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

Genomics-informed captive breeding in zoos.

Samuel A Speak[1,2,3*], Thomas Birley[1], Chiara Bortoluzzi[4,5], Matthew D Clark[2,6], Lawrence Percival-Alwyn[7], Hernán E Morales[8] and Cock Van Oosterhout[1].

1. University of East Anglia, Norwich Research Park, Norwich, United Kingdom.
2. Natural History Museum, London, United Kingdom.
3. North of England Zoological Society, Chester Zoo, Chester, United Kingdom.
4. University of Cambridge, Trinity Lane, Cambridge, United Kingdom.
5. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom.
6. Earlham Institute, Norwich Research Park, Norwich, United Kingdom.
7. NIAB, Lawrence Weaver Road, Cambridge, United Kingdom.
8. Globe Institute, University of Copenhagen, Øster Farimagsgade, Copenhagen, Denmark.

* Corresponding Author

Corresponding Author Email: s.speak@uea.ac.uk

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 lift-over 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 realised load of potential offspring. We thus identify the optimal mate pairs that are theoretically expected to reproduce offspring with the least inbreeding depression. We use computer simulations to show how genomics-informed conservation can reduce the genetic load and maintain genome-wide diversity, arguing this will become instrumental in maintaining the long-term viability of zoo populations.

Keywords

Genomics-informed conservation, Inbreeding depression, Genetic load, *Nesoenas mayeri,* CADD, Captive populations.

Introduction

More than 28% of the 150,388 species on the Red List of the International Union for Conservation of Nature (IUCN) are threatened with extinction (IUCN, 2022). A relatively small subset of these species are 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). Deleterious mutations create harmful genetic variants in the genome, collectively known as genetic load (Bertorelle et al., 2022). High genetic load can compromise population viability and recovery potential of species, especially if they experienced a recent population size decline (Jackson et al., 2022; Sachdeva et al., 2022). In declining populations, the impact of genetic load on fitness is not immediately apparent. It can take many generations before the harmful effects of mutations become expressed in homozygous loci (Pinto et al., 2023). Consequently, the long-term viability of many zoo populations could be at risk, despite individuals and populations thriving now.

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 (Willi, van Buskirk and Hoffmann, 2006; Charlesworth, 2009; Harrisson et al., 2014, but see Wood, Yates and Fraser, 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). This may be an oversight. However, recent advances in genomics and bioinformatics could change that.

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), CADD 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 ranks genetic variants such as single nucleotide polymorphisms (SNPs) and insertions and deletions (indels) throughout the genome. This analysis integrates surrounding sequence context, gene model annotation, evolutionary constraints (e.g., GERP scores), epigenetic measurements, and functional predictions into CADD scores. CADD was employed to investigate conserved elements into the chicken Combined Annotation-Dependent Depletion (chCADD) (Groß, Bortoluzzi, et al., 2020), and 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, CADD can estimate the genetic load within an individual’s genome.

Presently, we cannot translate the impact scores of mutations such as CADD into fitness effects. Nevertheless, we can calculate CADD scores for all 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 realised load, and the proportion whose fitness effects remains masked as an inbreeding load or masked load (Bertorelle et al., 2022). The realised load comprises the genetic load that reduces fitness when the harmful effect of the mutations come to light. Inbreeding increases the realised load because more deleterious mutations become fully expressed as homozygous. By minimising realised load, conservation managers can reduce inbreeding depression. This could be particularly useful in captive-bred populations where breeding pairs can be manipulated to improve the fitness of offspring.

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 of 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 are areas of the genome phylogenetically conserved across diverged taxa (Bejerano et al., 2004). 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), with UCEs being 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, comparative genomic approaches can be used to obtain a proxy of the genetic load, building on the knowledge of model organisms and humans. Studying UCEs in reference genomes allows for between-species comparisons of the proxies of genetic load, realised 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 realised 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 to reduce inbreeding depression and maximise the long-term viability of zoo populations.

Materials and Methods

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 (Supplementary Figure S1) and have a high level of relatedness (F=0.064 to 0.346) (Supplementary Table 2), which is typical of many zoo populations (Boakes et al., 2007). See Supplementary Information for further details.

Genome sequencing and bioinformatics

DNA was extracted from blood, using Qiagen MagAttract, linked read library preparation was 10x Genomics Chromium technology, which were then sequenced on an Illumina HiSeq X with 2x150bp reads (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).

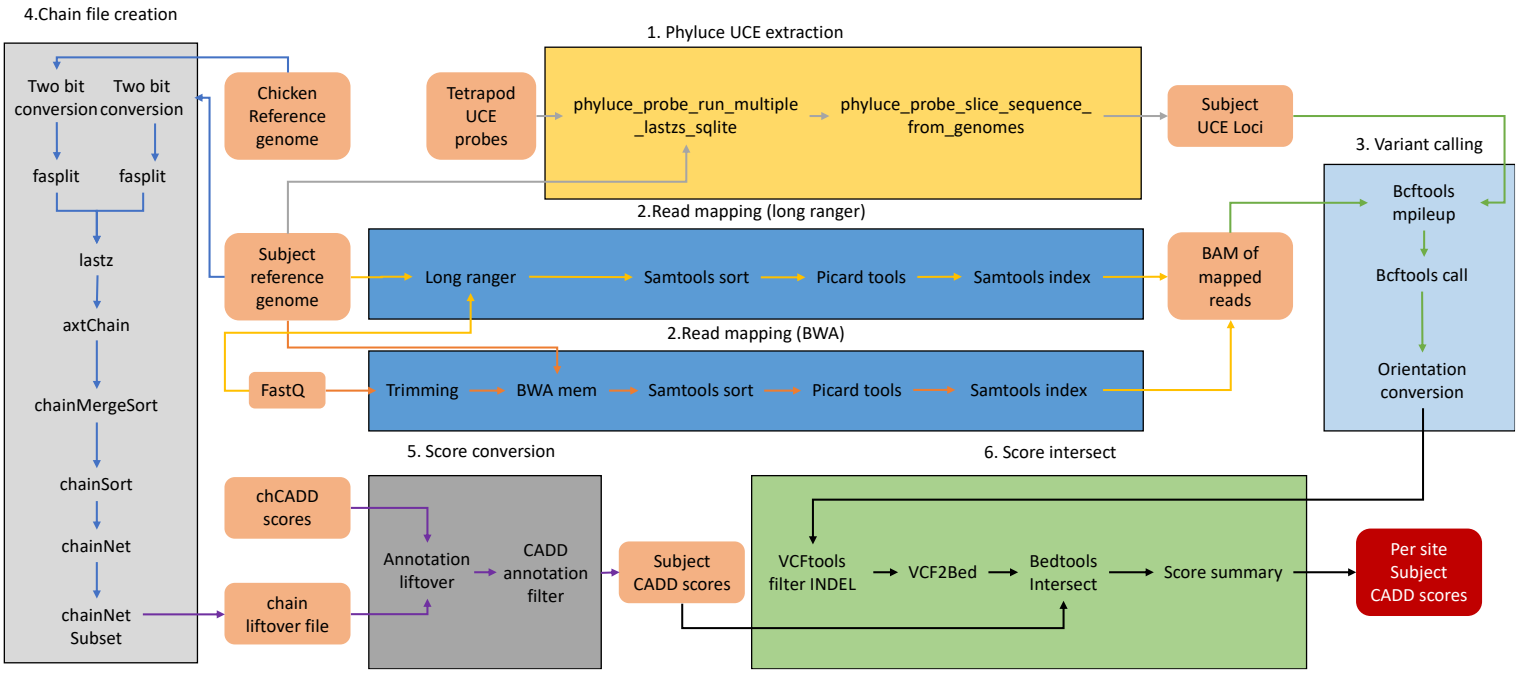


Figure 1 - The pipeline for the creation of per Single Nucleotide Polymorphism (SNP) pink pigeon Combined Annotation Dependent Depletion (ppCADD) scores from raw reads of individual pink pigeons. The Snakemake (Mölder et al., 2021) pipeline uses as input the sequencing reads of the subject individuals, 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 liftover 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).

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 lift-over 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. Genotypes of each locus were assessed to calculate the genetic load components. Individual's genetic load, realized load and masked load were calculated using the following formulas (Bertorelle et al., 2022):

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

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

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

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. Noted that part of the realised 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 Supplementary Information.

Computer simulations of breeding regimes

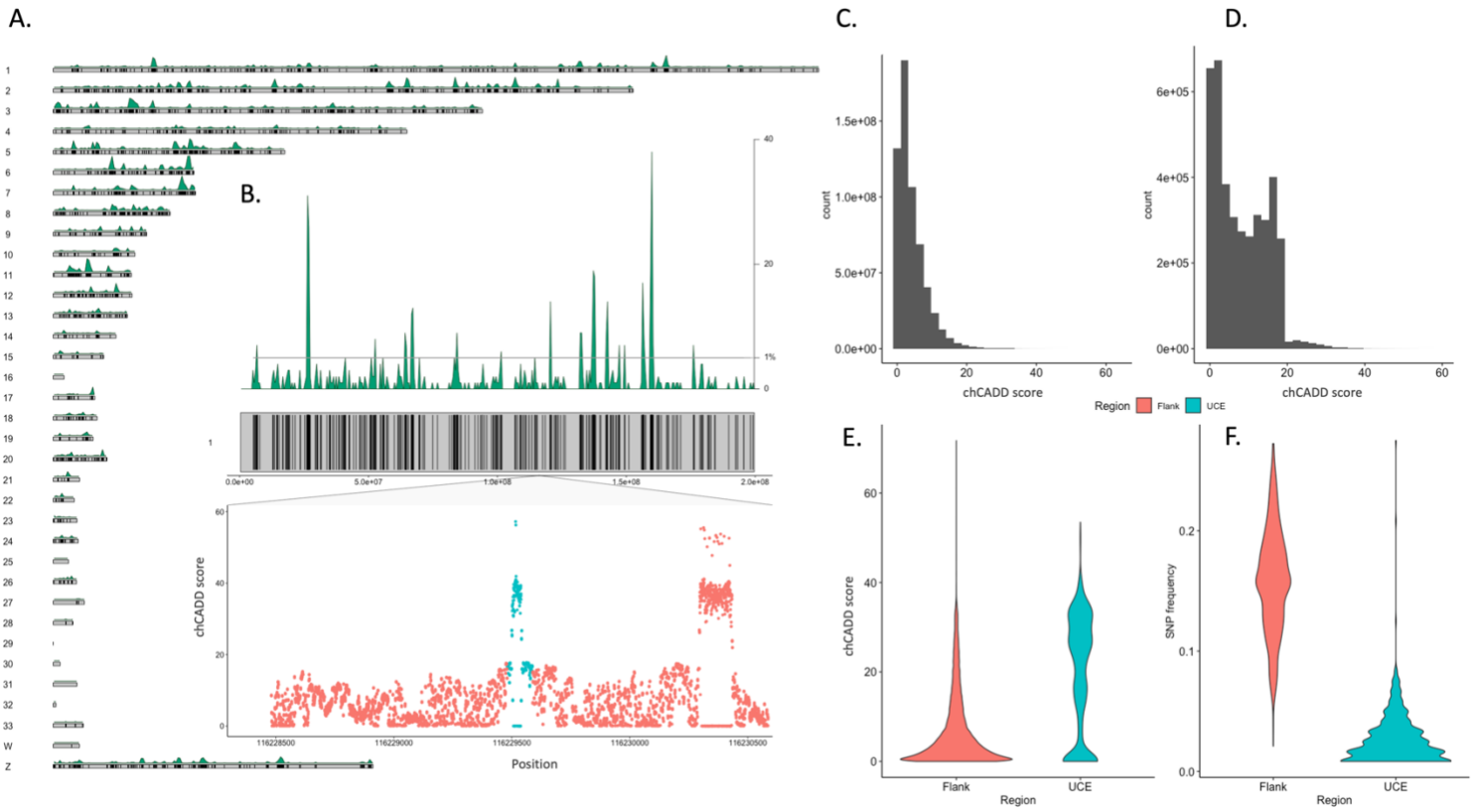
We conducted computer simulations in SLiM3 (Haller & Messer, 2019) to examine the impact of four breeding regimes on genetic and realised load, neutral genetic diversity, and fitness. In the “Minimise load” regime we examined whether mate pair selection can reduce the realised load of the offspring and alleviate 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 reducing relatedness (or kinship) of parents, and this was simulated in the “Minimise relatedness” regime. Additionally, we simulated a regime that aimed to minimise realised load of the offspring whilst maintaining genetic diversity, “Minimise 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 minimise realised load of their offspring. Finally, we simulated random mating “Random mating” regime. In each regime we

randomly sampled 20 monogamous pairs of males and females and allowed each pair to produce a brood of 64 offspring per generation. We ran 100 replicates for each regime for 50 generations. Further detail about the breeding regimes and SliM model are given in Supplementary Information.

Results

Distribution of UCEs and CADD scores

The 4976 UCEs along the 34 chromosomes of the chicken reference genome are not evenly distributed (Fig.2A), 15 chromosomes were significantly depleted for UCEs, whilst 9 chromosomes were significantly enriched for UCEs (Supplementary Table 1). Figure 2B shows the distribution of all chCADD scores along a single UCE (UCE-2729) and its 2000 bp 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 116230300 – 116230450 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).



238 **Figure 2– Distribution of ultraconserved elements (UCEs) and their mutation**
239 **impact scores (CADD scores). (A)** Karyotype plot of the chicken genome with the
240 distribution of UCEs (black bars) and density of UCEs (green peaks). (**B)** Karyotype
241 plot of chicken chromosome 1 showing the distribution of UCE-dense regions. Green
242 peaks above the 1% horizontal line are significantly enriched for UCEs ($p<0.01$). At the
243 bottom of Panel B, zoomed in at a single UCE and its 2000bp flanking regions (i.e.,
244 UCE2729), the CADD scores of every possible substitution at each site. The UCE is
245 shown in blue. The CADD scores in flanking regions are shown in red. Distribution of
246 all CADD scores for (**C)** the entire chromosome 1 of the chicken genome, and (**D)** 620
247 UCEs in chromosome 1 and their 2000bp flanking regions. **(E)** The CADD score
248 distribution of the flanking regions and the UCEs within the six pink pigeon genomes.
249 (**F)** SNP frequency at flanking regions and the UCEs. (See main text for test results).
250

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 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 = 4541885925, p-value < 0.0001). Whilst the frequency of derived mutations is significantly lower at UCEs compared to that at the flanking regions (Mann-Whitney test W = 13010970, p-value < 0.0001), consistent with the effect of purifying selection.

Genetic load components and kinship load

We analysed the genetic load in the hypothetical offspring of our six pink pigeons. This kinship load is calculated by theoretically crossing all possible combinations of individuals assuming mendelian segregation ratios. As kinship between two individuals increases, homozygosity of their offspring increases (Figure 3). Similarly, increased kinship between parents elevates offspring's' realised load and reduces masked load (Figure 3). Optimal mate pairing can significantly reduce the realised load of the offspring (R^2=0.258, $F_{1,13}$ = 8.32, p=0.00918).

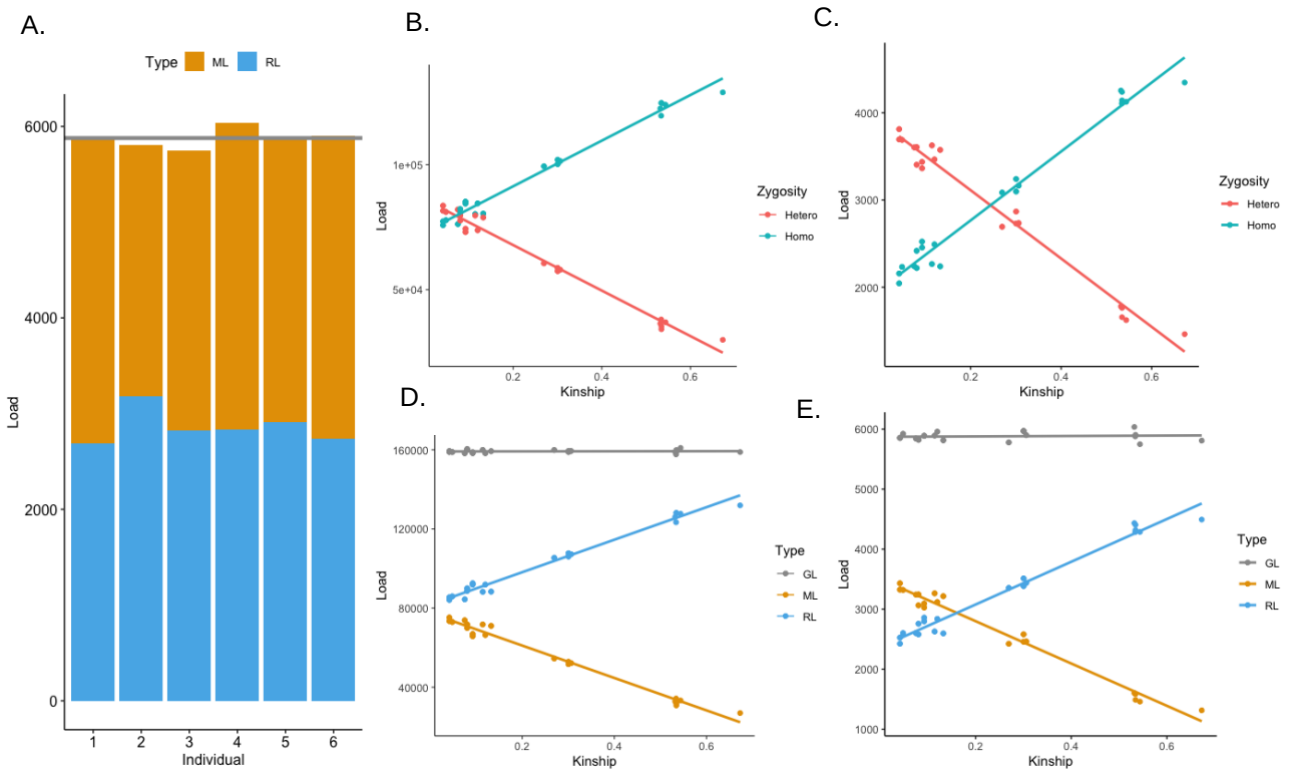


Figure 3 – The composition of the genetic load in six pink pigeon individuals and their hypothetical offspring. (A) The total realised load (Blue) and masked load (Orange) in each of the six pink pigeon individuals within their UCEs. **(B and C)** The realised load at heterozygous loci (Red) and homozygous loci (Teal) of the offspring is shown for the total region (B) and UCEs only (C). **(D and E)** The genetic load (Grey), realised 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)**.

Next, we performed an analysis to identify optimal crosses to minimise genetic load (Figure 4). Figure 4A shows average genetic load of potential offspring. In essence, these are the deleterious mutations that offspring are predicted to inherit from both

parents, 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.

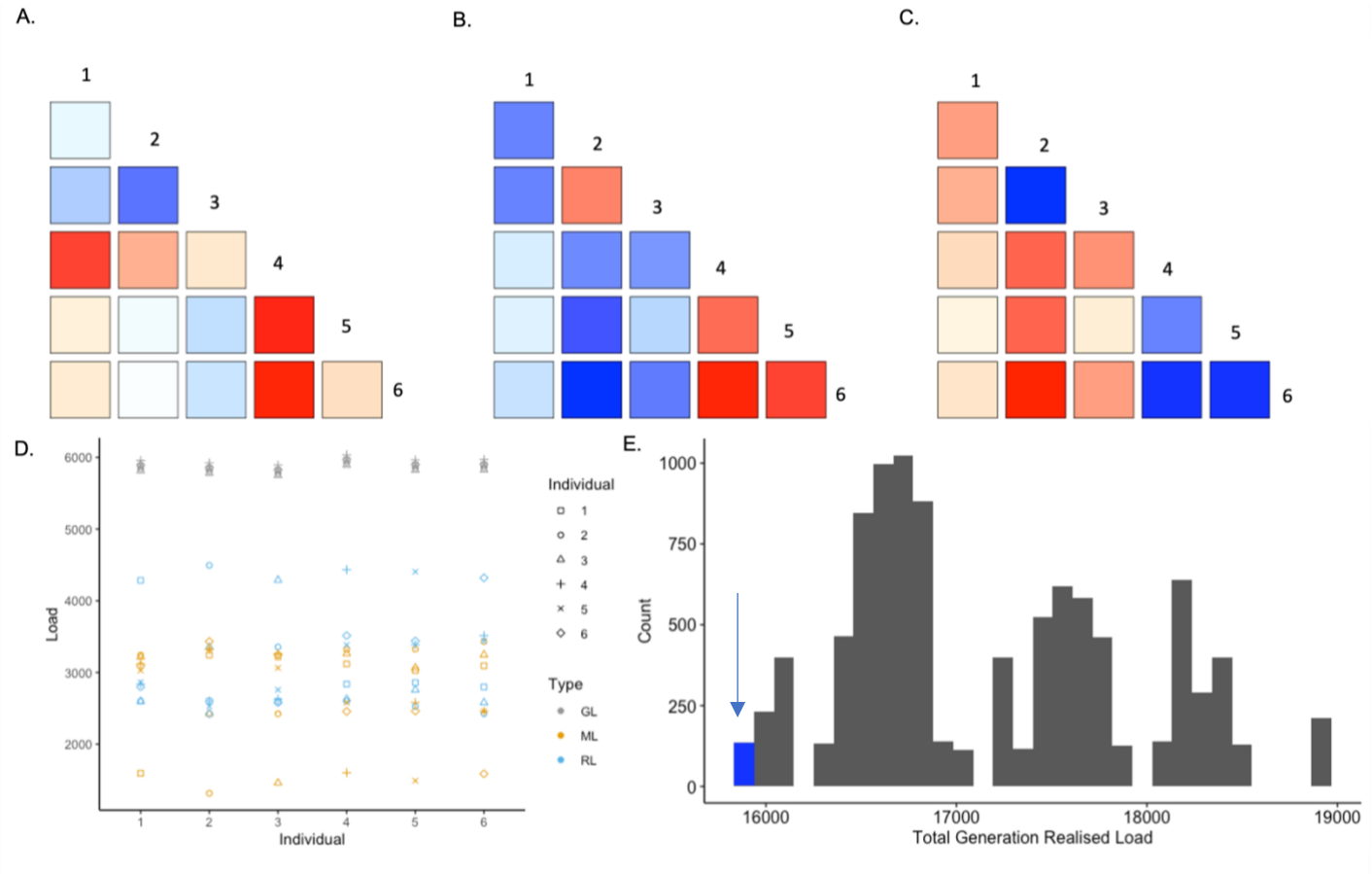


Figure 4 – The genetic load at UCEs of six pink pigeons calculated using cross-mapped chCADD scores. Correlogram showing the total load of potential offspring between six individuals of the captive pink pigeon population. The colour of the tile is relative to the load of the offspring when compared to 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)** realised load and **(C)** masked load. **(D)** The genetic load (grey), realised load (blue) and masked load (orange) of the

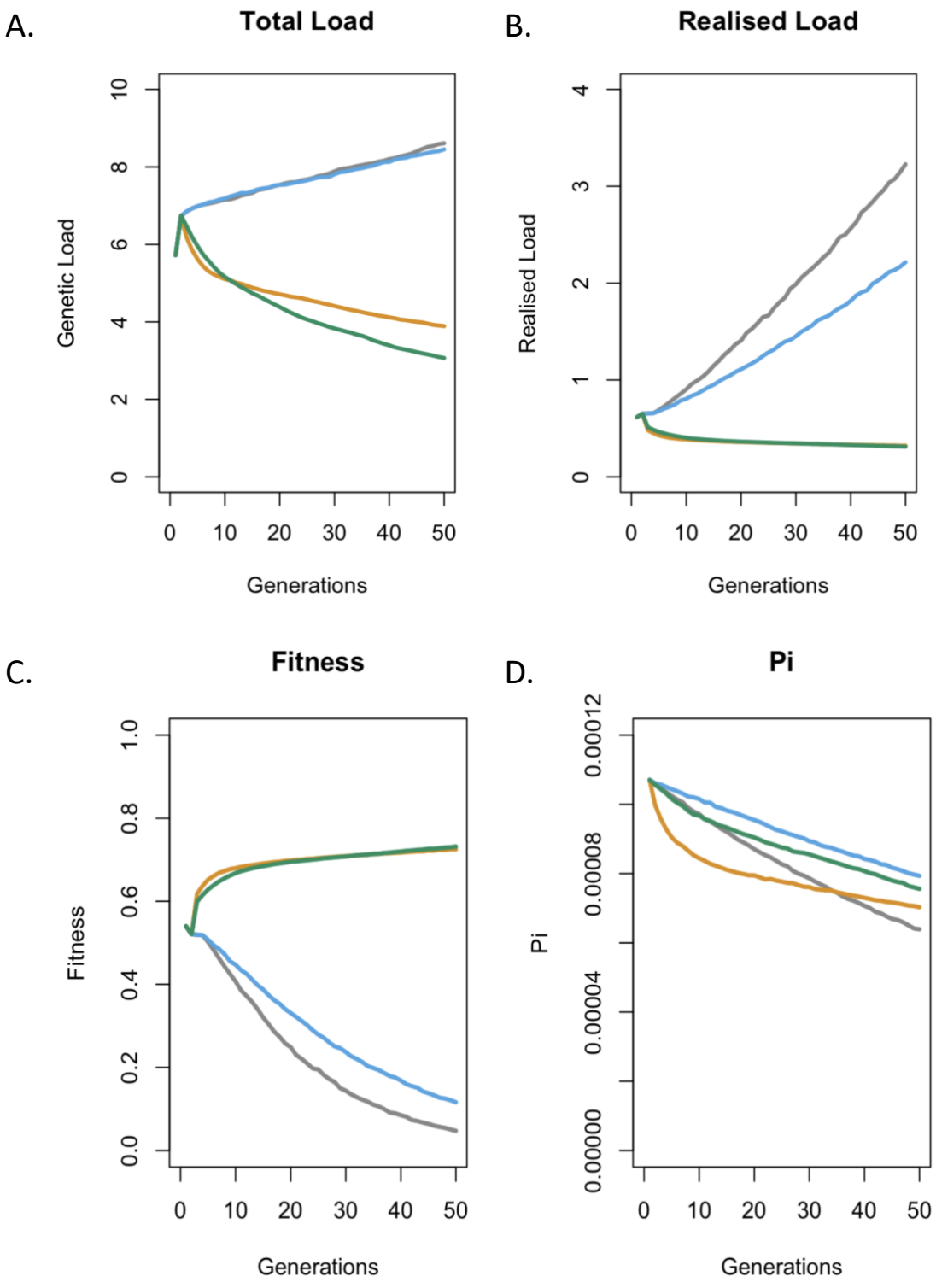
hypothetical offspring of all possible crosses (including “selfing”). **(E)** The distribution of total realised load in the offspring generation calculated by crossing all individuals at random. In this procedure, each individual was crossed twice without self-mating or repeating the same crosses, and this was repeated 10,000 times. The optimal crossing combination is shown in blue.

To predict degree of inbreeding depression, the realised load of the offspring of different crosses was calculated. Blue tiles in the correlogram in Figure 4B show the realised load of the offspring of the optimal crosses. The realised 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 x 3 cross with the lowest genetic load possesses a relatively high realised load. Individuals 2 and 3 were closely related (Aunt and Niece), but they each possess a low genetic load. However, because they are related, their offspring expresses a high realised load, even though their genetic load is low.

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, realised load, and fitness of individuals. The "Random mating" and “Minimise relatedness” regimes showed a steady increase in genetic (Fig. 5A) and realised (Fig. 5B) load over generations. Both regimes also suffered from a large decline in fitness due to a mutation meltdown (Fig. 5C). In contrast, both the genetic load and realised load were reduced in “Minimise load” and “Minimise load and relatedness” regimes (Fig. 5A,B).

Therefore, genomics-informed captive breeding can effectively purge deleterious mutations and reduce their homozygosity, independently of consideration of relatedness. Consequently, mean fitness remained high in these regimes, increasing during the first ten generations (Fig. 5C). However, populations lost neutral genetic diversity at a relatively fast rate in the “Minimise load” regime (Fig. 5D). Such loss in diversity was not observed in the “Minimise load and relatedness” regime, and after ~10 generations, this regime maintained more diversity than the “Random mating” regime (Fig. 5D).



329 **Figure 5- Impact of the four breeding regimes, simulated over 50 generations.**

330 Showing the impact on **(A)** the genetic load, (**B)** the realised load of offspring, **(C)** the

fitness of adults, and **(D)** neutral nucleotide diversity (π). Each coloured line corresponds to a specific mating regime: "Random mating" (grey), "Minimise relatedness" (blue), "Minimise load" (orange), and "Minimise load and relatedness" (green). The genetic load and realised load are expressed in lethal equivalents calculated using equations [1] and [2] in the Material & Methods (see Bertorelle et al., 2022). The values presented in the figure represent the mean results obtained from 100 replicas.

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 viability. We developed a novel bioinformatics pipeline to estimate the genetic load using CADD sores 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 realised 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 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 reduces long-term viability. Since the early 1970s, conservation biologists have used pedigrees and neutral genetic markers to assess and minimise 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 does 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 realised load than random crosses. These offspring are expected to show less inbreeding depression. This reduction in realised load was modest because after nearly 10 generations in captivity, all pink pigeon individuals are relatively related. Crosses between closely related individuals have been minimised 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 realised 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 (Dussex et al., 2023),with considerable differences between individuals.

We do not know how CADD scores translate in fitness effects, and hence, we cannot calculate the exact benefits of genomics-informed breeding for survival rates. If a population carries a realised load of one lethal equivalent (LE), a reduction of 7.4% in realised load results in an increase of survival rate from 36.8% to 39.6%. This is a 7.7% relative increase. With a higher realise load of 2 LEs, the survival rate improves from 13.5% to 15.7%, which amounts to a relative increase of nearly 16%. More generally, reducing the realised load is likely to reduce inbreeding depression and increase fitness (Bertorelle et al., 2022).

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

inbreeding depression in the offspring generation. Computer simulations show that genomics-informed breeding can reduce the genetic load and realised load, and this can be accomplished without significantly reducing nucleotide diversity in the population. Genomics-informed management can increase the long-term viability of captive populations and help to select the optimal individuals for reintroduction and genetic rescue programs.

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), 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). SAS was funded by a NERC ARIES

PhD studentship (T209447) at the UEA and a Research Training Support Grant (RTSG; 100162318RA1). TB was funded by a BBSRC (BB/M011216/1) PhD studentship at the UEA. C.B. was funded by the Wellcome grant WT207492.

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

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*. https://doi.org/10.1016/j.tree.2023.05.008

Faircloth, B. C. (2016). PHYLUCE 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/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

Harrisson, 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/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/https://doi.org/10.1111/brv.12852

IUCN. (2022). *The IUCN Red List of Threatened Species. Version 2022-2. https://www.iucnredlist.org. Accessed on [04/08/2023].* 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/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*, *118*(48). 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*. https://doi.org/https://doi.org/10.1002/9780470015902.a0020842.pub2

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*. 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/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

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

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., & others. (2017). A new chicken genome assembly provides insight into avian genome structure. *G3: Genes, Genomes, Genetics*, *7*(1), 109–117.

Willi, Y., van Buskirk, J., & Hoffmann, A. A. (2006). Limits to the Adaptive Potential of Small Populations. *Annual Review of Ecology, Evolution, and Systematics*, *37*, 433–458. http://www.jstor.org/stable/30033839

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/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

Conflict of interest statement

The authors have no conflict of interest to declare.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

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.

The Chicken bGalGal6 genome is publicly available on NCBI (GCF_016699485.2).

The chCADD scores are publicly available on the OSF (DOI 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.

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 the manuscript and approved.