inbreeding depression and maximize the long-term viability of zoo populations.

2 | MATERIALS AND METHODS

2.1 | Study species

Six pink pigeon (*Nesoenas mayeri*) individuals from the captive-bred population of Jersey Zoo ($n=4$) and Bristol Zoo ($n=2$) were genome sequenced. Birds shared common ancestry within the last 3–6 generations (Figure S1) and produce offspring that are moderately to highly inbred (inbreeding coefficient, $F=0.064$ – 0.346) (Table S1), which is typical of many zoo populations (Boakes et al., 2007) (see Supporting Information for further details).

2.2 | Genome sequencing and bioinformatics

DNA was extracted from blood, using Qiagen MagAttract; linked read library preparation was 10× Genomics Chromium technology, which were then sequenced on an Illumina HiSeq X with 2×150 bp reads; mean depth of the six samples was between 16.51 and 19.41 (Ryan, 2021). The sequencing read data was mapped to a previously

generated pink pigeon reference genome (Albeshr, 2016). The variant calls were used to create a per-SNP pink pigeon CADD (ppCADD) score calculated for the UCEs of each individual's genome (Figure 1). A Snakemake pipeline (Mölder et al., 2021) allowing for reproduction of this approach can be found on GitHub (<https://github.com/saspeak/LoadLift>).

Previously published tetrapod ultraconserved element (UCE) probes based on the chicken reference genome (Warren et al., 2017) and the Tibetan ground-jay (*Pseudopodoces humilis*) (Faircloth et al., 2012) were used to harvest UCEs from the pink pigeon reference genome, using the Phyluce workflow (Faircloth, 2016). A chain file was created for annotation conversion, and the CADD scores of the chicken genome (Groß, Bortoluzzi, et al., 2020) were cross-mapped to the reference pigeon genome using CrossMap.py (Zhao et al., 2014). CADD scores were filtered to remove non-scoring and fixed sites (a non-scoring site is a site that is homozygous for the chicken reference allele with a CADD score equal to zero). Genotypes of each locus within the UCEs and flanking regions were assessed to calculate the genetic load components. Individuals' genetic load, realized load and masked load were calculated using the following formulas (Bertorelle et al., 2022):

$$\text{Genetic load (individual } k) = \sum_{i=1}^{L(\text{hom})} s_i + \sum_{j=1}^{L(\text{het})} 0.5s_j \quad (1)$$

$$\text{Realized load (individual } k) = \sum_{i=1}^{L(\text{hom})} s_i + \sum_{j=1}^{L(\text{het})} h_j s_j \quad (2)$$

$$\text{Masked load (individual } k) = \sum_{j=1}^{L(\text{het})} (0.5 - h_j) s_j \quad (3)$$

Here, s_i (and s_j) is the ppCADD score at locus i (and j), and they are summed across all homozygous (or heterozygous) loci at the UCEs of individual k . In the computer simulations (see below), s and h stand for the selection and dominance coefficients, and the fitness impact of the load can be expressed in lethal equivalents (Bertorelle et al., 2022). For simplicity, the dominance coefficient (h_j) is assumed to be $h_j=0.1$. It should be noted that part of the realized load comprises heterozygous mutations that are assumed to be partially dominant. Inbreeding coefficients (F_{RoH}) of the six pink pigeons were calculated using runs of homozygosity (RoH) with bcftools roh (Narasimhan et al., 2016). For further details, see Supporting Information.

2.3 | Computer simulations of breeding regimes

We conducted computer simulations in SLiM3 (Haller & Messer, 2019) to examine the impact of four breeding regimes on genetic and realized load, neutral genetic diversity and fitness. In

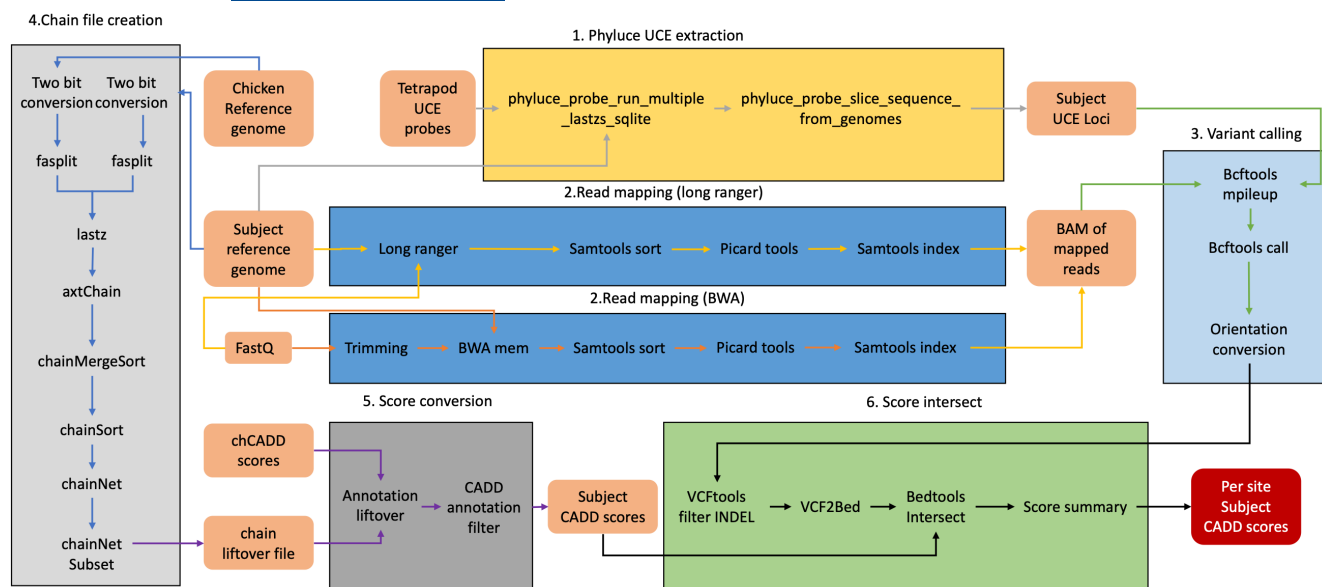


FIGURE 1 The pipeline for the creation of per Single-Nucleotide Polymorphism (SNP) pink pigeon Combined Annotation-Dependent Depletion (ppCADD) scores from raw reads of individual pink pigeons. The Snakemake (Mölder et al., 2021) pipeline uses as input the sequencing reads of the subject individuals, the subject species reference genome, and the CADD scores and reference genome of a model species (i.e. chicken, chCADD scores (Groß, Bortoluzzi, et al., 2020) and the Galgal6 reference genome (Warren et al., 2017)). The pipeline is separated into six sections, corresponding to sections of the pipeline (<https://github.com/saspeak/LoadLift>). (1) (Yellow) Extraction of UCEs from the reference genome using Phyluce. (2) (Dark Blue) Mapping the sequencing reads for individuals to the reference genome indicating two parallel approaches for 10x Chromium read data (used in this paper) and for Illumina read data. (3) (Light Blue) Variant calling for SNPs within the UCEs. (4) (Light grey) Creation of a chain file for the conversion of annotation from the chicken genome. (5) (Dark Grey) chCADD scores conversion to pink pigeon (subject species) annotation. (6) (Green) Intersection of BED files and UCE sites to output per site ppCADD (subject species) scores (Red).

the “Minimize load” regime, we examined whether mate pair selection can reduce the realized load of the offspring and minimize inbreeding depression. However, purifying selection against the genetic load can reduce genetic diversity (Cvijović et al., 2018) and result in the fixation of mildly deleterious mutations (Chen et al., 2020). To address this concern, we explored the impact of reducing relatedness of parents, and this was simulated in the “Minimize relatedness” regime. For this regime, only one male and female per brood are selected to be mates. Thus, relatedness within the population was minimized as all families were represented by exactly two individuals in the next round of breeders. Additionally, we simulated a regime that aimed to minimize realized load of the offspring whilst maintaining genetic diversity, the “Minimize load and relatedness” regime. Here, exactly one male and one female from each family were selected to mate with an optimal partner from another family, to minimize realized load of their offspring. This regime uses pairwise genetic similarity values to minimize the relatedness of parents. Using exactly two offspring per family as the breeders in the next generation minimizes genetic drift and maximizes the effective population size. Hence, this regime reduces relatedness in future generations. Furthermore, by crossing these individuals with an optimal (i.e. least-related) mating partner, this regime also minimized the realized load.

Finally, we simulated random mating “Random mating” regime. In this regime, mates are selected at random except that mates who share at least one grandparent are rejected and replaced. Mates with common grandparents are avoided in all four regimes as zookeepers would likely refrain from such crosses. In each regime, we randomly sampled 20 monogamous pairs of males and females and allowed each pair to produce 64 offspring during their lifetime (i.e. one generation). Although this may seem a relatively high fecundity, the total lifetime reproductive potential of many species significantly exceeds this number. We ran 100 replicates for each regime for 50 generations. Further details about the breeding regimes and SLiM model are given in [Supporting Information](#).

3 | RESULTS

3.1 | Distribution of UCEs and CADD scores

The 4976 UCEs along the 34 chromosomes of the chicken reference genome are not evenly distributed (Figure 2a), 15 chromosomes were significantly depleted for UCEs, whilst 9 chromosomes were significantly enriched for UCEs (Table S2). Figure 2b shows the

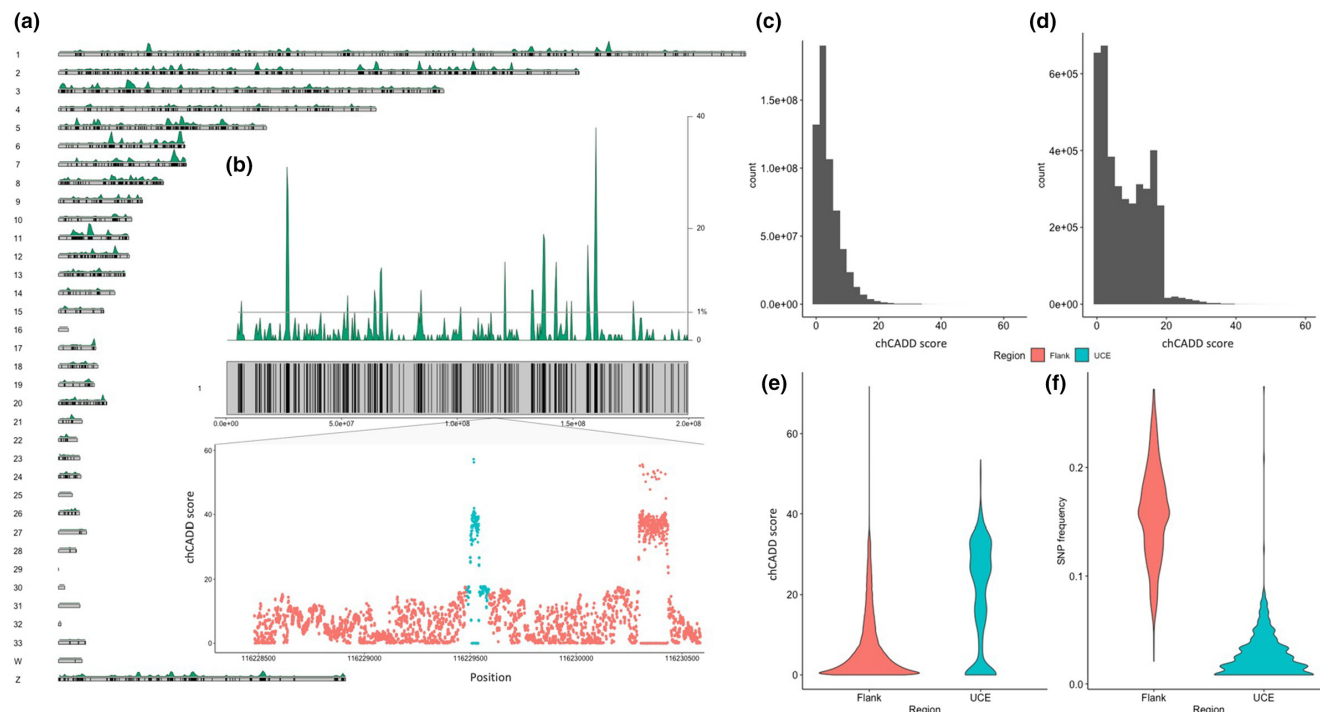


FIGURE 2 Distribution of ultraconserved elements (UCEs) and their mutation-impact scores (CADD scores). (a) Karyotype plot of the chicken genome with the distribution of UCEs (black bars) and density of UCEs (green peaks). (b) Karyotype plot of chicken Chromosome 1 showing the distribution of UCE-dense regions. Green peaks above the 1% horizontal line are significantly enriched for UCEs ($p < .01$). At the bottom of Panel B, zoomed in at a single UCE and its 2000bp flanking regions (i.e. UCE2729), the CADD scores of every possible substitution at each site. The UCE is shown in blue. The CADD scores in flanking regions are shown in red. Distribution of all CADD scores for (c) the entire Chromosome 1 of the chicken genome, and (d) 620 UCEs in Chromosome 1 and their 2000bp flanking regions. (e) The CADD score distribution of the flanking regions and the UCEs within the six pink pigeon genomes. (f) SNP frequency at flanking regions and the UCEs (see main text for test results).

distribution of all chCADD scores along a single UCE (UCE-2729) and its 2000bp flanking region on Chromosome 1. The chCADD scores in the flanking region are lower than those within the UCE, except for a potential coding region (e.g. position 116,230,300–116,230,450 in Figure 2b). Protein-coding genes are typified by a combination of high chCADD scores (representing the first and second codon position substitutions) and low chCADD scores (third codon position substitutions).

Figure 2c shows the distribution of chCADD scores along Chromosome 1 of the chicken genome. Most chCADD scores fall below 10, which per definition represent 90% of all scores. The right-hand tail represents a few high chCADD scores of highly deleterious mutations. In contrast, the UCEs and their flanking regions in Chromosome 1 have a bimodal distribution of chCADD scores, with a second peak of chCADD scores ranging between 17 and 18 (Figure 2d). These chCADD scores represent the worst, ~2% of all possible substitutions in the genome. The median chCADD score of UCEs is significantly higher than that of the flanking regions (Mann–Whitney test $W=4,541,885,925$, p -value $< .0001$). The frequency

of derived mutations is significantly lower at UCEs compared to that at the flanking regions (Mann–Whitney test $W=13,010,970$, p -value $< .0001$), which is consistent with the effect of purifying selection.

3.2 | Genetic load components and kinship

We analysed the genetic load in the hypothetical offspring of our six pink pigeons. This is calculated by theoretically crossing all possible combinations of individuals assuming Mendelian segregation ratios. When the kinship coefficient, between two individuals, is higher, homozygosity of their offspring increases (Figure 3), which elevates the offspring's realized load and reduces the masked load (Figure 3). Optimal mate pairing can significantly reduce the realized load of the offspring compared to random mating ($R^2=0.258$, $F_{1,13}=8.32$, $p=.00918$).

Next, we performed an analysis to identify optimal crosses to minimize genetic load (Figure 4). CADD scores for the potential

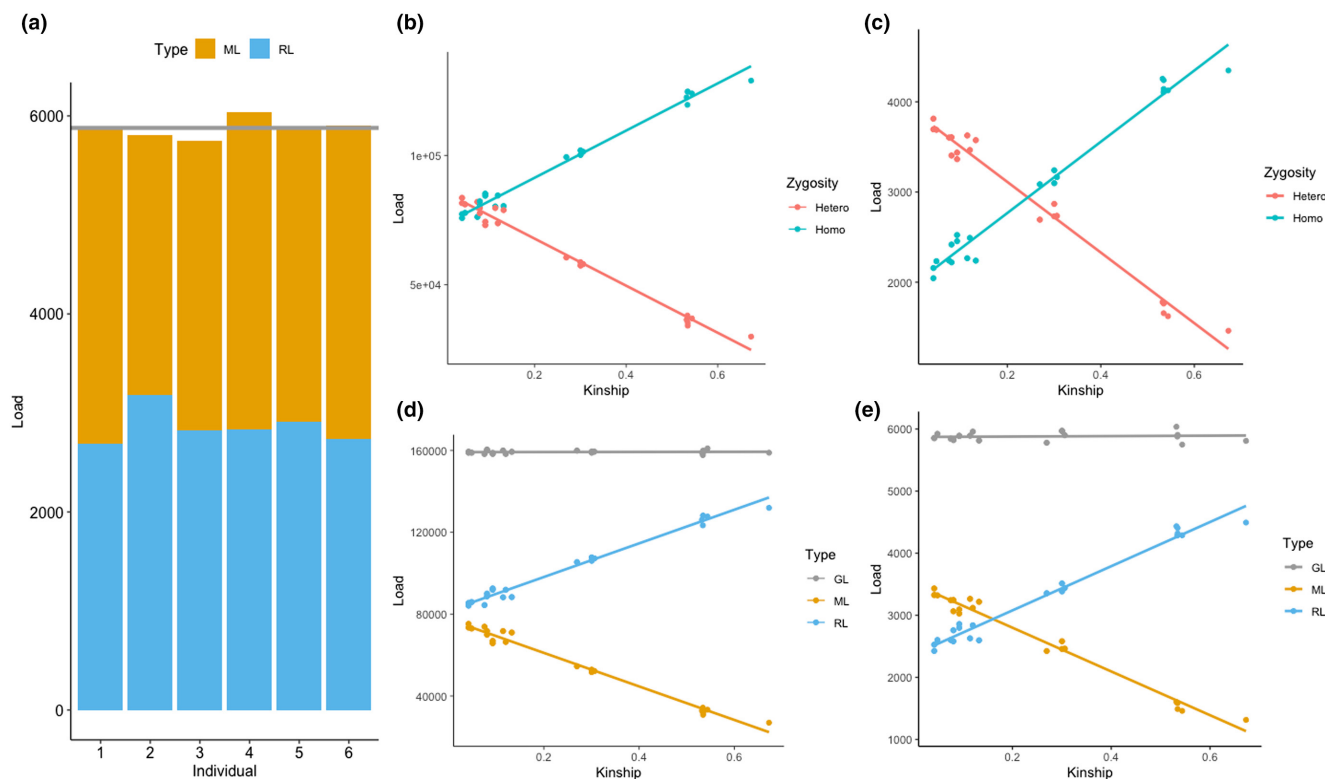


FIGURE 3 The composition of the genetic load in six pink pigeon individuals and their hypothetical offspring. (a) The total realized load (Blue) and masked load (Orange) in each of the six pink pigeon individuals within their UCEs. (b and c) The realized load at heterozygous loci (Red) and homozygous loci (Teal) of the offspring are shown for the total region (b) and UCEs only (c). (d and e) The genetic load (Grey), realized load (Blue) and masked load (Orange) of the hypothetical offspring of all possible crosses between the six pink pigeons for the total region (d) and the UCE only (e).

offspring of each cross, including self-mating, are provided for the genetic load (Table S3), realized load (Table S4) and masked load (Table S5). Figure 4a shows the average genetic load of potential offspring. In essence, these are the deleterious mutations that offspring are predicted to inherit from both parents. This can be distinguished for captive breeding managers with blue tiles representing offspring with low genetic load and red tiles offspring with high genetic load. The genetic load is lowest in the offspring from a cross between individuals 2 and 3.

To predict the degree of inbreeding depression, the realized load of the offspring of different crosses was calculated. Blue tiles in the correlogram in Figure 4b show the realized load of the offspring of the optimal crosses. The realized load of these offspring is 7.4% less than that of offspring of random crosses (Figure 4e), and these offspring are predicted to show less inbreeding depression. Note that the offspring from the 2 × 3 cross with the lowest genetic load possesses a relatively high realized load. Individuals 2 and 3 were closely related (Uncle and Nephew), but they each possess a low genetic load. However, because they are related, their offspring express a high realized load, even though their genetic load is low.

3.3 | Computer simulations of the genetic load

Finally, we performed computer simulations examining the impact of genomics-informed captive breeding on the neutral nucleotide diversity, genetic load, realized load and fitness of individuals. The “Random mating” and “Minimize relatedness” regimes showed a steady increase in genetic load (Figure 5a) and realized load (Figure 5b) over generations. Both regimes also suffered from a large decline in fitness due to a mutation meltdown (Figure 5c). In contrast, both the genetic load and realized load were reduced in the “Minimize load” and “Minimize load and relatedness” regimes (Figure 5a,b). Therefore, genomics-informed captive breeding can successfully reduce the realized load and homozygosity of deleterious mutations, independently of consideration of relatedness. Consequently, mean fitness remained high in these regimes, increasing during the first 10 generations (Figure 5c). However, populations lost neutral genetic diversity at a relatively fast rate in the “Minimize load” regime (Figure 5d). Such loss in diversity was not observed in the “Minimize load and relatedness” regime, and after ~10 generations, this regime maintained more diversity than the “Random mating” regime (Figure 5d).

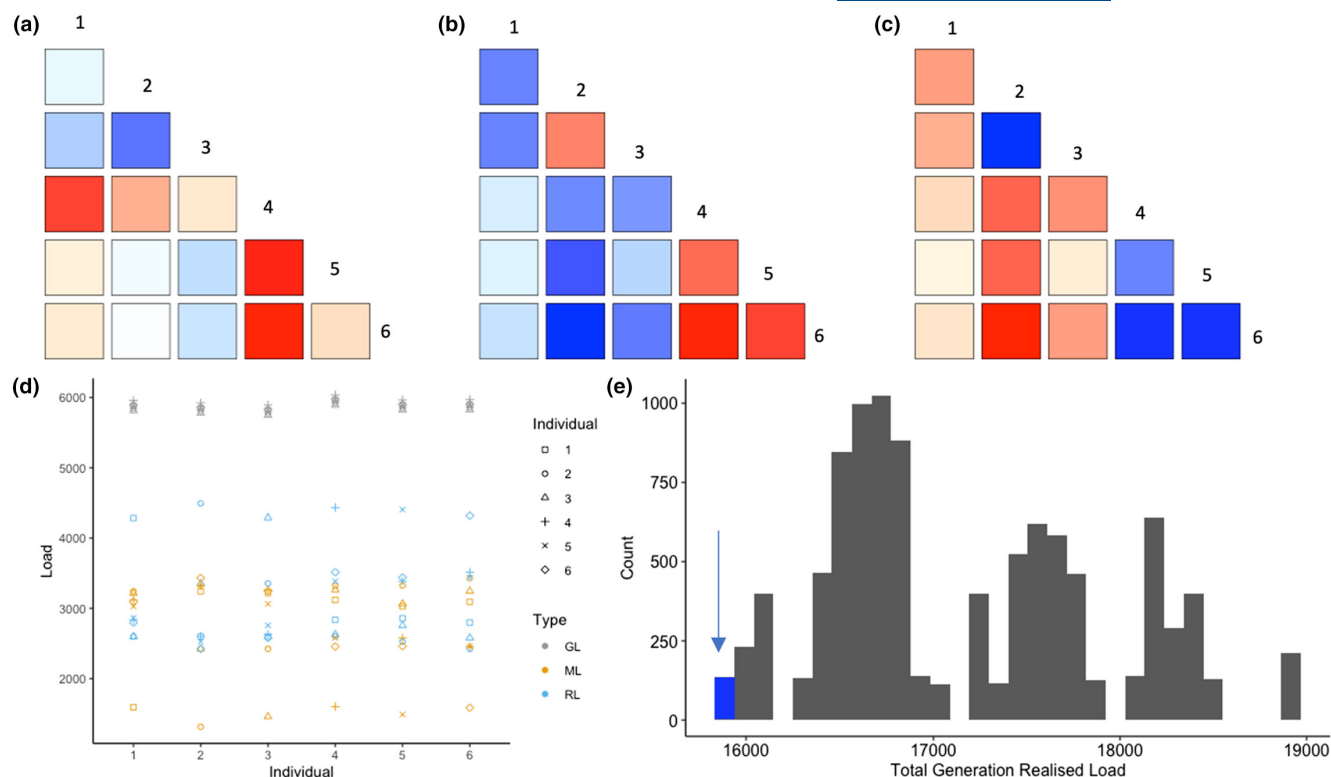


FIGURE 4 The genetic load at UCEs of six pink pigeons was calculated using cross-mapped chCADD scores. Correlogram showing the total load of potential offspring between all combinations of six individuals of the captive pink pigeon population. The colour of the tile is relative to the load of the offspring when compared to other potential offspring, and it is ranked on a gradient from high load (red) to low load (blue). (a) genetic load of the offspring between two potential parents, (b) realized load and (c) masked load. (d) The genetic load (grey), realized load (blue) and masked load (orange) of the hypothetical offspring of all possible crosses (including “selfing”). (e) The distribution of the summed realized load in all offspring was calculated by crossing all six individuals at random. In this procedure, each individual was crossed twice without self-mating or repeating the same crosses, and this was repeated 10,000 times. The optimal crossing combination is shown in blue.

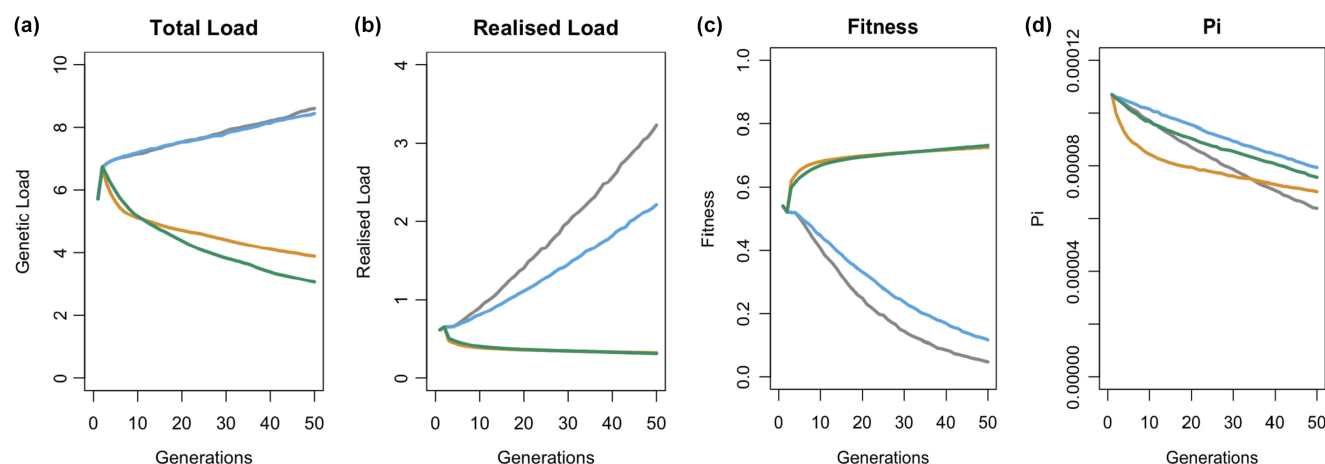


FIGURE 5 Impact of the four breeding regimes, simulated over 50 generations. Showing the impact on (a) the genetic load, (b) the realized load of offspring, (c) the mean fitness of adults and (d) neutral nucleotide diversity (π). Each coloured line corresponds to a specific mating regime: “Random mating” (grey), “Minimize relatedness” (blue), “Minimize load” (orange) and “Minimize load and relatedness” (green). The genetic load and realized load are expressed in lethal equivalents calculated using Equations (1) and (2) in Section 2 (see Bertorelle et al., 2022). The values presented in the figure represent the mean results obtained from 100 replicates.

4 | DISCUSSION

We conducted a proof-of-concept study to evaluate the utility of genomics-informed conservation for the management of captive

populations in zoos. Our aim was to examine whether we could use genomic data to reduce the level of inbreeding depression and genetic load, thereby increasing both the short- and long-term population viabilities. We developed a novel bioinformatics pipeline

to estimate the genetic load using CADD scores calculated for a model species (the chicken). We piloted our bioinformatics pipeline on the genomes of six pink pigeons from the captive-bred population from two UK zoos (Jersey Zoo and Bristol Zoo). We quantified realized load in hypothetical offspring by crossing these six individuals, showing that inbreeding depression may be reduced in the captive pink pigeon population. We furthermore found that UCEs possess the most severely deleterious mutations with the highest CADD scores and that mutations in UCEs occur at a lower SNP density and frequency compared to polymorphisms in the flanking regions. These observations are consistent with purifying selection.

Substantial genetic drift and inbreeding in zoo populations reduce long-term viability. Since the early 1970s, conservation biologists have used pedigrees and neutral genetic markers to assess and minimize inbreeding (Rabier et al., 2020). However, genetic load cannot be effectively measured or managed using this approach because neither markers nor pedigrees contain information about the segregation of deleterious mutations. Furthermore, pedigree data do not capture the possible relatedness between founder individuals. This can be especially problematic in populations that experienced a bottleneck before being sampled.

We showed our bioinformatics pipeline can identify optimal crosses that produce offspring with on average 7.4% lower realized load than random crosses. These offspring are expected to show less inbreeding depression. This reduction in realized load was modest because after nearly 10 generations in captivity, all pink pigeon individuals are relatively related. Crosses between closely related individuals have been minimized in the captive management of this species by exchanging pigeons between different zoos. However, this means that all individuals are similarly related. More substantial gains can be made in reducing the realized load using genomics-informed breeding in zoo populations with individuals that are less closely related. Genomics-informed breeding will be especially efficient in reducing inbreeding depression in captive populations founded by many individuals, fewer generations in captivity, non-bottlenecked species and species with a large ancestral population size (Bertorelle et al., 2022). These are all scenarios of populations that are likely to possess a high genetic load of segregating deleterious mutations not yet purged (Dusseux et al., 2023), with considerable differences in genetic load between individuals.

We do not know how CADD scores translate into fitness effects, and hence, we cannot calculate the exact benefits of genomics-informed breeding for survival rates. For example, two mutations with a CADD score of 10 each do not necessarily have the same fitness impact as one mutation with CADD score of 20. Theoretically, if a population carries a realized load of one lethal equivalent (LE), the probability of an individual surviving equals $\text{Fitness} = e^{-1} = 0.368$. Hence, a reduction of 7.4% in realized load results in an increase in survival rate from 36.8% to 39.6% ($\text{Fitness} = e^{-0.926} = 0.396$). This is a 7.7% relative increase in survival probability. With a higher realized load of two LEs, the survival probability is expected to improve from 13.5% to 15.7%, which amounts to a relative increase of nearly

16%. More generally, reducing the realized load is likely to reduce inbreeding depression and increase fitness (Bertorelle et al., 2022). Furthermore, the higher the realized load, the more gain can be made with genomics-informed management, at least theoretically. Future investigations into CADD scores and fitness effects will be able to provide valuable insights into the correlation between the realized load and individual fitness. Additionally, estimating the fitness effects of variants with known CADD scores would also help to improve the assessment of the extinction risk in population viability analysis (Jackson et al., 2022) and in artificial intelligence-informed conservation genomics (van Oosterhout, 2023). Such studies are required before our method of genomics-informed breeding can be formally implemented in the management of zoo populations.

Our simulations indicate that the genetic load and realized load can be reduced by the “Minimized load regime” and the “Minimized load and relatedness regime”. This resulted in a substantial increase in fitness compared to the “Random mating regime” and the “Minimized relatedness regime”. Although the “Minimized load regime” resulted in a substantial loss in nucleotide diversity, this was avoided by reducing relatedness in the “Minimized load and relatedness regime”. Theoretically, this regime is the optimal approach to maximize the long-term viability of captive populations, in terms of both reduced genetic load and maintaining adaptive potential.

To conclude, CADD scores for model species can be successfully lifted over to provide an initial assessment of the genetic load from whole-genome sequence data of non-model species. Optimal mate pairs can be identified to manage the realized load and inbreeding depression in the offspring generation. Computer simulations show that genomics-informed breeding can reduce the genetic load and realized load, and this can be accomplished with little reduction in nucleotide diversity. Genomics-informed conservation holds real potential for the management of captive populations, and it could also help to select the optimal individuals for reintroduction and genetic rescue programs.

AUTHOR CONTRIBUTIONS

Cock van Oosterhout and Samuel Speak conceived the study; Samuel Speak and Chiara Bortoluzzi developed the CADD analysis methods; Samuel Speak developed the LoadLift Snakemake and analysed the genomic data; Thomas Birley and Hernán Morales conducted the SLIM simulations; Chiara Bortoluzzi, Matthew Clark, Lawrence Percival-Alwyn, Hernán Morales and Cock van Oosterhout supervised the study; Matthew Clark and Lawrence Percival-Alwyn contributed to DNA sequencing; Samuel Speak, Hernán Morales and Cock van Oosterhout wrote the paper; all authors contributed to and approved the manuscript.

ACKNOWLEDGEMENTS

The authors would like to thank Harriet Whitford of the Durrell Conservation Wildlife Trust in Jersey and Bristol Zoo for providing samples. Library construction and sequencing reagents were a gift from 10x Genomics and Illumina, respectively. Sequencing was performed by the Genomics Pipeline Group at Earlham Institute (EI) with support from the UK Research and Innovation (UKRI) Biotechnology

and Biological Sciences Research Council (BBSRC). Research at EI was funded by the following grants: BBS/E/T/000PR6193, BBS/E/T/000PR5885, BBS/E/T/000PR9818 and BBS/E/T/000PR9816. The authors acknowledge the Research/Scientific Computing teams at the James Hutton Institute and NIAB for providing computational resources and technical support for the "UK's Crop Diversity Bioinformatics HPC" (BBSRC grant BB/S019669/1), the use of which has contributed to the results reported within this paper. H.M. was funded by an EMBO fellowship (grant 1111-2018) and the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie (grant 840519). C.v.O. was awarded the Royal Society International Collaboration Awards 2020 (ICA/R1/201194). S.A.S. was funded by a NERC ARIES PhD studentship (T209447) at the UEA and a Research Training Support Grant (RTSG; 100162318RA1). T.B. was funded by a BBSRC (BB/M011216/1) PhD studentship at the UEA. C.B. was funded by the Wellcome grant WT207492.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

Genetic data: The Raw sequence reads for the six pink pigeon individuals have been deposited in the NCBI SRA (BioSample: PRJNA1018937, Accessions: SAMN37457073, SAMN37457074, SAMN37457075, SAMN37457076, SAMN37457077, SAMN37457078). The pink pigeon reference genome used for this project has been submitted to the NCBI BioSample: PRJNA1018937 (Speak et al., 2023). The Chicken bGalGal6 genome is publicly available on NCBI (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000002315.5/).

The chCADD scores are publicly available on the OSF (<https://doi.org/10.17605/OSF.IO/8GDK9>).

Scripts: The LoadLift Snakemake pipeline is available on GitHub (<https://github.com/saspeak/LoadLift>).

BENEFIT-SHARING STATEMENT

Benefits Generated: Benefits from this research accrue from the sharing of our data and results on public databases as described above.

ORCID

Samuel A. Speak  <https://orcid.org/0000-0002-4207-7562>

Hernán E. Morales  <https://orcid.org/0000-0002-2964-020X>

Cock van Oosterhout  <https://orcid.org/0000-0002-5653-738X>

REFERENCES

- Albeshr, M. (2016). *Conservation biology and genomics of a flagship endangered species: The Mauritian pink pigeon Nesoenas mayeri*. <https://ueaeprints.uea.ac.uk/id/eprint/59458/>
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved elements in the human genome. *Science*, 304(5675), 1321–1325. <https://doi.org/10.1126/science.1098119>
- 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. <https://doi.org/10.1038/s41576-022-00448-x>
- Boakes, E. H., Wang, J., & Amos, W. (2007). An investigation of inbreeding depression and purging in captive pedigreed populations. *Heredity*, 98(3), 172–182. <https://doi.org/10.1038/sj.hdy.6800923>
- Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3), 195–205. <https://doi.org/10.1038/nrg2526>
- Chen, J., Glémin, S., & Lascoux, M. (2020). From drift to draft: How much do beneficial mutations actually contribute to predictions of Ohta's slightly deleterious model of molecular evolution? *Genetics*, 214(4), 1005–1018. <https://doi.org/10.1534/genetics.119.302869>
- Cvijović, I., Good, B. H., & Desai, M. M. (2018). The effect of strong purifying selection on genetic diversity. *Genetics*, 209(4), 1235–1278. <https://doi.org/10.1534/genetics.118.301058>
- de Jong, M. J., van Oosterhout, C., Hoelzel, A. R., & Janke, A. (2024). Moderating the neutralist–selectionist debate: Exactly which propositions are we debating, and which arguments are valid? *Biological Reviews*, 99(1), 23–55. <https://doi.org/10.1111/brv.13010>
- DeWoody, J. A., Harder, A. M., Mathur, S., & Willoughby, J. R. (2021). The long-standing significance of genetic diversity in conservation. *Molecular Ecology*, 30(17), 4147–4154.
- 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, 961–969. <https://doi.org/10.1016/j.tree.2023.05.008>
- Faircloth, B. C. (2016). PHYLUC is a software package for the analysis of conserved genomic loci. *Bioinformatics*, 32(5), 786–788. <https://doi.org/10.1093/bioinformatics/btv646>
- Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, 61(5), 717–726. <https://doi.org/10.1093/sysbio/sys004>
- García-Dorado, A., & Caballero, A. (2021). Neutral genetic diversity as a useful tool for conservation biology. *Conservation Genetics*, 22(4), 541–545. <https://doi.org/10.1007/s10592-021-01384-9>
- Gilbert, T., Gardner, R., Kraaijeveld, A. R., & Riordan, P. (2017). Contributions of zoos and aquariums to reintroductions: Historical reintroduction efforts in the context of changing conservation perspectives. *International Zoo Yearbook*, 51(1), 15–31. <https://doi.org/10.1111/izy.12159>
- Groß, C., Bortoluzzi, C., de Ridder, D., Megens, H.-J., Groenen, M. A. M., Reinders, M., & Bosse, M. (2020). Prioritizing sequence variants in conserved non-coding elements in the chicken genome using chCADD. *PLoS Genetics*, 16(9), e1009027. <https://doi.org/10.1371/journal.pgen.1009027>
- Groß, C., de Ridder, D., & Reinders, M. (2018). Predicting variant deleteriousness in non-human species: Applying the CADD approach in mouse. *BMC Bioinformatics*, 19(1), 373. <https://doi.org/10.1186/s12859-018-2337-5>
- Groß, C., Derks, M., Megens, H.-J., Bosse, M., Groenen, M. A. M., Reinders, M., & de Ridder, D. (2020). pCADD: SNV prioritisation in *Sus scrofa*. *Genetics Selection Evolution*, 52(1), 4. <https://doi.org/10.1186/s12711-020-0528-9>
- Habic, A., Mattick, J. S., Calin, G. A., Krese, R., Konc, J., & Kunej, T. (2019). Genetic variations of ultraconserved elements in the human genome. OMICS: A journal of. *Integrative Biology*, 23(11), 549–559. <https://doi.org/10.1089/omi.2019.0156>

- Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Molecular Biology and Evolution*, 36(3), 632–637. <https://doi.org/10.1093/molbev/msy228>
- Harrison, K. A., Pavlova, A., Telonis-Scott, M., & Sunnucks, P. (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications*, 7(9), 1008–1025. <https://doi.org/10.1111/eva.12149>
- Hoban, S., Archer, F. I., Bertola, L. D., Bragg, J. G., Breed, M. F., Bruford, M. W., Coleman, M. A., Ekblom, R., Funk, W. C., Grueber, C. E., Hand, B. K., Jaffé, R., Jensen, E., Johnson, J. S., Kershaw, F., Liggins, L., MacDonald, A. J., Mergeay, J., Miller, J. M., ... Hunter, M. E. (2022). Global genetic diversity status and trends: Towards a suite of essential biodiversity variables (EBVs) for genetic composition. *Biological Reviews*, 97(4), 1511–1538. <https://doi.org/10.1111/brv.12852>
- IUCN. (2023). IUCN. 2023. *The IUCN red list of threatened species. Version 2023-1*. <https://www.iucnredlist.org>
- Jackson, H. A., Percival-Alwyn, L., Ryan, C., Albeshr, M. F., Venturi, L., Morales, H. E., Mathers, T. C., Cocker, J., Speak, S. A., Accinelli, G. G., Barker, T., Heavens, D., Willman, F., Dawson, D., Ward, L., Tatayah, V., Zuël, N., Young, R., Concannon, L., ... Van Oosterhout, C. (2022). Genomic erosion in a demographically recovered bird species during conservation rescue. *Conservation Biology*, 36(4), e13918. <https://doi.org/10.1111/cobi.13918>
- 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>
- Kircher, M., Witten, D. M., Jain, P., O’Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, 46(3), 310–315. <https://doi.org/10.1038/ng.2892>
- Lee, A. P., & Venkatesh, B. (2013). Ultraconserved elements (UCEs) in the human genome. In *Encyclopedia of life sciences*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470015902.a0020842.pub2>
- Mathur, S., Mason, A. J., Bradburd, G. S., & Gibbs, H. L. (2023). Functional genomic diversity is correlated with neutral genomic diversity in populations of an endangered rattlesnake. *Proceedings of the National Academy of Sciences of the United States of America*, 120(43), e2303043120. <https://doi.org/10.1073/pnas.2303043120>
- Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). Sustainable data analysis with Snakemake. *F1000Research*, 10, 33. <https://doi.org/10.12688/f1000research.29032.2>
- 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. <https://doi.org/10.1093/bioinformatics/btw044>
- 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>
- Rabier, R., Robert, A., Lacroix, F., & Lesobre, L. (2020). Genetic assessment of a conservation breeding program of the houbara bustard (*Chlamydotis undulata undulata*) in Morocco, based on pedigree and molecular analyses. *Zoo Biology*, 39(6), 422–435. <https://doi.org/10.1002/zoo.21569>
- Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 47(D1), D886–D894. <https://doi.org/10.1093/nar/gky1016>
- Ryan, C. (2021). *Better tools, better resources, better conservation: Integrating genome data into the conservation of the pink pigeon Nesoenas mayeri*. University of East Anglia.
- Sachdeva, H., Olusanya, O., & Barton, N. (2022). Genetic load and extinction in peripheral populations: The roles of migration, drift and demographic stochasticity. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 377(1846), 20210010. <https://doi.org/10.1098/rstb.2021.0010>
- Silla, T., Kepp, K., Tai, E. S., Goh, L., Davila, S., Ivkovic, T. C., Calin, G. A., & Voorhoeve, P. M. (2014). Allele frequencies of variants in ultra conserved elements identify selective pressure on transcription factor binding. *PLoS One*, 9(11), e110692. <https://doi.org/10.1371/journal.pone.0110692>
- Speak, S., Percival-Alwyn, L., Clark, M., & van Oosterhout, C. (2023). *Genetic load analysis of captive-bred pink pigeons*. NCBI. BioProject: PRJNA1018937.
- van Oosterhout, C. (2020). Mutation load is the spectre of species conservation. *Nature Ecology & Evolution*, 4(8), 1004–1006. <https://doi.org/10.1038/s41559-020-1204-8>
- van Oosterhout, C. (2023). AI-informed conservation genomics. *Heredity*, 132, 1–4. <https://doi.org/10.1038/s41437-023-00666-x>
- Visel, A., Prabhakar, S., Akiyama, J. A., Shoukry, M., Lewis, K. D., Holt, A., Plajzer-Frick, I., Afzal, V., Rubin, E. M., & Pennacchio, L. A. (2008). Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nature Genetics*, 40(2), 158–160. <https://doi.org/10.1038/ng.2007.55>
- Warnefors, M., Hartmann, B., Thomsen, S., & Alonso, C. R. (2016). Combinatorial gene regulatory functions underlie ultraconserved elements in drosophila. *Molecular Biology and Evolution*, 33(9), 2294–2306. <https://doi.org/10.1093/molbev/msw101>
- Warren, W. C., Hillier, L. W., Tomlinson, C., Minx, P., Kremitzki, M., Graves, T., Markovic, C., Bouk, N., Pruitt, K. D., Thibaud-Nissen, F., Schneider, V., Mansour, T. A., Brown, C. T., Zimin, A., Hawken, R., Abrahamsen, M., Pyrkosz, A. B., Morisson, M., Fillon, V., ... Cheng, H. H. (2017). A new chicken genome assembly provides insight into avian genome structure. *G3: Genes, Genomes, Genetics*, 7(1), 109–117.
- Wood, J. L. A., Yates, M. C., & Fraser, D. J. (2016). Are heritability and selection related to population size in nature? Meta-analysis and conservation implications. *Evolutionary Applications*, 9(5), 640–657. <https://doi.org/10.1111/eva.12375>
- Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., & Wang, L. (2014). CrossMap: A versatile tool for coordinate conversion between genome assemblies. *Bioinformatics*, 30(7), 1006–1007. <https://doi.org/10.1093/bioinformatics/btt730>

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: Speak, S. A., Birley, T., Bortoluzzi, C., Clark, M. D., Percival-Alwyn, L., Morales, H. E., & van Oosterhout, C. (2024). Genomics-informed captive breeding can reduce inbreeding depression and the genetic load in zoo populations. *Molecular Ecology Resources*, 24, e13967. <https://doi.org/10.1111/1755-0998.13967>