Genomics-informed conservation of endangered bird species in zoos

Samuel Andrew Speak

A thesis submitted to the School of Environmental Sciences of the University of East Anglia in partial fulfilment of the requirements for the degree of Doctor of Philosophy

August 2024

Abstract

Globally, the biodiversity crisis threatens ~28% of species with extinction according to the Red List of the International Union for the Conservation of Nature (IUCN). Captive breeding programmes can act as insurance populations against extinction and to preserve genetic diversity. However, due to their small size, the survival of these populations is threatened by inbreeding depression resulting from high genetic load. I developed the LoadLift pipeline to assess the genetic load of individuals. LoadLift utilises Combined Annotation-Dependent Depletion (CADD) scores from model species to assess the impact of mutations and estimate the genetic load within ultraconserved elements (UCEs). Six pink pigeons (*Nesoenas mayeri*) were analysed with LoadLift and *in silico* crossings, to identify optimal mate pairings expected to show the least inbreeding depression. The CADD scores of three model species (humans, pigs and chickens) were highly comparable, giving confidence that the LoadLift approach can be applied across vertebrates. However, CADD scores cannot be summed as they represent a rank value of the predicted impact of a mutation relative to all other mutations within the genome. Hence, I converted CADD scores to selection coefficients (s) that were simulated in SLiM by aligning the rank scores of both CADD and s. LoadLift was also used to assess the genetic load of the whooping crane (*Grus americana*), which revealed that this species possesses a greater realised load than masked load, which is consistent with extensive inbreeding during the population bottleneck. I also compared the methods of LoadLift and SNPeff, showing good correspondence in their classifications of deleterious mutations. LoadLift enables captive-breeding managers to maintain long-term viable populations and reduce inbreeding depression. LoadLift can also be applied to assist reintroduction programmes in identifying optimal candidates to provide genetic rescue, thereby maximising the potential of *ex situ* populations for species conservation and restoration.

Table of contents:

List of figures

List of tables

Acknowledgements

I would not have been able to produce this thesis without the incredible support of countless people who have guided and developed me as a researcher. I would like to thank:

My Primary Supervisor Cock van Oosterhout for all the incredible support he has provided, as well as his knowledge and boundless enthusiasm for conservation genomics which has inspired me throughout my PhD. I cannot thank him enough for his guidance over my MScR and PhD, and I am confident I would not be the researcher I am today without his help. Also, to my Secondary Supervisors Professors Simone Immler and Diana Bell, and to Professor David Richardson, for their guidance over the last four years and for accepting me into their groups at UEA.

ARIES DPT team and the School of Environmental Sciences for funding this project and my training over my PhD and providing a supportive network of friends and researchers. Also, to The Natural History Museum, Professor Matt Clark, Dr Eileen Cox and Dr Helen Robertson and the PGR students, for their support and invaluable advice.

My CASE partner Chester Zoo, for their funding and support throughout my PhD as well as Andrew Owen for his enthusiasm for bird conservation, Dr Leah Williams for her support, and Dr Claire Raisin for her knowledge of Mauritius conservation. I am incredibly grateful to Penny Rudd and also the Parrots and Penguins team at Chester Zoo for the amazing training and time that I spent with them as part of my placement, the fond memories I created during my time there, especially those of feeding Wiggly the penguin, which kept me going during my write-up.

Dr Hernán Morales for his knowledge, advice, simulation work and hospitality as well as Dr Claudia Fontsere for her work on the whooping crane. Also, to the team at the International Crane Foundation for their conservation work. With thanks also to Harriet Whitford from Jersey Zoo for her incredible work on the pink pigeon and for providing the samples for my analysis. Also, to Dr Zjef Pereboom, Raymond van der Meer and EAZA for their support and knowledge of EEPs.

Professor Jim Groombridge, Dr Simon Tollington and Rangsinee Sankhom from DICE, at the University of Kent, for their knowledge, laboratory facilities and support. Also, to Dr Lawrence Percival-Alnwyn for his bioinformatics guidance and support throughout my PhD, and Dr Chiara Bortoluzzi for their help and advice with CADD scores and assistance in the development of LoadLift. And with thanks to Dr Iain Milne and the UK's Crop Diversity Bioinformatics HPC for providing me with computational support for my research. Also, to Thomas Birley for his incredible SliM simulations and Edward Martin for his work on the sonification of CADD scores for science communication.

Professor Carl Jones for his work conserving Mauritian species, and the advice and amazing stories shared. To Dr Vikash Tatayah, the Mauritian Wildlife Foundation and to Tom Everly for his photographs of pink pigeons. To Kevin Ruhomaun and the NPCS for allowing me to experience the pink pigeons in their natural habitat, as well as to Dr Nicolas Zuël for his hospitality and guided tours of Ebony Forest.

Professor Thomas Mock and Dr Andrew Toseland for introducing me to bioinformatics during my MScR and for their teaching and guidance. Also, to Professor Iain Barr for his support whilst at UEA, and for inspiring me to get into conservation.

I would not have been able to achieve anything without the dedication and support of my parents Margret and John Speak and my grandparents Joan Houghton and George Speak for encouraging me to follow my passion for conservation. Also, to my brothers, Will and Tom, for their constant support and healthy competition that made me who I am. Who, along with Louise and Alice provided me with many pigeon-based gifts, and to the Peers Clan for their steady supplies of coffee and chocolate.

My friends have always inspired me both from far away and those in Norfolk who have made Norwich and UEA my home where I have created so many incredible memories over the last eight years. Most importantly Esme Peers for her constant support, motivation, and enthusiasm, I cannot express how thankful I am for everything she has done to help me along the way.

Publications and contributions

The development and initial proof of concept of the LoadLift pipeline was published in Molecular Ecology Recourses, “Genomics-informed captive breeding can reduce inbreeding depression and the genetic load in zoo populations”. This work is included in this thesis in Chapter 1 – “An introduction to conservation genomics and its applications for threatened species” and Chapter 2 – “The development and application of LoadLift for genomics-informed captive breeding”. The SLiM simulations in Chapter 2 were performed by Hernán E Morales Globe Institute, Copenhagen University and Thomas Birley UEA.

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, 00, e13967. https://doi-org.uea.idm.oclc.org/10.1111/1755-0998.13967

The results of the analysis of Chapter 2 were used in the sonification of genomic data to communicate the threats of genetic load. This work has been published in the Journal Zoo Biology:

Martin, E.J., Speak, S.A., Urban, L., Morales, H.E. and van Oosterhout, C. 2024, Sonification of Genomic Data to Represent Genetic Load in Zoo Populations. Zoo Biology, 43: 513-519. https://doi.org/10.1002/zoo.21859

The analysis in Chapter 3 – “Investigating the similarity of CADD scores in three model species”, was adapted to be included in the publication, Genomic erosion in the assessment of species extinction risk and recovery potential. Which is currently available as a pre-print:

Van Oosterhout, C., Speak, S. A., Birley, T., Bortoluzzi, C., Percival-Alwyn, L., Urban, L. H., Groombridge, J. J., Segelbacher, G., & Morales, H. E. 2022. Genomic erosion in the assessment of species extinction risk and recovery potential. bioRxiv, 2022.09.13.507768. https://doi.org/10.1101/2022.09.13.507768

In Chapter 4 – “Conversion of CADD scores to selection coefficients for load calculations”, the SLiM simulations of selection and dominance coefficients were performed by Hernán E

Morales, Globe Institute, Copenhagen University. This Chapter is currently in preparation for publication.

In Chapter 5 – “Conservation genomics of the previously endangered whooping crane (*Grus americana*)”, Library preparation and sequencing was carried out by Molly Cassatt-Johnstone of the University of California, Santa Cruz. Read mapping was carried out by Hernán E Morales, Globe Institute, Copenhagen University whilst Variant Calling and SNPeff analysis was carried out by Claudia Fontsere of the Globe Institute, Copenhagen University. This chapter is currently in preparation for publication and is currently available as a pre-print:

Claudia Fontsere, Samuel A. Speak, Andrew J. Caven, Juan Antonio Rodríguez, Xuejing Wang, Carolina Pacheco, Molly Cassatt-Johnstone, Georgette Femerling, Brigid Maloney, Jennifer Balacco, Joanna Collins, Ying Sims, Linelle Abueg, Olivier Fedrigo, Erich D. Jarvis, Barry K. Hartup, Beth Shapiro, M. Thomas P. Gilbert, Cock van Oosterhout & Hernán E. Morales. 2024. Persistent genomic erosion in whooping cranes despite demographic recovery. bioRxiv 2024.12.12.628160; doi: https://doi.org/10.1101/2024.12.12.628160

1. An introduction to conservation genomics and its applications for threatened species

1.1 Abstract

In this Chapter, I give an overview of the current issues facing global species, and how we can apply genomics to attempt to reduce the extinction risk of a species. Direct threats to population viability, including habitat loss and fragmentation, invasive species, and environmental change alter the interactions between the evolutionary forces. I explain how these altered interactions fundamentally affect fitness and population viability, and how this can be assessed by studying genetic and genomic variation. I review the classical approaches to conservation genetics that have been applied for the last 50 years and explain recent developments in conservation genomics, and the importance of the genetic load when assessing the health of endangered species. Furthermore, I provide a review of the ideas of reintroduction and genetic rescue programmes and the benefits that they can bring to populations, as well as the potential risks from outbreeding depression, deleterious alleles and genetic swamping. I outline the history of zoos, how they can be used for the conservation of species, and the critiques they face. After introducing the concepts of mutation impact scores and ultraconserved elements (UCEs), I explain how they have been used in genomic studies. I summarise the history and conservation of the threatened species that I assess in Chapter 2, the pink pigeon (*Nesoenas mayeri*), and in Chapter 5, the whooping crane (*Grus americana*). This Chapter aims to introduce the threats facing endangered species globally as well as how genomic techniques and theories can assist in the assessment and applications of species recovery.

1.2 Global species extinction risk

Globally we are currently suffering a biodiversity crisis with 28% of the 163,040 species on the Red List of the International Union for Conservation of Nature (IUCN) threatened with extinction (IUCN, 2024). 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). 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 the population viability and recovery potential of species, especially if they have experienced a recent decline in population size (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., 2024). Consequently, the long-term viability of many zoo populations could be at risk, despite individuals and populations thriving at present. Analysis of genetic load enables us to assess this indirect threat of inbreeding on the viability of current and future populations.

1.3 Classical conservation genetics

1.4 Evolutionary forces

In the current global biodiversity crisis, it is critical to understand how genetics and genomics have developed and how they are being used to conserve threatened species globally. Core to the study of conservation genetics, and how it can be applied to species conservation, are the five evolutionary forces that affect both the fitness of individuals and the dynamics and viability of populations. These five forces are mutation, selection, genetic drift, gene flow and recombination (Arenas et al., 2018). The combination and interplay of these forces acting upon genes, individuals and populations are essential to the evolution and maintenance of threatened species and populations.

Mutation is the first of the two primary evolutionary forces responsible for introducing new alleles and genetic diversity into populations (Berger et al., 2017; Arenas et al., 2018). When genes are replicated errors occur by chance, creating new genetic variation within the individual. Although seen as a random process, the rate at which mutations are acquired is

not uniform across the genome and differs between species (Drake et al., 1998; Lynch et al., 1999; Lynch et al., 2016). As a means of introducing novel variation into the population, mutation is the driving force for all the other forces, without which evolution would not be possible. Under the neutral theory of evolution, the majority of mutations that do occur within the genome are neutral (Kimura, 1968; Ohta, 2000) whilst a relatively small percentage will provide a fitness benefit to the individual, in turn, some may lead to a reduction in fitness. Recent studies suggest that rising temperature influences the mutation rate of species (Berger et al., 2017). Therefore, with rising global temperatures, due to climate change, it is likely species mutation rates will vary. This could become a potential issue for many threatened species as shifts in mutation rate can lead to populations suffering a mutation meltdown (Lynch et al., 1995). In which a population accumulates deleterious mutations at a steadily increasing rate, greater than they can be removed by selection, causing a decrease in fitness resulting in a rapid population decline and eventual extinction.

Selection acts upon the alleles present within a population, increasing or decreasing the frequency depending on whether the allele increases or decreases the fitness of the individual. Selection is, and often as such subjectively, quoted as one of the best theories in biology, whilst also being one of the most misunderstood (Gregory, 2009). It was famously theorised by Charles Darwin in his work “On the Origin of Species by Means of Natural Selection” (Darwin, 1859), as the process which acts upon traits to determine if they will be inherited in subsequent generations. Selection acts on phenotypic and/or behavioural differences between individuals that is encoded by heritable (genetic or epigenetic) variation of an individual’s genotype. Therefore, if a trait increases or decreases the fitness of an individual, selection will act upon it. Many forms of selection focus on different biological processes. For example, over-dominant selection describes where heterozygotes are in the optimal state, with a higher fitness than either homozygote (Saccheri and Hanski, 2006). Sexual selection focuses on the selection of traits that arises from a fitness difference due to competition for access to gametes for reproduction (Shuker and Kvarnemo, 2021). Natural selection may also be considered as either hard or soft (Wallace, 1975; Saccheri and Hanski, 2006). Soft selection is when the selection for a trait value is relative to the other trait values of individuals within a population. Hard selection is when selection is determined solely by a trait value and is independent of the values of others within the population (Bell

et al., 2021). The term “fitness” as coined in the phrase “the survival of the fittest” by Herbert Spencer is also misleading (Gregory, 2009). It can be misinterpreted and assumed in the current vernacular to refer to the strongest or healthiest individuals, or that individual’s survival is equal to fitness. However, fitness truly should be seen in terms of reproductive fitness or the ability to ensure that genetic material is passed on. It is therefore common for investigations to use a proxy for fitness of individuals for example the lifetime reproductive success (the number of independent offspring produced in an individual's lifetime) whilst the specific chosen proxy for fitness can infer different information on the selection pressures (Van de Walle et al., 2022). In this way, Darwinian fitness does not necessarily equate to survival, but rather, as the success of an individual to propagate genetic material into future generations.

Recombination is a process that takes place within sexual organisms to rearrange the genes present within the gametes that individuals produce. Recombination can be seen as the main reason for sexual reproduction (Burt, 2000), as without recombination, an individual would always have more deleterious genetic mutations than their parents (Lynch et al., 1995). Over generations these would build in number, eventually leading to the expression of these mutations, thereby leading to a reduction in fitness and eventually the species extinction. This process is known as Muller’s ratchet (Muller, 1964). Recombination therefore allows for the rearrangement of chromosomes to create new combinations, thus increasing the genetic variation of the population. This process also breaks any linkage blocks within the genome, which is when a section of DNA can be transferred to subsequent generations due to being in close vicinity to a gene that increases fitness. Similar to mutation, the rate of recombination differs greatly between species and is not uniform along the entire genome (Butlin, 2005). This creates areas of high recombination known as hotspots, as well as regions of linkage disequilibrium referred to as haplotype blocks.

Genetic drift is an entirely random process by which the frequency of alleles within a population change (Lynch et al., 2016; Arenas et al., 2018). This, especially in small, isolated populations, can lead to the loss of genetic variation due to the fixation of alleles within a population (Frankham, 1997). This makes genetic drift one of the major threats to the

survival of many endangered species globally. This is because, if a species has suffered a reduction in population size, genetic drift can hamper its ability to recover due to the loss of variation suffered during its population decline through the drift and fixation of alleles.

Gene flow acts in contrast to genetic drift: it is the introduction of genetic variation from one population to another (Slatkin, 1987). A common and natural means of gene flow is due to interbreeding between individuals from two separated gene pools due to migration, but it can also result from founder events, as well as extinction and recolonisation events. In this way, individuals increase the genetic diversity of the destination population and ensure that they do not become geographically isolated. Geographical isolation and an absence of gene flow are responsible for population divergence, in this way, a population's gene pool is mediated by the balance between gene flow and genetic drift (Clegg and Phillimore, 2010). Many endangered species face risks due to low levels of gene flow between natural populations (Magro Moraes et al., 2023). Gene flow may be influenced by man-made obstructions such as housing transport links, canals or dams which may break up natural migration routes, while for other species it may increase (Fusco et al., 2021). Climate change can also cause shifts in natural weather events which can disrupt migration routes (Seri and Rahman, 2021) and impact food supplies (Kubelka et al., 2022) which can lead to genetic isolation of populations. Understanding the effects of these evolutionary forces on conservation efforts, as well as the interplay between these forces, is key in the formation of the fields of conservation genetics and genomics.

1.5 The development of conservation genetics

The field of classical conservation genetics was developed in the early 1970s within conservation biology to study and combat the issues facing small or low-density populations (Frankel and Bennett, 1970; Frankel, 1974). Key concepts in conservation genetics are built on nearly 50 years of population genetics work that started in the late 1920s. The studies by Sewall Wright on genetic drift and fragmented populations were particularly instrumental (Wright, 1922) and largely shaped our understanding of the effects of demographic changes on the genetics of a population (Willi et al., 2022). Much of this early conservation genetic

research, and even the later work by Sewall Wright focussed on the risk of population extinction (Wright, 1949). In the 1970s, conservation genetic research increasingly focused on combatting the threats to species that had undergone, or were currently suffering, population declines. At this time, researchers used population genetic techniques – initially using isozyme and allozyme electrophoresis for the assessment of the population viability of threatened species.

This relationship between genetic drift (as discussed in section 1.4 of this thesis), population size and extinction became the heart of population genetics in the cycle termed an "extinction vortex" (Benson et al., 2019; Gilpin & Soulé, 1986). This is the theory that small populations increase the likelihood of genetic drift thereby reducing genetic variation and increasing homozygosity. This will in turn increase inbreeding depression, leading to a reduction in the fitness of individuals, resulting in a reduction in offspring production or survival leading to further reduction in population size and, if left unchecked, eventual extinction.

1.6 Genetic diversity

Genetic diversity describes the number of alleles present within a population for a gene. This is an extremely important aspect of species conservation and recovery potential. For example, genetic diversity is very low for many island populations of endangered bird species (Ando, 2019). This is often attributed to the severe population bottlenecks that these species have undergone, reducing their genetic diversity, and in turn their ability to adapt to challenges such as climate change (Ceballos et al., 2017). This can result in severe population decline. Applied conservation genetics aims to combat these issues, often by studying the changes within natural populations in the context of human activities. Habitat alteration, overexploitation and invasive species are considered the main drivers of biodiversity loss, and these operate alongside other lower-ranked threats (Maxwell et al., 2016). These direct threats lead to reductions in genetic diversity and an increase in the expressed genetic load, i.e., realised load (Dussex et al., 2023). Genomic analyses can help in the design and implementation of management plans for endangered species to estimate

factors such as effective population size (N_e) (Wilder et al., 2023) and past bottlenecks and through tools such as Vortex (Lacy et al., 2019), to predict potential future declines. In the past 50 years, conservation geneticists have focused on maintaining genetic variation (DeWoody et al., 2021; García-Dorado and 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, Yates and Fraser, 2016).

1.7 Genetic load

1.8 Calculating genetic load

Estimates of genetic load were originally used to quantify the loss of fitness within a population relative to the fittest individual (Crow, 1970; reviewed in Bertorelle et al., 2022). A statistical standardisation for calculating genetic load was developed by Morton et al. (1956). This method expresses the magnitude of inbreeding depression in populations without information on the number of loci that segregate for deleterious mutations (L), the frequency of those mutations (*q*), their selection coefficients (*s*), or dominance coefficients (*h*) (Morton et al., 1956). Therefore, the load of a population was expressed by grouping genes or mutations that together caused, on average, the death of a given number of individuals. The original wording by Morton et al (1956) of a lethal equivalent, in particular *"[…] would cause on average one death"*, has created considerable confusion. What does it mean if, on average, one individual dies? The definition tells us that in this case, the expressed genetic load (i.e., realised load) of this population is equal to one Lethal Equivalent (LE). Morton et al. (1956) also explain that this can be thought of as either two heterozygous sites that both reduce fitness by 50% or a single homozygous site that causes death. The biological meaning can be understood when realising this is the mean of a Poisson distribution of a population, and there is a zero-class of individuals that do not inherit an LE. That group are the survivors, whereas the individuals in the other groups (with 1, 2, 3, … LEs) all die. This also clarifies how a population can have a realised load exceeding one LE. For example, in a population with 2 LEs, most individuals inherit multiple LEs, but

there still is a zero-class that is mutation-free. Bertorelle et al. (2022) explain that the proportion of this class is equal to e^{-2}=13.5% of individuals.

Calculations of genetic load are a key aspect of future conservation and the long-term viability of declining populations (Pinto et al., 2024; van Oosterhout et al., 2022). Due to inbreeding and genetic drift, the realised load of homozygous mutations increases, thereby reducing fitness and population viability (Pinto et al., 2024).

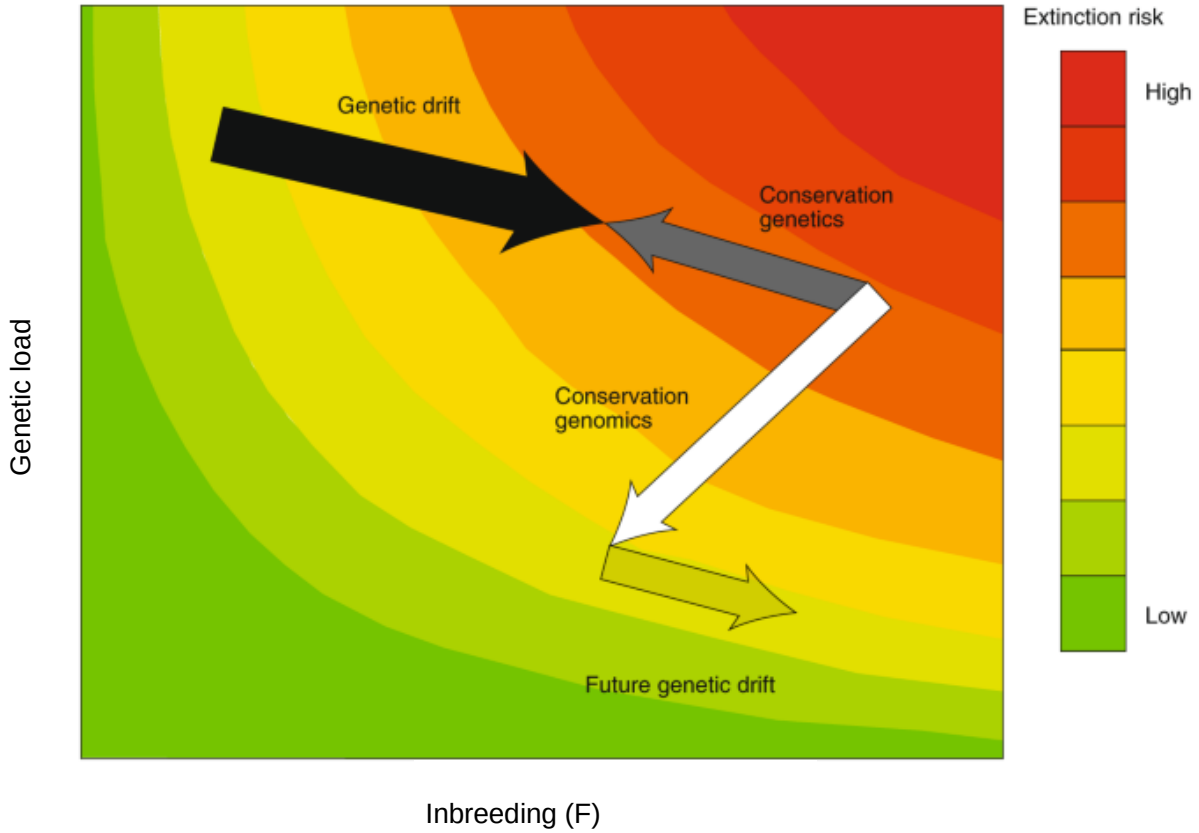


Figure 1 – Extinction risk of a species increases when the amount of inbreeding, genetic load and the number of deleterious mutations segregating within the population are high. Genetic drift (Black arrow) increases the species inbreeding coefficient and extinction risk. However, due to purging, the genetic load decreases during population size decline (Dussex et al., 2023). Traditional conservation methods aim to combat inbreeding and drift to maximise genetic diversity (Grey arrow), but in turn, this counteracts the purging of deleterious mutations and increases the genetic load relative to an unmanaged population. Genomics-informed conservation (white arrow), aims to reduce the future risk of extinction by reducing inbreeding and the genetic load of the population, making populations more resilient to future genetic drift (transparent arrow). Figure adapted from van Oosterhout, 2020.

A recent study using Vortex models (Lacy et al., 2019) with both genetic and demographic data has shown that without any genetic management, the current population of pink pigeons is likely to go extinct within the next 100 years (Jackson et al., 2022). Considering the recent reintroduction of individuals and genetic diversity (in the form of alleles not found in the wild population) from the captive-bred population to Isle aux Aigrettes and the Black River Gorges in Mauritius, it is vital that genomic techniques are used to calculate genetic load and alleviate inbreeding depression. The loss of fitness due to inbreeding is a function of both the rate of inbreeding as well as the genetic load. Genomic analyses of the genetic load would allow zoo managers to remove deleterious mutations from the gene pool (Figure 1, White arrow). As such, management takes on the role of natural selection which, due to management practices such as supplementary feeding, has been significantly relaxed compared to the ancestral population. Therefore, the relaxed environment does not pose a significant challenge to the individual for selection to act (Trask et al., 2019). By screening the genome for deleterious mutations, we will be able to identify individuals with the best 'genetic health', thereby minimising the level of inbreeding depression, and ensuring the long-term survival of the species.

The genome of every individual is likely to be affected by mutations. A very small fraction of these mutations may be potentially beneficial, whilst the vast majority are neutral or (slightly) deleterious (Kimura, 1968; Kimura, 1991; Ohta, 2000). As mutations occur over evolutionary time, many deleterious dominant variants are removed by selection. However, mutations that are (partially) recessive can remain in the population in heterozygote conditions (Henn et al., 2015). This accumulation of recessive deleterious mutations is referred to as the mutation load or genetic load (Bertorelle et al., 2022). There are several definitions of the genetic load (Henn et al., 2015; Galeota-Sprung et al., 2020). At the population level, the genetic load is the sum of all selection coefficients of deleterious mutations, each multiplied by their frequency in the population. The genetic load can also be calculated at an individual level. In that case, the genetic load is the sum of all selection coefficient (s) of deleterious mutations that are in homozygous condition, plus half the sum of all selection coefficient (s) of deleterious mutations that are in heterozygous condition (Bertorelle et al., 2022). Note that in these definitions of the genetic load, the dominance coefficient (h) of mutations is irrelevant (Bertorelle et al., 2022).

1.9 Genetic load components

The dominance coefficient (*h*) only becomes important when analysing the fitness effects of deleterious mutations. The genetic load is comprised of two elements: the masked load and the realised load (Bertorelle et al., 2022). The masked load comprises of the fitness effects of all (partially) recessive mutations that are in heterozygous conditions, and which are not completely expressed. In contrast, the realised load is the proportion of deleterious mutations that are present as homozygotes, plus the fitness effects of the partially recessive, additive and dominant mutations that are in heterozygous loci (Mathur and DeWoody, 2021; Bertorelle et al., 2022). Only the realised load reduces the fitness of individuals, whereas the masked load (also known as the inbreeding load, or potential load) remains hidden from selection. The masked load has the potential to become expressed in future generations, for example when loci become homozygous due to inbreeding following a population bottleneck. Hence, the alternative names “potential load” and “inbreeding load”.

It is worth noting that there is no masked load if the fitness effects of alleles of all loci are completely additive. In that case, there is no inbreeding depression. Given that most species do suffer a loss of fitness during inbreeding (and assuming that the number of over-dominant loci is small), we can conclude that $h<0.5$, and that deleterious mutations tend to be recessive. However, the actual distribution of *h* is only estimated and modelled for some species (García-Dorado and Caballero, 2000; Balick et al., 2015). This shows that with a low *h*, masked load may be accrued in a large population without harmful phenotypes being expressed for long periods of time, leading to populations possessing a large latent masked load.

1.10 The dark history of genetic load

It is crucial when investigating and discussing the effects of mutational load on a species, to understand and acknowledge the troubled past of the history of eugenics in the formation of

ideas of heritability and the deleterious load of genes on a population. For it is not purely as a means of measuring the effects of radiation on individuals that Muller in 1950 published his critical paper “Our Load of Mutations” in which he first used the term “mutational load” (Muller, 1950). Muller, along with many of his contemporaries, was a lifelong believer in the eugenics movement.

From the mid-nineteenth century, the idea that heritable traits were responsible for a variety of nervous and mental disorders became widespread, despite no understanding of the mechanism of hereditary transmission (Teicher, 2018). This mainly focused on “degeneracy”, the idea that individuals inherited traits relating to crime, poverty, alcohol dependency, and mental illnesses (Carlson, 2001). This led to Francis Galton’s proposed term “eugenics”, with the aim to control human evolution by redirecting it towards a “better path”. This was proposed to be achieved by limiting the reproduction of those deemed “less fit”, whilst increasing the reproductive output of individuals thought to be of a “higher quality” (Teicher, 2018). These theories became popular globally during the early 20th century creating the foundations of the American eugenics movement as well as with the work of Wilhelm Weinberg (known for the Hardy-Weinberg equilibrium), who assessed the hereditary risk of mental illness (Weinberg, 1912). In Germany, research into the consequences of marrying into families with mental illness was institutionalised in 1917, and by the 1920s German psychiatrists were advising their patients not to reproduce. After Adolf Hitler’s election in 1933, the “Law for the Prevention of Hereditarily Diseased Offspring” was passed, resulting in hereditary tribunals for citizens with “mental or physical defects”. Current estimates suggest there were 436,000 court cases between 1934 and 1945, resulting in the sterilization of 300,000 individuals (Benzenhöfer & Ackermann, 2015). This figure does not include the >70,000 mentally and physically disabled people murdered as part of the Nazi “euthanasia” programme during the Second World War (Burleigh, 1994; Proctor, 1988). In a post-war world Muller, Crow, Dobzhansky and Wallace debated and transformed the foundations of genetic mutational load on a theoretical basis. They modelled the changes in the frequency of genotypes and the resulting phenotypes rather than focusing on the individuals and the traits or diseases that they possessed.

1.11 The past, present and future of human inbreeding and genetic load

Throughout human history, the risks and deleterious effects of inbreeding with close family relatives has been understood long before the knowledge of the causes for such traits were discovered. Inbreeding avoidance within humans has been long theorised with the hypothesis that proximity to kin during development leads to sexual aversion known as the "Westermarck effect" (Westermarck, 1922). Studies on the identification of facial morphology of kin suggest that females prefer non-kin familiar faces, whilst males showed a preference for individuals resembling kin (Marcinkowska et al., 2013). This study suggested that facial identification is a biological deterrent to inbreeding depression in females, where the cost associated with inbreeding may be greater. Inbreeding avoidance may also be influenced by behaviours developed due to the benefits of increasing genetic diversity, such as olfactory stimuli, which have been suggested to influence mate choice (Tizard and Skow, 2021). The classical study of the "T-shirt test" found that pheromones in sweat were most attractive to females with the most diverse major histocompatibility complexes (MHC) (Wedekind et al., 1997). Offspring with increased diversity of MHC genes would have a better immune response to disease, increased fitness, and thus preferentially selected for during mate-choice in humans.

Despite the adaptation of these behaviours, inbreeding and consanguineous marriages have been common throughout history in many civilisations globally. This was especially common in aristocratic groups, where land rights were passed to one's offspring. Thus, intra-family marriages were financially and politically advantageous to ensure that power, prestige and control of kingdoms and empires remained within family groups. There have been many examples of this throughout the ancient empires such as the 18th Dynasty of Egypt, where Pharaoh Tutankhamun's parents were siblings (Hawass et al., 2010), and the Ptolemaic dynasty's repetitive sibling marriages (Ager, 2005). One of the most notorious families for consanguineous marriage was the Hapsburgs. In the span of 300 years (1450- 1750), 67% of marriages within the family had a higher inbreeding coefficient than between second cousins (Ceballos & Álvarez, 2013). This led to the inheritance of phenotypic traits such as the "Hapsburg jaw", but also resulted in the eventual extinction of the Spanish line as Charles

II was infertile, as well as potentially suffering multiple genetic diseases (Alvarez et al., 2009).

We may consider concerns of inbreeding depression to be historical issues arising from powerful medieval families attempting to ensure they reattained their power over Europe and the wider world, however, they are by no means uncommon. Charles Darwin himself worried about the effects of his own marriage to his wife Emma Wedgewood, who was his first cousin (Álvarez et al., 2015). Inbreeding within the Darwin-Wedgewood family was prevalent, and it has been suggested that it is the cause of the high levels of infertility within the male line of the family, with Darwin's sons having lower rates of fertility when compared to the rest of the Wedgewood family (Álvarez et al., 2015). In the modern day, understanding the effects of frequent consanguineous marriages is extremely important due to the association with infant mortality, congenital birth defects, cardiovascular risks and many complex genetic disorders (Badaruddoza et al., 1994; Tadmouri et al., 2009; Fareed and Afzal, 2016). A comprehensive report on fertility and mortality, in the context of inbreeding and sociodemographic factors in India was conducted (Fareed et al., 2017). This investigation found that inbred populations often had increased infant mortality and high homozygosity for autosomal deleterious genes, as well as increased fertility, theorised as "reproductive compensation" for high infant mortality (Fareed et al., 2017).

The potentially harmful impacts of genetic load from increased inbreeding in populations have been of high concern across the Middle East, where the rate of consanguineous marriages is estimated to be between 20 and 50% of all marriages (Fareed and Afzal, 2014). Reports suggest that the financial costs of common genetic disorders across the Arab world account for USD 13 billion annually (Alhosain, 2018). It has therefore been proposed that premarital genetic screening could be implemented to advise couples on the potential risks of any genetic disorders. However, the utilisation of freely available screening in Oman is low, and the recorded willingness to access such services was low among individuals with a family history of hereditary disease and consanguinity between parents (Alkalbani et al., 2022). However, the benefits to public health from such screenings should not be used to infringe on the human rights of individuals, with particular attention to Article 23 of the

International Covenant on Civil and Political Rights which states: “The right of men and women of marriageable age to marry and to found a family shall be recognized” (General Assembly resolution 2200A (XXI), 1966).

Similarly to the advances to enable genetic screening for diseases, the advances in modern reproductive health have allowed for assisted reproductive technology (ART) which in 40 years has facilitated the conception of over 8 million babies (Wennerholm and Bergh, 2020). Whilst an extremely important tool for people who have struggled with reproduction to have a family, such methods may lead to an increase in genetic load of the offspring due to the reduced selection of zygotes (van Oosterhout, Marcu, et al., 2022). The ethical and scientific debate surrounding genetic screening and the use of ART in the context of humans and animals is extremely complex and beyond the scope of this thesis. Whilst it is important to understand and communicate the risks associated with inbreeding depression and reduced selection pressures, it is paramount to carefully consider the ethical implications and to put the rights of the individual first and foremost.

1.12 Reintroductions and genetic rescue

1.13 Genetic rescue

One of the most promising applications of conservation genomics for the long-term survival of endangered species is through genetic rescue. Genetic rescue is a means of gene flow (see section 1.4 of this thesis) and can be seen as the increase in a population's fitness due to the introduction of genetic variation that had previously been lost from the population (Whiteley et al., 2015; Waller, 2015; Frankham et al., 2017; Bell et al., 2019). Novel genetic variation can increase the evolutionary potential of populations, enabling them to better respond to environmental change (Frankham et al., 2017; Hoffmann et al., 2021). In addition, novel genetic variation can help to mask recessive deleterious mutations that have become fixed in the recipient population (Tallmon et al., 2004; Waller, 2015). However, there are also risks associated with genetic rescue, including the introduction of harmful genetic

variants (Bertorelle et al., 2022) that could reduce the fitness of individuals in the recipient population.

Since 2015, there have been a number of studies focusing on genetic rescue for conservation (Bell et al., 2021; Fitzpatrick et al., 2023; Robinson et al., 2021; White et al., 2023), with the focus on much of this on the increased viability of the population (Frankham et al., 2017), or on the population growth rates due to its association with persistence probability (Bell et al., 2019). Genetic rescue is a means of manipulating the gene pool of a threatened species to provide it with new allelic variation that allows it to adapt to future threats (Samuel et al., 2020). The success of genetic rescue also depends on the genetic load of the source and recipient (i.e., target) population (Hoffmann et al., 2021). Genetic rescue can increase the fitness of individuals by masking the effects of recessive deleterious alleles that form part of the masked load (Bertorelle et al., 2022). It thereby reduces the realised load (Speak et al., 2024) and alleviates inbreeding depression. When used effectively, controlled reintroduction programmes can act not only as a means of combating low genetic diversity within isolated populations, but also to rescue small, inbred populations through masking the harmful effects of recessive deleterious mutations. Here and throughout reintroduction programmes are defined as the release of individuals either raised or rehabilitated in captivity into their natural environment, to stabilise, reestablish, or increase in-situ populations that have suffered significant declines (AZA, 2024). Despite over a decade of publications which overwhelmingly support the positive impact of genetic rescue for animal conservation, there have been a mere twenty or so genetic rescue programmes focused on conservation efforts (Adams et al., 2011; Frankham, 2015). Why then have reintroductions of individuals and alleles using captive breeding programmes as a resource been significantly underutilised in the field of conservation?

This hesitancy to widely adopt genetic rescue may be due to three distinct possible risks. The first perceived risk is that of outbreeding depression, whereby genetic incompatibility between the source and recipient gene pool reduces the fitness of offspring derived from inter-population crosses (Waller, 2015). In addition, gene flow may increase the genetic load of rare deleterious alleles of the recipient populations (Bell et al., 2019; Bertorelle et al.,

2022). Furthermore, the genetic variation of new immigrants may replace the variation in the local gene pool, undermining local adaptation. This phenomenon can lead to the extinction of a unique conservation unit or population, and it is sometimes referred to as genetic swamping (Moerman et al., 2020).

1.14 Outbreeding depression

Outbreeding depression, or reproductive isolation, is a term that is used to describe a reduction in reproductive fitness of a population after gene flow due to either pre or postzygotic isolation, or a combination of both (Frankham et al., 2011). This is often due to: gene flow occurring between different species; populations with fixed genetic differences; populations being adapted to different environments; the two populations being isolated from each other for a long time; or a combination of these situations (Frankham et al., 2017). The harmful effects of outbreeding depression are sometimes evident within the F1 generation but are, in most cases observed within the population by the F3 generation (Frankham, 2015; Frankham, 2016). Some effects result in viable sterile F1 progeny such as hybridisation between horses and donkeys. Despite some hybrid vigour (Proops et al., 2009), breeding results in a sterile F1 generation due to a difference in chromosome number (31 in horses and 32 in donkeys) that leads to multiple fixed chromosomal differences resulting in only nine chromosomes remaining unchanged between the two species (Yang et al., 2004; Carbone et al., 2006). A classic case study of outbreeding depression resulting from an attempted genetic rescue is in the Czechoslovakian ibex (*Capra ibex*) (Templeton et al., 1986). Individuals from genetically distinct subpopulations were mixed resulting in the hybrid offspring from these crosses being born early, this caused a decrease in birth rates and eventually the loss of the population.

The risks associated with the effects of outbreeding depressions on genetic rescue attempts have been greatly reduced through the implementation of screening (Frankham et al., 2011). This screening process uses a decision tree to determine if the mixing of two isolated populations would potentially increase the risk of outbreeding depression. This decision tree is based on five questions:

Question 1. Has the taxonomy of the species has been resolved? To reduce gene flow between different species crosses of this kind are most likely to result in outbreeding depression. It is therefore critical to first determine if the two populations are of the same species, separated and how each is classified.

Question 2. Are there fixed chromosomal differences? This can be differences in ploidy, translocations inversions, and centric fusions (Frankham et al., 2017). Differences in ploidy within a species are found in some plant species although the crosses between individuals are harmful. Most commonly when crossing between plants of different ploidy levels the offspring are sterile (Frankham 2011). Chromosomal translocations occur when segments of non-homologous chromosomes are swapped during meiosis. The resulting gametes are therefore chromosomally unbalanced causing the zygotes to be inviable. Inversions are when segments of a chromosome break off and reattach within the same chromosome but are now re-orientated 180^{o}, causing a modest reduction in gamete viability. Centric fusions are when two non-homologous chromosomes are fused at their centromere, this results in unbalanced gametes (Baker and Bickham, 1986). Issues can occur when gene flow between populations with different centric fusions occur, referred to as monobrachal fusions, which are often harmful.

Question 3. Has there been gene flow between populations within the last 500 years? This allows for the identification of unrecognised speciation between populations as well as local adaptation to environmental conditions.

Question 4. Are there any substantial environmental differences? Differences in environmental conditions can often lead to reproductive isolation, such as when crosses were made between benthic and limnetic forms of three-spined stickleback fish, resulting in reduced spawning rates (Rundle et al., 2000). This is often due to pre-zygotic reproductive isolation either through linkage disequilibrium between alleles linked to adaptive evolution

and reproduction, pleiotropic effects, or changes in timing or location of reproduction (Frankham et al., 2017).

Question 5. Has the population been separated for over 20 generations? This is to reduce the chances that the two subpopulations have become reproductively isolated, although when the environmental conditions are similar between isolated populations there have been many cases of no reduction in fitness up to ~6,000 generations (Rundle et al., 2000; Frankham et al., 2017)

Despite the perceived fears of outbreeding depression, for almost all reported cases of observed outbreeding depression between reproductively isolated populations, the populations recover once natural selection acts upon the genetic variation within the population (Frankham et al., 2017). The effects of this within ameliorated environments such as captive-bred populations where natural selection is less effective, are less well reported and should be avoided. Although Frankham (2015) showed that 93% of the species listed showed a low risk of outbreeding depression (Frankham, 2015), conservation managers remain worried about attempting genetic rescue due to fear of losing their conservation unit (CU) (Funk et al., 2012; Barbosa et al., 2018; Forester et al., 2022).

1.15 Rare deleterious alleles

A second major perceived risk that can result from a genetic rescue attempt is the reintroduction of rare deleterious alleles present within the rescuing population or source population. The frequency of these mutations can increase within large populations without serious negative consequences due to remaining in a heterozygote state. In this way, they remain part of the masked load and are not converted to the realised load (see section 1.9). When a species undergoes a population bottleneck some of these deleterious alleles can become fixed due to chance due to genetic drift (Charlesworth, 2009). However, most of these alleles are lost from the population, even in the absence of purifying selection. In the absence of selection, the probability of loss or fixation is equivalent to the initial allele

frequency. Given that harmful mutations tend to be rare, most of these are lost due to drift. In addition, purifying selection removes the most severely deleterious mutations, resulting in purging and a reduction in the total genetic load.

As a consequence of genetic drift and purifying selection, isolated populations are likely to fix deleterious alleles at the different genetic loci. Therefore, during genetic rescue attempts, in an endeavour to increase the genetic diversity (see section 1.6) of the population, rare deleterious alleles that had been previously lost can be reintroduced (Whiteley et al., 2015). Once reintroduced to the population, these alleles may increase in frequency over time due to attempts to maximise the potential number of offspring from rescuing individuals. In addition, if the effective population size (N_e) of the rescued population is small, the efficacy of purifying selection is much reduced, which means that slightly deleterious mutations (with $s<1/(4\ N_e)$) can drift and increase in frequency in the population (Wilder et al., 2023; Dussex et al., 2023). Alleles may also increase in frequency due to genetic hitchhiking (Gillespie, 2000; Smith & Haigh, 1974). This is where potential fitness benefits to the F1 offspring are conveyed by other reintroduced loci, meaning that rare deleterious alleles increase in frequency within the population. Once these deleterious alleles have accumulated to a high enough frequency, they are more likely to become expressed in homozygote condition, forming part of the realised load (Mathur and DeWoody, 2021; Bertorelle et al., 2022). This leads to a decrease in the fitness of the population. Despite the potential threat to the population, the risks, if measured and controlled, can likely be mitigated.

In natural populations, purifying selection can remove harmful mutations with a large effect on fitness, i.e., $s>1/(4N_e)$ (Dussex et al., 2023). Once the deleterious alleles are expressed in a homozygote condition, they will become exposed to natural selection and purged from the population (Hedrick, 1994; Pérez-Pereira et al., 2022). However, this may take a long time to achieve and is most effective in small populations (van der Valk et al., 2021) where the chances of homozygotes occurring is increased. On the other hand, small N_e reduces the efficacy of purifying selection because of more intense genetic drift (Wilder et al., 2023). Therefore, the best potential means to combat the risk of increasing the genetic load is to limit the number of deleterious alleles reintroduced through conservation-genomic

management (Speak et al., 2024; Jensen et al., 2024). This can be done by implementing genetic screening of the individuals to be introduced, as well as the threatened population. In this way, the number of deleterious alleles reintroduced can be minimised. Such screening using impact scores derived from Combined Annotation Dependant Depletion (CADD) (Kircher et al., 2014; Rentzsch et al., 2019) (this will be discussed further in section 1.22 of this thesis) scores, can determine and rank how deleterious potential mutations may be. These values can then be used to ensure specific mutations are not reintroduced into threatened populations during genetic rescue programmes, to minimise the risk of increasing the genetic load (discussed further in section 1.7 of this thesis).

1.16 Source population size

When designing a genetic rescue program, it is important to consider not just the demographic and genetic history of the threatened population but also that of the source population being used for the rescue (Whiteley et al., 2014). There are two contrasting views on the optimal population size to use as the source for genetic rescue programmes. Is it more effective to use a historically small population that has undergone the purging of deleterious alleles (Kyriazis et al., 2021; Robinson et al., 2019) or a historically large source population with high genetic diversity (Ralls et al., 2020).

Studies suggest the optimal method of supplementation is to use individuals from a large population source which has undergone a lot of mixing (Ralls et al., 2020). The theory being genetic rescue aims to alleviate the risk of extinction caused by a low genetic diversity within the threatened population. Therefore, to reduce the risk of inbreeding depression, the population should be supplemented with individuals with a high genetic diversity. This boosts the population's genetic diversity, viability, and future adaptive potential to environmental changes and genetic drift (Hoffmann et al., 2021). The limitation of this proposed method is that although the genetic diversity is increased, the large source population will have accumulated a very high genetic load if it has remained large for a long period of time (Bertorelle et al., 2022). This can result in the accumulation of a large number of recessive deleterious alleles, hidden from selection as the masked load (DeWoody et al., 2021). However, when these mutations are introduced to the new far smaller population, the

chance that these mutations are expressed in the homozygous state increases, leading to an increase in the realised load and thereby causing a reduction in both individual fitness as well as population viability.

A recent example of this is in the Scandinavian wolf (*Canis lupus*), where descendants of immigrants from a genetic rescue programme had 844 deleterious mutations not previously found within the population (Smeds and Ellegren, 2023). These risks are elevated in genetic rescue programmes due to the unequal distribution of mate pairs due to an overrepresentation of introduced individuals (Robinson et al., 2019). Despite this, proponents of this method argue that, in observed cases, reintroduction programmes using individuals from large populations free from inbreeding are more effective when compared to reinforcement with individuals from smaller populations (Frankham, 2015; Ralls et al., 2020). They also suggest that supplementations of this type would require fewer and less frequent translocations. This makes their application more feasible for many conservation programmes where translocations are a costly endeavour both in their monetary expenses as well as the large amount of time, resources, and personnel required.

Alternatively, to attempt to avoid the risks of increasing the genetic load of the threatened populations, it has been proposed to reintroduce individuals from historically small populations (Kyriazis et al., 2021; Robinson et al., 2019). These populations, due to their consistent small size, will have undergone purifying selection (Hedrick, 1994; van der Valk et al., 2021; Khan et al., 2021; Dussex et al., 2023). In this way the masked load within these source populations will be reduced as deleterious alleles will have been exposed to selection in the homozygous state and purged from the population (Bertorelle et al., 2022). Therefore, these populations would act as very safe source populations as the risk that they will introduce rare deleterious alleles is far lower. But due to the small size of these source populations, they are expected to have a lower genetic diversity. In this way, reintroduction programmes using individuals from these populations will not have as substantial an impact on increasing the genetic diversity of the threatened populations and could leave them vulnerable to future genetic drift (see section 1.4). This means that to increase the genetic diversity multiple expensive translocations of individuals may be required and repeated

regularly to maximise their effects (White et al., 2023). If using a historically small population, the length of time that the species has been bottlenecked is extremely important to consider (Bortoluzzi et al., 2020). If a population has been bottlenecked for a long time, then purging of deleterious alleles will have been more successful. However, populations with a recent (sharp) bottleneck should be avoided as these are likely to hold a high genetic load as well as a low genetic diversity (Pérez-Pereira et al., 2022).

This leaves potential conservation managers who wish to implement a genetic rescue programme with a conundrum. Should they increase genetic diversity at the risk of reintroducing deleterious alleles? It is here that the application of conservation genomics can be most beneficial. Genetic screening of both the threatened and source populations allow for the identification of harmful deleterious variants present as well as shared masked load and identifies pairings that should be avoided (Speak et al., 2024). Individuals could also be selected that would result in the largest increases in the genetic diversity within the threatened population, meaning that the species would be more resilient to future genetic drift or environmental changes. It should also be considered that genetic rescue programmes should be reciprocal, with movement between populations being in both directions (White et al., 2023). In this way, both populations can be managed as one allowing for an increased genetic diversity for both populations together. This will also mitigate any potential negative effects caused by reducing the population size of the source populations (Dennis et al., 2016). This is particularly important in circumstances where the source population is also small, or when using individuals from captive breeding programmes and zoos.

1.17 Genetic swamping

Genetic rescue using a captive source population could result in gene swamping within the rescued population. Gene swamping (Haldane, 1956) is a theory that gene flow between populations (commonly from the centre of a species range into populations at the edge), may inhibit local adaptation at the edge ranges, limiting range expansion (Kottler et al., 2021). This can be problematic issue as it not only limits the potential adaption of the species

to fragmented and changing environments, which is often caused by human activity (Dirzo et al., 2014), but also can result in the loss of these edge populations. This is an issue as edge populations are often useful reservoirs for genetic diversity that could further species recovery (Hampe and Petit, 2005). The risks surrounding genetic swamping and genetic rescue in edge environments are multifaceted, as preserving genotypes within peripheral populations may result in them having a higher genetic load (Angert et al., 2020). Hybridisation between feral domesticated and undomesticated rock dove (*Columba livia*) initially improves the genomic diversity of isolated populations (Smith & Clegg, 2023). However, gene swamping may lead to the eventual replacement of "wild" genotypes causing an overall reduction in genetic diversity (Smith, 2023). As such, hybridisations should be limited to ensure the unique diversity of the wild population is maintained. Therefore, genetic rescue programmes involving populations in edge environments should undergo careful consideration. This is with the aim of balancing the benefits of preserving unique adaptations that can maximise the potential for range expansion or recovery, against the heightened risks of increasing genetic load.

1.18 The role of zoos in conservation

1.19 Zoo populations and the One Plan approach

Many species globally have insurance populations both in captive breeding programmes and within zoos around the world. These populations have the potential to be an excellent resource for conserving individual species and importantly genetic diversity (Al Hikmani et al., 2024; Jackson et al., 2022; Lott et al., 2020). This provides a vital resource for potential genetic rescue attempts for populations, as a pool from which conservationists can supplement endangered or recovering species. Despite many in the field calling for increases in genetic rescue attempts (Ralls et al., 2018), there has been a limited number of publications attempting this, using techniques such as outcrossing and assisted migration for genetic rescue (Frankham, 2015; Fitzpatrick et al., 2023). Nevertheless, in the relatively small number of instances in which genetic rescue has been attempted, the overall effect of fitness and population viability were positive (Frankham, 2015; Whiteley et al., 2015;

Frankham, 2016). Examples include the Florida panther (Johnson et al., 2010), whereby eight wild-caught female Texas pumas were translocated to the Florida Everglades. This resulted in a threefold increase in panther numbers, a doubling in heterozygosity and increases in fitness and survival rates (Johnson et al., 2010). Frankham (2015) reviewed the impact of an assisted migration dataset in 156 wild species across 77 taxa that were suffering from low genetic diversity or inbreeding depression, to determine how consistent the effects of outbreeding were. This comprehensive review not only showed the range of benefits of outbreeding in previously published investigations but also highlighted the perceived issues that face the widespread implementation of genetic rescue (Frankham, 2015).

The primary aim of the Frankham review was to solve the issues of limited quantitative information on the consequences of outbreeding (Frankham, 2015). The conclusions of which were overwhelmingly beneficial across a wide range of fitness effects including fecundity, survival, population persistence and growth rates. Outbreeding benefited 145 of the species, whilst only nine suffered negatively. However, these nine were likely as a result of a combination of low statistical power, benign environments, or due to the study species being naturally inbred or from an inbred founder population. Only one species showed convincing signs of outbreeding depression. The composite fitness of species was 148% higher after outcrossing in stressful conditions and 45% higher in benign (Frankham, 2015). This indicates that outcrossing is most beneficial when applied to wild species in stressed environments, such as those of many endangered populations (Jackson et al., 2022; Wilder et al., 2024), but there are benefits to outcrossing inbred domestic (Smith & Clegg, 2023) or captive-bred populations (Howell, Frankham, et al., 2021) to increase the fitness of species. One of the most intriguing results of Frankham's 2015 paper was that for six species the effects on the evolutionary potential of the species were monitored, showing that in all cases genetic rescue was beneficial(Frankham, 2015). This is in line with other studies of intraspecific gene flow between populations that show that it is possible to conserve the evolutionary potential of a species and provide it with a resource to adapt to changing environments (Hamilton and Miller, 2016).

1.20 The history of zoos

Collections of animals have been found in ancient Mesopotamian and Chinese civilizations, and the oldest known zoo is thought to be that of Queen Hatshepsut II, founded around 1490 BCE (Foster, 1999). These collections were used by ancient rulers to display wild animals from the lands they controlled, thereby demonstrating their power and prestige. Similarly, the Roman Empire would capture wild animals from across its empire to display in the Coliseum or for hunts. During the medieval period, the practice of keeping exotic animals evolved into the menageries of many European monarchs (O'Regan et al., 2006). One notable example is the menagerie founded by Henry III of England at the Tower of London(Parnell, 1999). It is reported to have been founded after the gift of three lions (listed at the time as "leopards") from the Holy Roman Emperor to the King in honour of his coat of arms. Later, the collection was expanded to include a polar bear, gifted by the King of Norway. The menagerie became home to a variety of animals over the centuries until it was moved to Regents Park in 1831 (Kisling, 2000). In the 1930s, archaeological excavations around the 'Lion Tower' revealed the skulls of two barbary lions (Barnett et al., 2008), which are now extinct, dating back to around the 14th and 16th centuries respectively (O'Regan et al., 2006). In the 18th century, aristocratic families throughout Europe began private collections of animals, and the first modern zoos opened to the public during the French Revolution when collections from all over Paris had been confiscated and amalgamated (Lindholm, 2013; Miranda et al., 2023).

Modern zoos have their origins in the private collections of Western European and American empires. However, during the 20th century, many began to recognize the potential of using these collections for conservation purposes (Roe et al., 2014). In the novel 'The Stationary Ark' (Durrell, 1977), Gerald Durrell aimed to inspire the creation of a new type of zoo that would focus on the biological study of animals and plants. This zoo would aid endangered species through the biological study of endangered species, breeding and reintroduction programmes, while also serving as a means to display to the public the need for conservation of our natural habitats. Durrell's vision was confident in its assertion that such a zoo would be a valuable tool in the fight to preserve our planet's biodiversity. These key values represent three main aspects of modern zoos. They are always paired to attract the

public to generate revenue and ensure the zoo's sustainability. Additionally, they serve as a showcase for animal welfare and management. Throughout their extensive and illustrious history, zoos have faced numerous criticisms and remain under constant scrutiny from the public and media alike (Kreger and Hutchins, 2010; Pierce and Bekoff, 2018). Here I will outline four of the main aspects for which zoos are scrutinised concerning their involvement in the conservation efforts of threatened species.

1.21 Critiques of modern zoos

Firstly, there are ethical concerns relating to keeping captive-bred populations on display as attractions for the public (Kreger and Hutchins, 2010). This is a problem that all zoos face currently, with many of their exhibits (especially those most often used for promotional material) being the descendants of animals caught during the 19th and 20th centuries by imperialists (Lindholm, 2013; Miranda et al., 2023). This sordid past of the collections and historic environment that some of these animals were kept in is in stark contrast to the way that modern-day zoos operate, with strict guidelines for the welfare and upkeep of the animals (Mellor et al., 2015; Kagan et al., 2015). Despite this, as a public attraction, the ethics of maintaining wild animals in captivity is often an emotive topic, especially when animals are kept far from their natural environment.

The use of zoos for their captive breeding programmes also creates ethical concerns surrounding the right to control the reproduction and deaths of animals. These concerns are very emotive and often focused on the individual welfare of specific animals. An example of this was the media response around Marius the giraffe killed in Copenhagen Zoo in 2014 (Cohen and Fennell, 2016). Marius was deemed by the managers of the European giraffe population to be surplus to requirement as his genes were being over-represented in the breeding population. This is managed due to the risk of inbreeding depression within captive populations. Therefore, as a non-breeding “surplus” individual, keeping Marius was not economically feasible for the Zoo. Many alternative methods to culling were considered but seen as unethical due to the inherent stresses and limitations that Marius would face. These included the risks from sedation to giraffe associated with procedures such as castration

and contraception which were therefore seen as too high risk (Cohen and Fennell, 2016). Furthermore, due to his acclimatisation to captive breeding, he was unfit for release (Schulte-Hostedde and Mastromonaco, 2015). It was therefore decided that the appropriate action was to euthanise Marius. This garnered almost instant media backlash and criticism for an action the zoo saw as a sad but necessary part of managing the large number of surplus animals produced by captive breeding (Bandoli and Cavicchio, 2021). This negative response may have been bred from the disparity between the way that zoos use young animals as part of their advertising campaigns with terms such as “adopted” (Cooke, 2011) and portrayed and humanised (i.e. with names) which evoke emotional responses from the public. In the case of Marius only then to be culled once they are no longer economically viable (Cohen & Fennell, 2016). Despite these criticisms, zoos are often global advocates for animal welfare, with global welfare efforts for animals regulated by accrediting bodies such as the World Association of Zoos and Aquariums (WAZA) and the Association of Zoos and Aquariums (AZA) (Kagan et al., 2015) and develop strategy guidelines such as the WAZA Caring for Wildlife: The World Zoo and Aquarium Animal Welfare Strategy (Mellor et al., 2015). Therefore, maintaining zoo populations could be seen as a constant ethical tightrope balancing the rights of the animals to enact wild behaviours such as mating and raising young, with the risks associated with managing the numbers and genetic health of captive populations.

Secondly, there are concerns that many species do poorly when kept in captivity. Many species in captive breeding programmes breed well and suffer no negative consequences from captivity, often benefiting from the constant keeper and veterinary attention. However, this is not true of all species, with many unable to establish self-sustaining populations within captivity with some suffering “boom-and-bust cycles” (Jones, 2015). Whilst other critics suggest that this is due to keeping charismatic animals which do not thrive in a zoo setting (Pierce & Bekoff, 2018). Many other species such as orca (*Orcinus orca*) suffer from behavioural issues within captive breeding environments and can become aggressive to other individuals or keepers (Anderson et al., 2016), as well as suffer mental stress without their natural environmental interactions (Marino et al., 2020).

Thirdly, there are the risks that the population may over time become domesticated. This can be a result of a combination of factors, especially if they are enacted over a prolonged period with multiple generations of captive breeding. Within captive breeding environments designed or inadvertent selective breeding can take place (Schulte-Hostedde & Mastromonaco, 2015). It can be helpful for zoos to have animals with traits that make them easy to manage, such actions could contribute to selection against aggression (Norscia et al., 2024) or self-domestication (Hare et al., 2012). This can mean that the animals can be kept in closer proximity to other individuals or species without causing harm to themselves or other animals. Whilst other species or individuals may be preferentially selected for their ability to withstand auditory stress from the surrounding environment (Morgan & Tromborg, 2007; Rose et al., 2022). This can also be beneficial to zookeepers who can feed and clean in a safer environment, whilst reducing the stress upon the captive animals. Alternately, captive breeding can apply unwanted selection pressures on morphological traits (O'Regan & Kitchener, 2005) and reproductive output (Schulte-Hostedde & Mastromonaco, 2015). This may result in populations that are biased to descendants of individuals who were able to breed year-round (Frankham, 2008), were most amicable with keepers or those with the largest reproductive output. Whilst some of these behaviours can be seen to have perceived benefits, they will result in more domesticated individuals who are unsuitable for reintroduction programmes as they may no longer survive in their natural habitats or have reduced fitness (Christie et al., 2012; Frankham, 2008).

It has been suggested that zoos should endeavour to breed individuals to maintain "wild" traits within populations to allow them to be most appropriate for reintroduction programmes. However, the interpretation of "wild" behaviours, traits or phenotypes is not straightforward and requires extensive research and study of true wild populations, which is not feasible for many species. An example of this issue can be seen in the conservation of the takhi or Przewalski's horse (*Equs przewalski*). To ensure captive individuals maintained truly wild phenotypes, hybridisations with other horse species and captive takhi were avoided. More than 1500 individuals were bred in captivity from just nine capable breeders (Turghan et al., 2022). Recent studies on Przewalski's horse genetics uncovered ancient introgression or hybridisation with domestic horses (Goto et al., 2011), as well as a relatively high genetic diversity for a species that went through a population bottleneck to just nine individuals. This

raises questions surrounding whether to separate pure and non-pure (those with recent domestic hybrid ancestry) populations from breeding and how this will influence the long-term conservation of founder genetic diversity. The characterisation and preservation of undomesticated lineages are a critical aspect of maintaining both genetic diversity and ecological functions (Smith et al., 2022).

Concerns have similarly been raised regarding temperature adaptation and melatonin availability within zoo populations (Fa, 2011; Schulte-Hostedde & Mastromonaco, 2015). Although captive breeding managers attempt to recreate species' natural habitats (i.e., through heaters or lamps), it is hard and costly to replicate the natural photoperiod of many species in captivity. It has been suggested that photoperiods influence melatonin secretion and circadian rhythms may influence reproductive cycles, which could explain the limitations in some species breeding success (Schulte-Hostedde & Mastromonaco, 2015). These effects may also be magnified by the ameliorated environment that has allowed genetic disorders or traits to survive in the population without the selection pressures provided by the natural environment (Trask et al., 2019). Thereby, individuals that may not have been able to reproduce naturally are represented (or over-represented) within captive populations.

A further criticism of zoos is that they may only be focused on conserving charismatic animals and not species that are endangered (Spooner et al., 2023). Zoos have limited space, which is preferentially used for species most likely to increase visitor numbers and thus profits, such as large historic species including lions, tigers and giraffes (Mooney et al., 2020). Therefore, species that may potentially be at a higher risk of extinction are not protected. It has therefore been shown that the current global zoo population is not an accurate representation of global threatened species (Conde et al., 2013). Due to the rising ethical concerns expressed by the general public, zoos will need to shift to displaying a greater number of smaller species such as amphibians, reptiles, small mammals and birds which would be more space-efficient (Spooner et al., 2023). Amphibians (41%) (IUCN 2024) in particular are more threatened globally than many large charismatic species. However,

the financial implications of such changes could hamper the widespread application of such changes.

Finally, critics believe captive breeding is too expensive and the money that is spent on maintaining the populations could be better served in conserving the natural environments of the species. They envision this as the best means to protect the species as a whole rather than maintaining captive-bred populations as an insurance against extinction (Miranda et al., 2023). Although most zoos do make financial contributions to conservation, it is often around 5-10% of their income (Gusset & Dick, 2011), a figure which makes them the third largest donators globally, this could be viewed by critics as greenwashing or "window-dressing" (Keulartz, 2015). Most of these profits are reinvested into the upkeep and development of the zoos themselves (Spooner et al., 2023). However, if global zoo populations are to be truly seen as insurance populations against species extinction and implemented as part of the One Plan Approach (Traylor-Holzer et al., 2019), combining *in situ* and *ex situ* populations for conservation, this development of their facilities could be seen as a conservation financial investment. Whilst zoos are indeed some of the largest tourist attractions internationally, with previous studies suggesting at least 700 million global visitors annually (Gusset & Dick, 2011) the effects of the covid pandemic highlighted how fragile zoos are financially. During the COVID-19 pandemic, zoos were forced to close to the public which halted their income, but the daily running costs of housing, keeping and feeding the species they held remained (Spooner et al., 2023). This situation meant that some zoos needed to resort to public donations such as the "Save our Zoo campaign" for Chester Zoo which raised over £3 million to cover the £1.6 million monthly cost of caring for 350,000 animals (Chester Zoo, 2020). This situation, although showing the support that the public still has for zoos, does highlight the risks that zoos potentially face in the future. They are foremost public attractions that rely upon a steady stream of visitors to fund the animal upkeep and welfare before any conservation efforts can be considered.

1.22 Mutation impact scores

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, CADD scores were developed in humans (Kircher et al., 2014), and then 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, 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% scores 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.

1.23 Ultraconserved elements

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 200 bp 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, i.e., 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), while 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 intra-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.

1.24 Pink pigeon conservation

Figure 2 – A free-living pink pigeon (*Nesoenas mayeri*) in the Black River Gorges National Park, Mauritius. Photograph by Tom Everley.

The pink pigeon (*Nesoenas mayeri*) is a member of the Columbidae family and is native to the tropical island of Mauritius (Figure 2). The wild pink pigeon population has suffered a gradual decline over the past three centuries (Jones et al., 2013; Swinnerton et al., 2004).

More recently, the population experienced a demographic bottleneck, collapsing to only nine or ten individuals in the early 1990s (Swinnerton et al., 2004). The main factors influencing the decline included native habitat degradation, fragmentation and the introduction of invasive species (Swinnerton et al., 2004). Eventually, this pushed the population into an extinction vortex from which it only managed to recover through intensive conservation management by Carl Jones, the Mauritius Wildlife Foundation (MWF) and National Parks and Conservation Service (NPCS). In the 1970s, a captive population was established at the Gerald Durrell Endemic Wildlife Sanctuary (GDEWS) in Mauritius, initially taking 11 individuals from the population at Pigeon Wood. The GDEWS gene pool was intensively managed to conserve genetic diversity (Figure 1, Grey Arrow). This was achieved through captive breeding with supplementation from the wild and captive-bred populations (Jackson et al., 2022). As a result of the intensive breeding and conservation efforts, the wild metapopulation is stable at over 400 – 500 individuals (Jackson et al., 2022; Jones et al., 2013). These efforts have resulted in the species being downlisted on the IUCN's Red List from 'critically endangered' to its current rating of 'vulnerable' (BirdLife International, 2021). However, despite these conservation efforts, the pink pigeon displays signs of inbreeding depression (Jackson et al., 2022; Swinnerton et al., 2004). This has resulted in reduced survivability and hatching success, likely as a result of the severe population bottleneck resulting in the expression of the genetic load (see section 1.7 of this thesis for further information).

1.25 Whooping crane conservation

The whooping crane (*Grus americana*) is a species of migratory crane native to North America currently listed as Endangered in the IUCN red list (BirdLife International, 2020). Prior to European settlement in America, the whooping crane is thought to have been abundant and it was estimated that the population of whooping cranes in the 1800s was approximately 1500 individuals (Allen, 1952). During the early 20th century, the species was pushed almost to extinction due to overhunting (Golden et al., 2022) with the population declining to between 16-20 individuals in the wild during the 1930s and 40s (Butler et al., 2013; Golden et al., 2022; Smith, 2019). This catastrophic population decline led to the

species being protected from hunting under the Migratory Bird Treaty Act of 1918. Due to the creation of the Aransas National Wildlife Refuge in 1937 to protect their vital wintering habitat (Golden et al., 2022) and in 1967 a captive breeding population, the species began to recover demographically and is now estimated to be at over 800 individuals across the wild and captive populations (International Crane Foundation, 2021). The wild population of whooping cranes, known as the Aransas-Wood Buffalo population, migrates in increasingly large groups of up to 76 individuals (Caven et al., 2020), from its breeding ground in Canada across North America to overwinter in the US Texas Gulf Coast. Currently, this population is estimated to be over 500 individuals and has remained stable over recent years (Golden et al., 2022; Smith, 2019). However, its dependence on the overwintering habitat puts the species a critical threat to habitat loss due to climate change (Golden et al., 2022). This is because much of the population overwinters within the Aransas National Wildlife Refuge and the adjacent areas. In addition to this Aransas-Wood Buffalo population, two experimental wild populations of birds were created by releasing juvenile captive-bred individuals (Smith, 2019) (Figure 3). These efforts created a migratory and a non-migratory population, in areas where historically whooping cranes had been recorded.

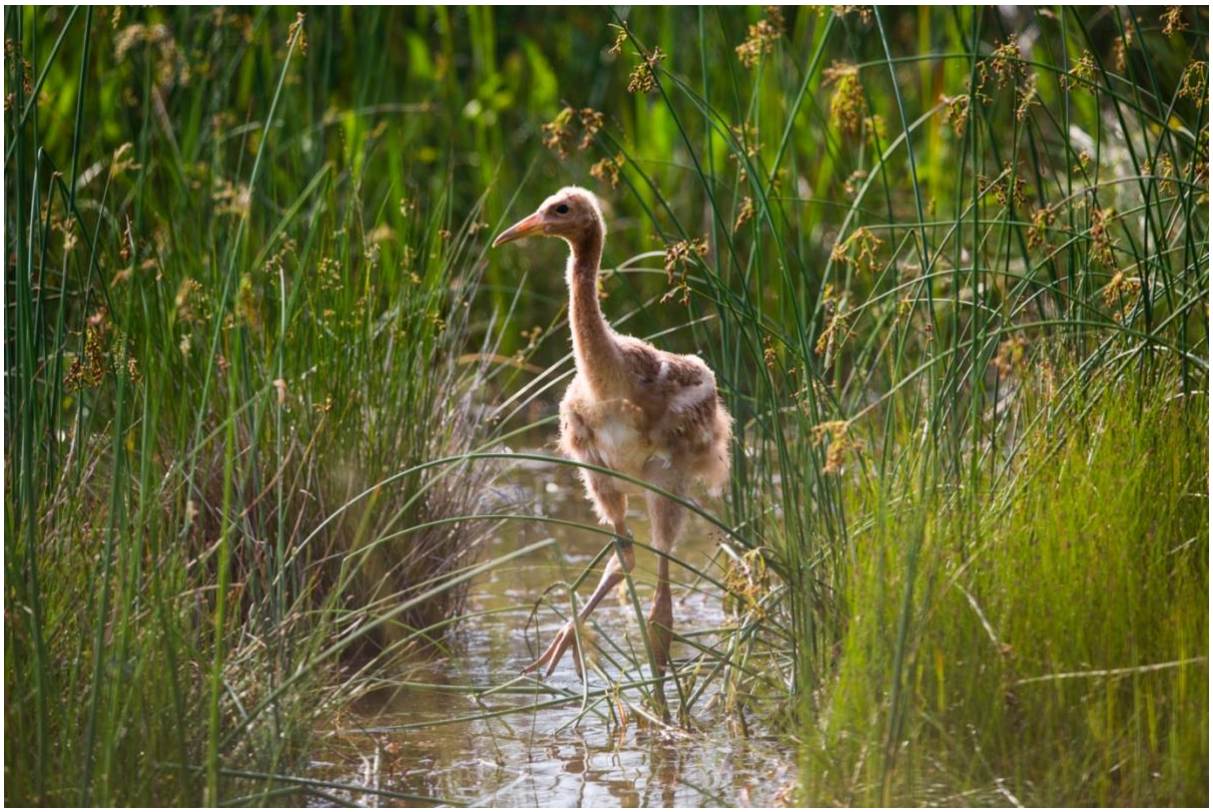


Figure 3 – Juvenile whooping crane at the International Crane Foundation. Photograph by Tom Lynn.

The Florida non-migratory population was founded between 1993 and 2006, releasing 289 whooping cranes (Folk et al., 2010). Juvenile cranes were raised in captivity using a combination of costume and parent-rearing techniques before being released. However, the population suffered from a lack of productivity caused by a combination of high adult mortality, predation by bobcat (*Lynx rufus*), powerline strikes, poor hatching success, and a possible lack of genetic diversity (Converse et al., 2019). Due to this, further planned captive releases were stopped, with the population in 2016 totalling 14 individuals. The Eastern migratory population was founded in 2001, releasing 268 juveniles between 2001 and 2016 (Smith, 2019). These individuals received flight training by ultralight aircraft to prepare them for their migration. These aircraft-guided migrants had a survival to one-year rate of 76% (Hartup, 2018). This population also suffered from poor productivity, with a nesting success of 46% (Smith, 2019), with few chicks surviving to fledge.

The whooping crane is still reported to face many threats including hunting, collisions with powerlines during migration, and predation of adults, chicks and eggs by corvids. This, combined with parasitism by black fly (Smith, 2019), may also be a cause of poor fledging success in the whooping crane populations. However, an assessment of the genetic load within the captive-bred and wild populations could investigate if poor hatching and fledging success is due to a high genetic load accrued during the population bottleneck. Population forecasts for the whooping crane have suggested that the current population growth rate is not resource-limited (Butler et al., 2013) and that the Aransas-Wood Buffalo population is likely to reach 400 individuals by 2040. Although these simulations were models assuming no changes to the habitat and the population growth rates reflecting those over the past 73 years of monitoring (Butler et al., 2013). Reduction in this growth rate due to factors such as sterilisation or low fecundity due to inbreeding depression should be evaluated to be included in future population evaluations and forecasts.

1.26 Research aims

The aims of this thesis will focus on designing a means to quantify the genetic load using whole genome sequenced data from individuals of the Mauritius pink pigeon. With this data, I will create a novel bioinformatics pipeline that will use CADD scores previously generated for the chicken genome, and apply them to calculate the genetic load within the UCEs of individual pink pigeon genomes. This will provide a means of quantifying genetic load within individuals, as well as enabling conservation genomticists to calculate the realised and masked load components that each bird possesses. This will allow for a step change in conservation by providing a framework for the analysis of the genetic load within endangered species globally, for which CADD scores are not currently available. CADD scores are currently available to be used to analyse humans, mice, pigs and chickens. However, there has not been a comparison of how similar the CADD scores between these species are. Understanding the similarities and differences of CADD scores in model species that are separated by millions of years of evolutionary history will give insights into how scores can be used once they have been transferred to non-model species. In this way, conservation managers of endangered populations will be able to make better assessments of the genetic

load that is present within threatened populations. This will also allow managers of conservation breeding programmes to assess the optimal mate pairings to reduce the effects of inbreeding depression within their populations. Using these approaches, we aim to maximise the effective use of available mutation impact scores for the assessment of threatened species globally. Thereby, maintaining the genetic health of threatened species to help ensure the species' long-term survival.

2. Development and application of LoadLift for genomics-informed captive breeding

2.1 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, by analysing the genetic load in the threatened pink pigeon *(Nesoenas mayeri)*. I lifted the mutation-impact scores that had been calculated for the chicken (*Gallus gallus*) to estimate the genetic load in six pink pigeons. Additionally, I perform *in silico* crossings to predict the genetic load and realised load of potential offspring. I thus identify the optimal mate pairs that are theoretically expected to produce offspring with the least inbreeding depression. I 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.

2.2 Introduction

Currently, 45,300 of the 163,040 (28%) species assessed on the Red List of the International Union for Conservation of Nature (IUCN) are threatened with extinction (IUCN 2024). Plans to combat this biodiversity loss were adopted by governments globally as one of the United Nations' Aichi Biodiversity Targets for 2010-20 and the 2022 Kunming-Montreal Global Biodiversity Framework (Langhammer et al., 2024), with USD121 billion annually invested in biodiversity conservation (Seidl et al., 2020). This global conservation effort is effective at improving the state of biodiversity (in 45.4% of applications) or slowing the decline (20.6%) (Langhammer et al., 2024). To help combat species loss, a small subset of threatened species have had either *in situ* or *ex situ* captive programmes established to act as "insurance populations" against their extinction (Gilbert et al., 2017). The European Association for Zoos and Aquaria (EAZA) currently maintains circa 400 species as part of EAZA *Ex situ* Programmes (EEPs) (EAZA, 2024b). These EEPs are managed with the aim

to maintain the health of *ex situ* zoo populations long-term, and they are a valuable resource for research and reintroduction programmes to establish or reinforce free-living populations.

Captive populations act to ensure that populations do not go extinct, however they themselves are threatened with the risks associated with inbreeding depression (Boakes et al., 2007). Guidelines suggest that *ex situ* breeding programmes are founded with wild individuals selected when the population is large enough to retain genetic diversity (McGowan et al., 2017). However, this has only recently become feasible to assess following the widespread implementation of genetic techniques. Captive populations are often started during population declines and from the limited number of individuals available, or from animals rescued from the wildlife trade (Owen et al., 2014). However, both approaches will lead to uncertainty about the relatedness of founders (Rabier et al., 2022). The combination of these factors and the small population size of captive-bred populations means that many species rapidly begin to show signs of inbreeding depression (Boakes et al., 2007; Rollinson et al., 2014). This occurs as harmful genetic variants increase in frequency over time, leading to a reduction in the fitness of individuals, known as the genetic load (see section 1.7 of this thesis) (Bertorelle et al., 2022). Here, the effects of a recent population decline (Jackson et al., 2022; Sachdeva et al., 2022) and population bottlenecks (Bortoluzzi et al., 2020) reduce the efficiency of selection and genetic purging to remove harmful variants. In addition, supplementary feeding and an artificially enhanced environment can also reduce the efficacy of natural selection (Hao et al., 2015), although this is partly offset by the increase in population number, which enhances the efficacy of selection (van Oosterhout pers. comm.). It can take multiple generations before the deleterious mutations become homozygous, therefore resulting in a reduction in fitness, in what is termed “drift debt” (Pinto et al., 2024). Altogether, this makes the conservation genomic assessment of the genetic load invaluable for the long-term viability of zoo populations.

Unfortunately, the risks posed by genetic load are not considered a conservation priority at present (van Oosterhout, 2020). Inbreeding depression in declining populations is a function of both the rate of inbreeding and the genetic load of recessive deleterious mutations that are 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 realised 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. Here, the genetic load of individuals can be quantified in terms of mutation impact scores (see section 1.22 of this thesis) (Bertorelle et al., 2022; van Oosterhout, 2020a; van Oosterhout et al., 2022). The Combined Annotation-Dependent Depletion (CADD) framework (Kircher et al., 2014; Rentzsch et al., 2019) produces scaled mutation impact scores for every potential mutation within the human genome relative to all variants within the genome. This has since been applied to other model species, including mouse (mCADD) (Groß et al., 2018), pig (pCADD) (Groß, Derks, et al., 2020), and chicken (chCADD) (Groß, Bortoluzzi, et al., 2020). Therefore, CADD scores provide a great opportunity to provide quantitative values to assess and compare the deleterious mutations present within threatened species.

Presently, we cannot translate the impact scores of mutations such as CADD scores into fitness effects. CADD scores equate to a type of rank score, and they are not the same as selection coefficients. 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. In this approach, we assume that the distributions of selection coefficients and CADD scores are similar between individuals. If the variances of the distributions are similar (i.e., homoscedasticity), this allows us to compare the mean (or median) of these distributions and assess the load of individuals relative to one another. Similarly, we can estimate the proportion of genetic load expressed as realised load, and the proportion whose fitness effects remain masked as an inbreeding load or masked load (Bertorelle et al., 2022). The realised load is the genetic load that leads to a reduction in fitness when the harmful effects of the mutations are expressed. Inbreeding increases the realised load because more deleterious mutations become fully expressed as homozygotes. By minimising the realised 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 (Speak et al., 2024).

Lifting over the CADD scores from the closest model species would therefore allow for the quantification of the amount and severity of deleterious mutations within sequenced individuals, as well as enabling comparisons of the genetic load between individuals. Most genetic variation within the genome that codes for quantitative traits are not always deleterious, and many species or lineage-specific mutations have evolved since the divergence between the subject species and the model species. Therefore, to ensure that only unconditionally deleterious mutations are being assessed in the calculations of genetic load we will focus on the ultraconserved elements (UCEs) of the genome. UCEs (see section 1.23 of this thesis), are regions of the genome that show 80% or greater conservation across 100bp in distantly diverged taxa (Faircloth et al., 2012). It is therefore likely that mutations within UCEs are unconditionally deleterious (Silla et al., 2014). This means that they are ideal regions to use to assess with CADD scores lifted over from a model species as the mutation's effects are likely to be deleterious irrespective of the species.

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

2.3 Materials and Methods

2.4 Study species

The pink pigeon (*Nesoenas mayeri*) declined to around ten individuals in the late 1990s and it has since recovered to over 500 individuals. The species is listed as "vulnerable" by the

IUCN Red List (last assessed: 15 July 2020). During the conservation efforts, captive-bred populations were started both in Mauritius and in zoos globally. Six pink pigeon (*Nesoenas mayeri*) individuals from the captive-bred population of Jersey Zoo (n = 4) and Bristol Zoo (n = 2) were whole genome sequenced. Birds share common ancestry within the last 3-6 generations (Supplementary Figure 1) and produce offspring that are moderate to highly inbred (inbreeding coefficient, F=0.064 to 0.346) (Supplementary Table 1), which is typical of many zoo populations (Boakes et al., 2007). See Supplementary Information for further details.

2.5 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, mean depth of the 6 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 4). A Snakemake pipeline (Mölder et al., 2021) allowing for reproduction of this approach can be found on GitHub (https://github.com/saspeak/LoadLift).

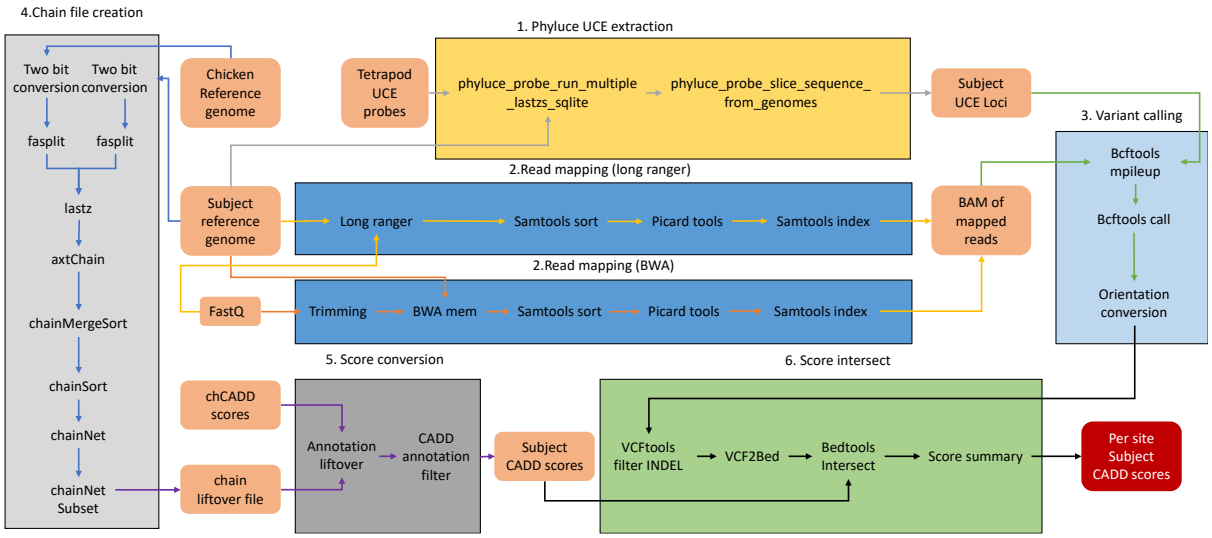


Figure 4 – The pipeline for the creation of per Single Nucleotide Polymorphism (SNP) pink pigeon Combined Annotation Dependent Depletion (ppCADD) scores from raw reads of individual pink pigeons. The Snakemake (Mölder et al., 2021) pipeline uses as input the sequencing reads of the subject individuals, 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 indicates 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).

2.6 Phyluce UCE Extraction

Previously published tetrapod UCE probes, based on the chicken reference genome (GalGal6) 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). Firstly, the genome was converted into a 2bit format using faToTwoBit to generate a tab file. After this, the probe set of 5K UCE loci (https://raw.githubusercontent.com/faircloth-lab/uce-probe-sets/master/uce-5k-probe-set/uce-5k-probes.fasta) of on average 120bp in length were downloaded and indexed using both SAMtools faidx v1.14 (Li et al., 2009) and BWA index v0.7.17 (Li and Durbin, 2009). Probes were then aligned to the pink pigeon genome using the *phyluce_probe_run_multiple_lastzs_sqlite* command. FASTA sequences matching the UCE loci were then extracted from the pink pigeon genome using the *phyluce_probe_slice_sequence_from_genomes* command, including a flanking region of 1000 bp surrounding the UCE loci. To enable SNP calling, each UCE's scaffold, start and end positions and orientation were extracted from the FASTA headers, and these data were separated into two files for UCEs in the forward and reverse orientation.

2.7 Chain file creation for annotation lift-over

To allow the accurate identification and scoring of single nucleotide variants, we decided to convert the chicken (i.e., chCADD) CADD scores to pink pigeon (i.e., ppCADD) CADD scores. This was achieved by lifting over the annotation of sites from the chicken GalGal6 genome to their positions in the pink pigeon reference genome. This was achieved using a chain crossover file. Firstly the chicken GalGal6 genome was downloaded from ftp://ftp.ensembl.org/pub/release-102/fasta/gallus_gallus/dna/Gallus_gallus.GRCg6a.dna_rm.toplevel.fa.gz (Warren et al., 2017). The reference pink pigeon genome was split into scaffolds and the GalGal6 genome was also split into 35 files (32 autosomes, MT, Z, W) using faSplit. Both the pink pigeon and GalGal6 genomes were converted to 2bit format using faToTwoBit and the chromosome size determined through twoBitInfo. We aligned all chromosomes to the pink pigeon genome using lastz v1.04.15 (Harris, 2007). Chain files were then created sorted and merged using axtChain, chainMergeSort, and chainSort respectively. Alignable regions were identified

from each chain file using chainNet and chainNetSubset. The output chain file allowed us to lift over the genomic coordinates of our set of UCEs from the chicken to the pink pigeon genome.

2.8 Converting chCADD scores to ppCADD

The chCADD scores of the chicken genome were downloaded from https://osf.io/c97ez/ (Groß, Bortoluzzi, et al., 2020). We then used the chain file to cross map the chicken CADD score to the reference pink pigeon genome using CrossMap.py (Zhao et al., 2014) . Only sites that were successfully mapped were retained, combined and sorted in BEDtools v2.30.0 following the pink pigeon scaffold annotation (Quinlan & Hall, 2010).

2.9 Read mapping

The pink pigeon reference genome was indexed using *long ranger make ref* (Marks et al., 2019). Raw read data for each of the six pink pigeon individuals was mapped to the pink pigeon reference genome using LongRanger align. Duplicates were removed from each of the mapped read BAM files using picard MarkDuplicates (Broad Institute, 2019).

2.10 SNP Calling

Variants within both sets of UCEs (forward and reverse orientated) were called using BCFtools mpileup v1.14 (Li, 2011) for all sites and all samples. Indels were removed using VCFtools v0.1.15 (Danecek et al., 2011). The SNP calls were then converted to BED format using vcf2bed and sorted using BEDtools sort (Quinlan & Hall, 2010). The reverse-orientated UCE BED file was then re-orientated to the forward orientation using a custom script.

2.11 Polymorphism scoring

Scores for forward and reverse UCEs were then extracted using BEDtools intersect (Quinlan & Hall, 2010) with the flags -wa -wb with the individuals UCE SNPs and ppCADD file, to output overlapping positions. Scores were then extracted using a custom script. This provided CADD scores for each SNP, identifying the chicken reference and alternative alleles for each position, the allele present in the pink pigeon UCE for each individual, and the corresponding ppCADD score for that polymorphism.

2.12 Filtering with R

Both files were then combined in R and filtered using the package *tidyverse* to remove any duplicated sites. The ppCADD scores were then quality checked using a custom R script including filtering of SNPs and recalling heterozygotes between an allele depth 17 - 83% of individual depth. Further filtering was used to remove fixed and non-scoring sites (a non-scoring site is a site that is homozygous for the chicken reference allele with a CADD score equal to zero).

2.13 Genetic load

Genetic load was calculated for individuals and their potential offspring (see below) using custom R scripts. Although we only sampled one female, calculations of the theoretical genetic load components were carried out for all potential crosses. These hypothetical crosses included selfing individuals (i.e., crossing an individual with itself), and these crosses were made irrespective of the sex of the individuals. We included these crosses for illustrative purposes to show the application of the method.

The genetic load, the realised load and the masked load of each individual were calculated using the following formulas, as described in (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. Note that part of the realised load comprises heterozygous mutations that are assumed to be partially dominant.

2.14 Pedigree of captive bred population

A pedigree of the captive bred pink pigeon population was created using the R package FamAgg (Rainer et al., 2016) containing all 849 living and deceased individuals. The kinship coefficient values for potential offspring from the crosses were calculated using the kinship function of FamAgg. The kinship coefficient is a measure of genetic relatedness. It is defined as the probability that a pair of randomly sampled alleles at a one locus are identical by descent. Therefore, it is the probability that an allele selected randomly from individual i, and an allele selected from another individual j are identical (autozygous) and from the same ancestor (Rainer et al., 2016).

2.15 Homozygosity

The inbreeding coefficient (F_{RoH}) of the six pink pigeon individuals was also calculated using runs of homozygosity (RoH) with bcftools roh (Narasimhan et al., 2016) which implements a hidden Markov model approach to detect tracks of continuous homozygote states. Once RoH had been identified the F_{RoH} was calculated for each individual using:

$$\mathrm{F_{RoH}} = \frac{\sum \mathrm{RoH} > 100\mathrm{KB}}{\sum \mathrm{BP}}$$

[4]

Here F_{RoH} is equivalent to the proportion of the genome that is identical by descent between both parents, and it is proxy of an individual's inbreeding coefficient. Short RoH under 100Kb were ignored and the sum length of all of the scaffolds in the pink pigeon genome in BP was used as the genome length (*n* = 1,294,301,718 bps).

2.16 Computer simulations of breeding regimes

This part of the analysis was conducted in collaboration with Thomas Birley and Hernán E. Morales. 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 minimise inbreeding depression. However, purifying selection against the genetic load can reduce genetic diversity (Bertorelle et al., 2022). The values presented in the figure represent the mean results obtained from 100 replicates.

2.17 Results

2.18 Distribution of UCEs and CADD scores

The 4976 UCEs along the 34 chromosomes of the chicken reference genome are not evenly distributed (Figure 5A), 15 chromosomes were significantly depleted for UCEs, whilst 9 chromosomes were significantly enriched for UCEs (Supplementary Table 2). Figure 5B 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 5B). 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).

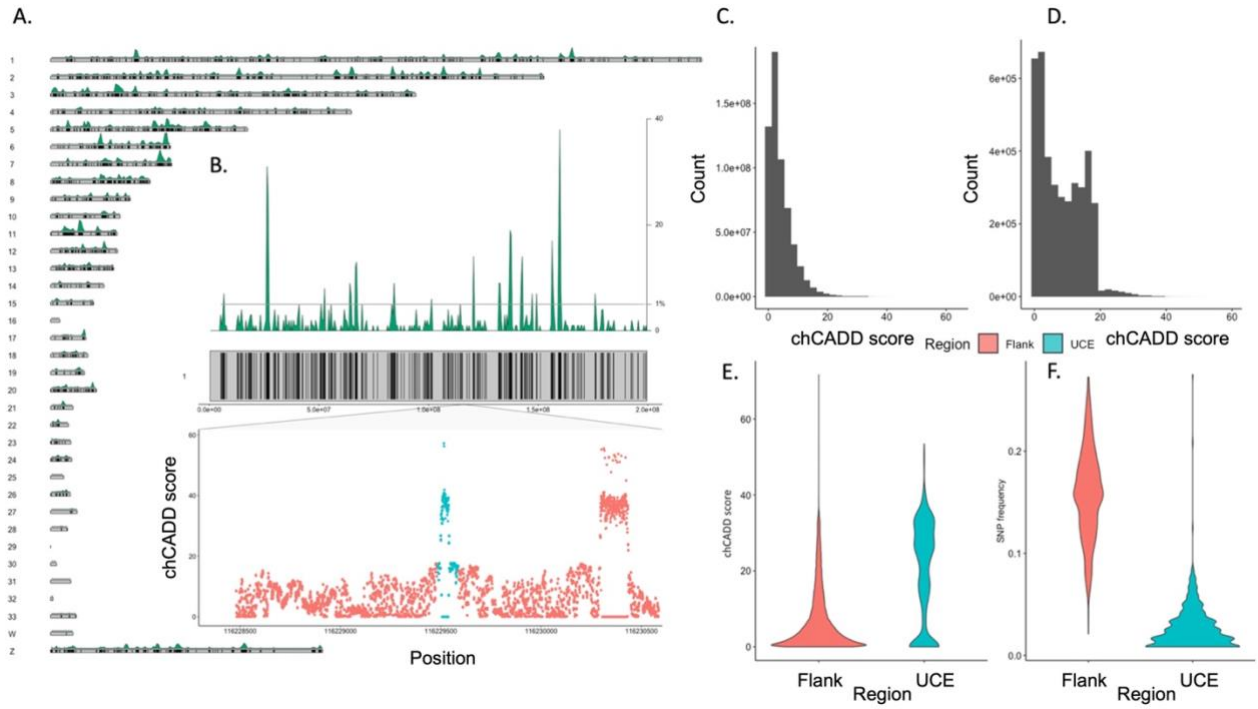


Figure 5 – Distribution of ultraconserved elements (UCEs) and their mutation impact scores (CADD scores). **(A)** Karyotype plot of the chicken genome with the distribution of UCEs (black bars) and density of UCEs (green peaks). (**B)** Karyotype plot of chicken Chromosome 1 showing the distribution of UCE-dense regions. Green peaks above the 1% horizontal line are significantly enriched for UCEs ($p<0.01$). 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).

Figure 5C 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 5D). 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). 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), which is consistent with the effect of purifying selection.

2.19 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 6), which elevates the offspring's realised load and reduces the masked load (Figure 6). Optimal mate pairing can significantly reduce the realised load of the offspring compared to random mating (R^2=0.258, $F_{1,13}$ = 8.32, p = 0.00918).

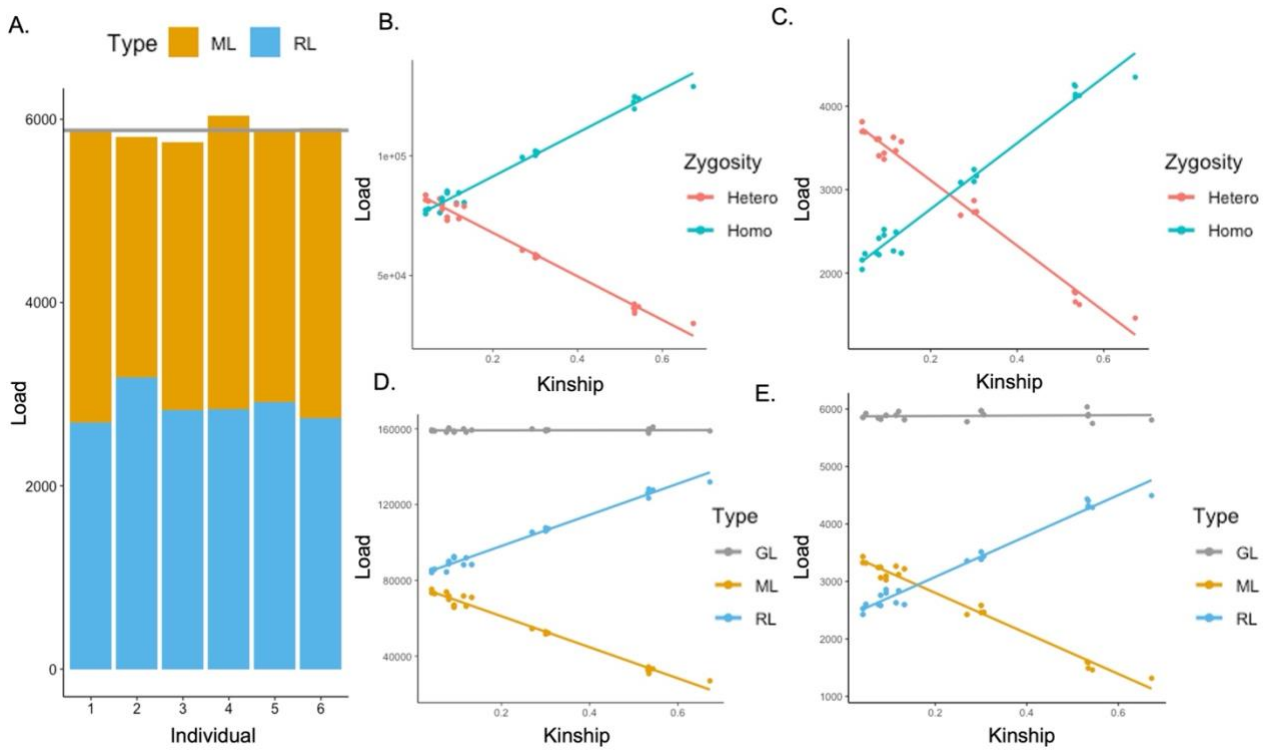


Figure 6 – The composition of the genetic load in six pink pigeon individuals and their hypothetical offspring. A) The total realised load (Blue) and masked load (Orange) in each of the six pink pigeon individuals within their UCEs. **(B and C)** The realised load at heterozygous loci (Red) and homozygous loci (Teal) of the offspring is shown for **B)** the total region and **C)** UCEs only. **(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 **D**) total region and the UCE only **E)**.

Next, we performed an analysis to identify optimal crosses to minimise genetic load (Figure 7). CADD scores for the potential offspring of each cross, including self-mating, are provided for the genetic load (Supplementary Table 3), realised load (Supplementary Table 4) and masked load (Supplementary Table 5). Figure 7A shows 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.

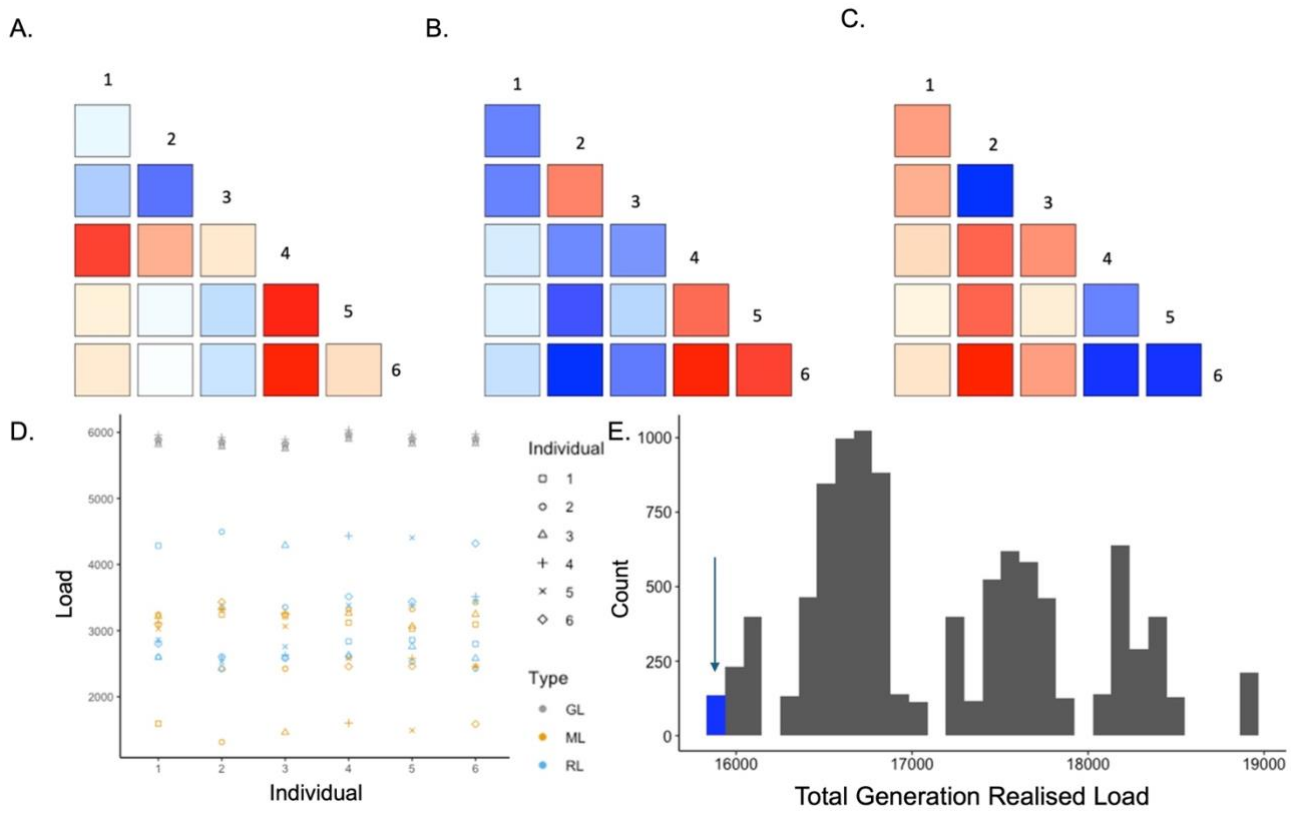


Figure 7 – The genetic load at UCEs of six pink pigeons calculated using cross-mapped chCADD scores. Correlogram showing the total load of potential offspring between all combinations of six individuals of the captive pink pigeon population. The colour of the tile is relative to the load of the offspring 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)** 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 the summed realised load in all offspring 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.

To predict the degree of inbreeding depression, the realised load of the offspring of different crosses was calculated. Blue tiles in the correlogram in Figure 7B 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 7E), 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 (uncle and nephew), but they each possess a low genetic load. However, because they are related, their offspring express a high realised load, even though their genetic load is low.

2.20 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 (Figure 8A) and realised (Figure 8B) load over generations. Both regimes also suffered from a large decline in fitness due to a mutation meltdown (Figure 8C). In contrast, both the genetic load and realised load were reduced in “Minimise load” and “Minimise load and relatedness” regimes (Figure 8A & B). Therefore, genomics-informed captive breeding can successfully reduce the realised load and homozygosity of deleterious mutations, independently of consideration of relatedness. Consequently, mean fitness remained high in these regimes, increasing during the first ten generations (Figure 8C). However, populations lost neutral genetic diversity at a relatively fast rate in the “Minimise load” regime (Figure 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 (Figure 8D).

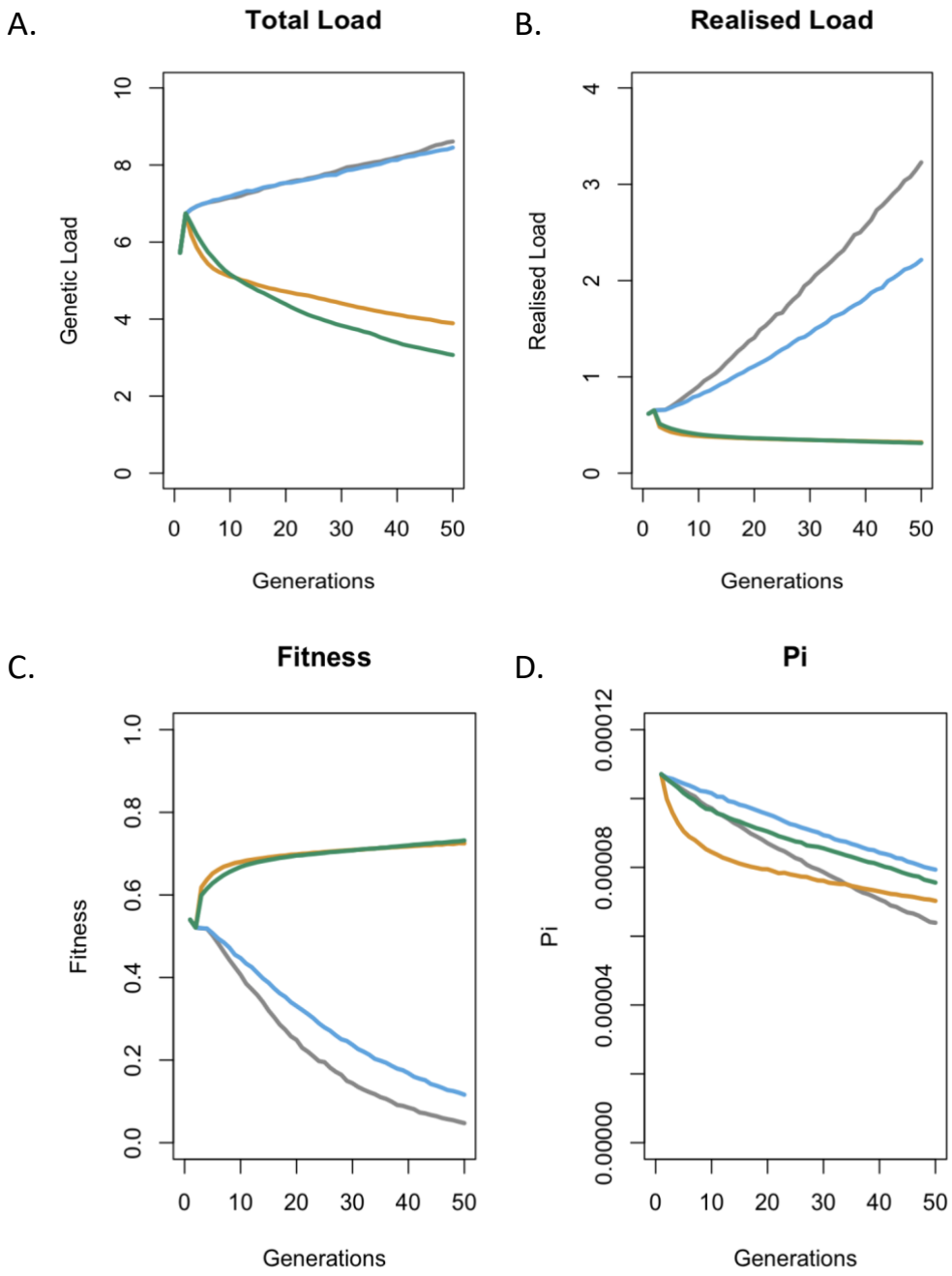


Figure 8 – Impact of the four breeding regimes, simulated over 50 generations. Showing the impact on **A)** the genetic load, **B)** the realised load of offspring, **C)** the mean fitness of adults, and **D)** neutral nucleotide diversity (π). Each coloured line corresponds to a specific mating regime: "Random mating" (grey), "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 replicates. Figure produced by Thomas Birley, UEA.

2.21 Discussion

We conducted a proof-of-concept study to evaluate the utility of genomics-informed conservation for the management of captive populations in zoos. The 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 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 the 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 ten 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 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 a CADD score of 20. Theoretically, if a population carries a realised load of one lethal equivalent (LE), the probability of an individual surviving equals: $Fitness = e^{-1} = 0.368$. Hence, a reduction of 7.4% in realised load results in an increase of survival rate from 36.8% to 39.6% ($Fitness = e^{-0.926} = 0.396$). This is a 7.7% relative increase in survival probability. With a higher realised load of 2 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 realised load is likely to reduce inbreeding depression and increase fitness (Bertorelle et al., 2022). Furthermore, the higher the realised 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 realised load and individual fitness. Investigation into the relationship between CADD scores and fitness impacts could be carried out by comparing the presence and absence of high CADD scores between viable and non-viable embryos. This would help to determine if possessing a high CADD score translates to an observable decrease in fitness. Furthermore, large scale investigations could also be carried out to compare the breeding success & offspring fledging of individuals to the calculated realised load of individuals and predicted offspring. 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, 2024). 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 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 maintaining adaptive potential.

2.22 Conclusion

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 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 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 programmes.

3. Investigating the similarity of CADD scores in three model species

3.1 Abstract

In Chapter 2, I showed how CADD scores can be lifted over from model species and used to assess the genetic load of threatened species. This enables conservation managers to assess the genetic health and determine optimal mate pairings within their populations. To illustrate that CADD scores calculated in a model species can be lifted over to a distantly related species, in this chapter I have analysed how CADD scores compare between three model species. To do this, I extracted the UCE and flanking regions from the reference genomes of three model species for which CADD scores have been calculated human (hCADD), pig (pCADD) and chicken (chCADD). I compared the CADD scores within the shared UCEs of all three species and showed that the CADD scores of UCEs are, on average, higher than those of the flanking regions. Moreover, I show that within individual UCEs the CADD scores are comparable between all three model species and that across all 2537 shared UCEs, the average CADD scores are significantly correlated (pigs and humans: $R^2 = 0.3273$, pigs and chickens: $R^2 = 0.1803$ and humans and chicken: $R^2 = 0.4089$). This suggests that it is appropriate to use CADD scores from model species to estimate the genetic load of non-model species within UCEs. Due to the high level of genetic conservation, I demonstrated that CADD lift-overs for UCEs can be performed, despite considerable phylogenetic distance between species.

3.2 Introduction

Like many other technological advances, the progress in genomic sequencing techniques and bioinformatics tends to first be developed for medical applications in humans and are then subsequently applied to model species and/or domesticated species (Arya et al., 2024). Combined Annotation Dependent Depletion (CADD) scores are a prime example of this, being first developed for use on the human genome (Kircher et al., 2014), before the techniques were applied to the chicken (chCADD)(Groß, Bortoluzzi, et al., 2020), pig

(pCADD)(Groß, Derks, et al., 2020) and mouse (mCADD)(Groß et al., 2018). The key driving factors behind developing CADD scores on these model species are firstly the benefits to human health through the advancement of medical technology (Kircher et al., 2014) and secondly the high economic value of identifying mutations with a negative impact on animal breeding (Groß, Bortoluzzi, et al., 2020). The vast quantity of projects and work on model species has allowed for a repository of data on the impacts of Single Nucleotide Polymorphisms (SNPs) on these model organisms. For instance, more than 20 years ago the human genome's initial release covered 94% of the human genome with 30,000–40,000 protein-coding genes and more than 1.4 million SNPs. From this developed the ENCODE (ENCODE, 2012) project that aimed to understand the functional coding regions of the genome, as well as the ClinVar (Landrum et al., 2014; Landrum et al., 2020) dataset of 2.3 million pathogenic variants within the human genome. This wealth of clinical data has been utilised to produce CADD scores for the human genome (Kircher et al., 2014). However, this vast quantity of genomic information is not a resource that is available for non-model species which has only a handful of sequenced individuals.

To overcome these issues, in Chapter 2 we used chCADD scores (Groß, Bortoluzzi, et al., 2020) designed for the chicken genome (Galgal6) lifted over to the non-model species in order to calculate the genetic load of individual pink pigeons (Speak et al., 2024). Currently, CADD scores exist for multiple model species and have great potential to be applied to a wide range of threatened species across birds and mammals. It has been shown that annotation lift over tools are 99.21% accurate when converting between human reference genomes hg19 and hg38 (Luu et al., 2020). This provides confidence that CADD scores are applicable between species when they are combined with the annotation liftover on ultraconserved elements (UCEs) as these are, by definition highly similar between species (Bejerano et al., 2004; Faircloth et al., 2012). However, it would be of interest to assess the similarity of the CADD scores within the UCEs of three distantly diverged model species. Only if there is a significantly high correlation, can we conclude that CADD scores of model species are applicable to be used in threatened species.

There is great potential for CADD scores to be used to quantify the genetic load within threatened species (van Oosterhout, Speak, et al., 2022; Speak et al., 2024). Applying these

approaches to UCEs allows for the detection of mutations that are likely deleterious to the subject species (Silla et al., 2014). Within the chicken genome, the highest chCADD scores were those which were the most potentially deleterious, namely stop-gained mutations and splice-site altering mutations (Groß, Bortoluzzi, et al., 2020). Similarly, when investigating the pCADD between coding positions of genes within the pig genome (Groß, Derks, et al., 2020), comparing the pCADD score of the third codon position against the second codon position and the first codon position, scored higher for 81.3% and 80.7% of genes respectively. This indicated that higher pCADD scores do relate to the more deleterious mutation. Therefore, a multi-species-wide comparison of the CADD score distribution within exons and introns can show how comparable high CADD scores are within species.

CADD scores assess the deleteriousness of mutations relative to all other potential variants within the genome of the focal species. But how do CADD scores of different model species compare and correlate? In this chapter, I will assess the correlation of CADD scores between three model species: human, pig and chicken. The three model species will be compared to examine the correlation between CADD scores and to determine whether CADD scores are higher within the UCEs than in other regions of the genome. I therefore use the existing published CADD score data available for each species, as well as the reference genomes, which allows me to compare the CADD scores of corresponding substitutions within the genomes of different model species.

3.3 Materials and methods

3.4 UCE extraction

The ultraconserved elements (UCEs) and flanking regions of 1000bp up and downstream form the UCE loci were extracted using the PHYLUCE pipeline (Faircloth, 2016) from the reference genomes of three model species with CADD scores available:

human hg38 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.40/),

pig Sscrofa11.1 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000003025.6/) (Warr et al., 2020) and chicken galGal6

(https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000002315.6/) (Warren et al., 2017). The extracted UCE sequences were aligned to the UCE probe set using BLAST, and filtered to remove sequences that aligned at less than 95% alignment for all species and where the UCE sequence that matched was under 90 bp in length. After filtering there were 2537 UCE loci, found in all three species.

3.5 CADD score extraction

Once the shared UCE loci had been identified, the CADD scores for the three model species: chicken (chCADD) (Groß, Bortoluzzi, et al., 2020) (available at: https://osf.io/c97ez/); pig (pCADD) (Groß, Derks, et al., 2020) (available at: https://www.bioinformatics.nl/pCADD/indexed_pPHRED-scores/) and humans (hCADD) (CADDv1.6) (Rentzsch et al., 2021) (available at: https://cadd.gs.washington.edu) were downloaded. The CADD scores for the shared UCEs and their 2000 bp flanking regions were extracted using Bedtools Intersect (Quinlan & Hall, 2010). The CADD scores were then orientated to that of the UCE probes.

3.6 Obtaining UCE annotation

The annotation files for the reference genomes of the three model species that the CADD scores were calculated on were downloaded from the NCBI database. These were imported to R and each species was converted to genomic ranges using the package GenomicRanges (Lawrence et al., 2013). The UCE and flanking regions were then converted to genomic ranges and intersected using the overlapRegions function. The resulting files were filtered to include only annotation data within the UCEs and flanking regions and joined to the CADD score data for each species.

3.7 Range comparison

The number of CADD scores for each species within the flanking regions and UCE were calculated (Supplementary Table 6), along with the number of CADD scores within the exons and intergenic regions for the UCEs and flanking regions for all three species (Supplementary Table 7). The CADD scores were binned into ranges between 0 and 60 with a bin width of five and the number of scores for each range for each species was counted. The percentage difference between exons and intergenic regions were calculated using the following formula:

$$|x| = 100\frac{|e - i|}{e + i} \quad [5]$$

Where *e* is the count of the number of CADD scores in the range for a species annotated within the exons and *i* is the count of the number of CADD scores in the range for a species annotated within the intergenic regions.

3.8 Results

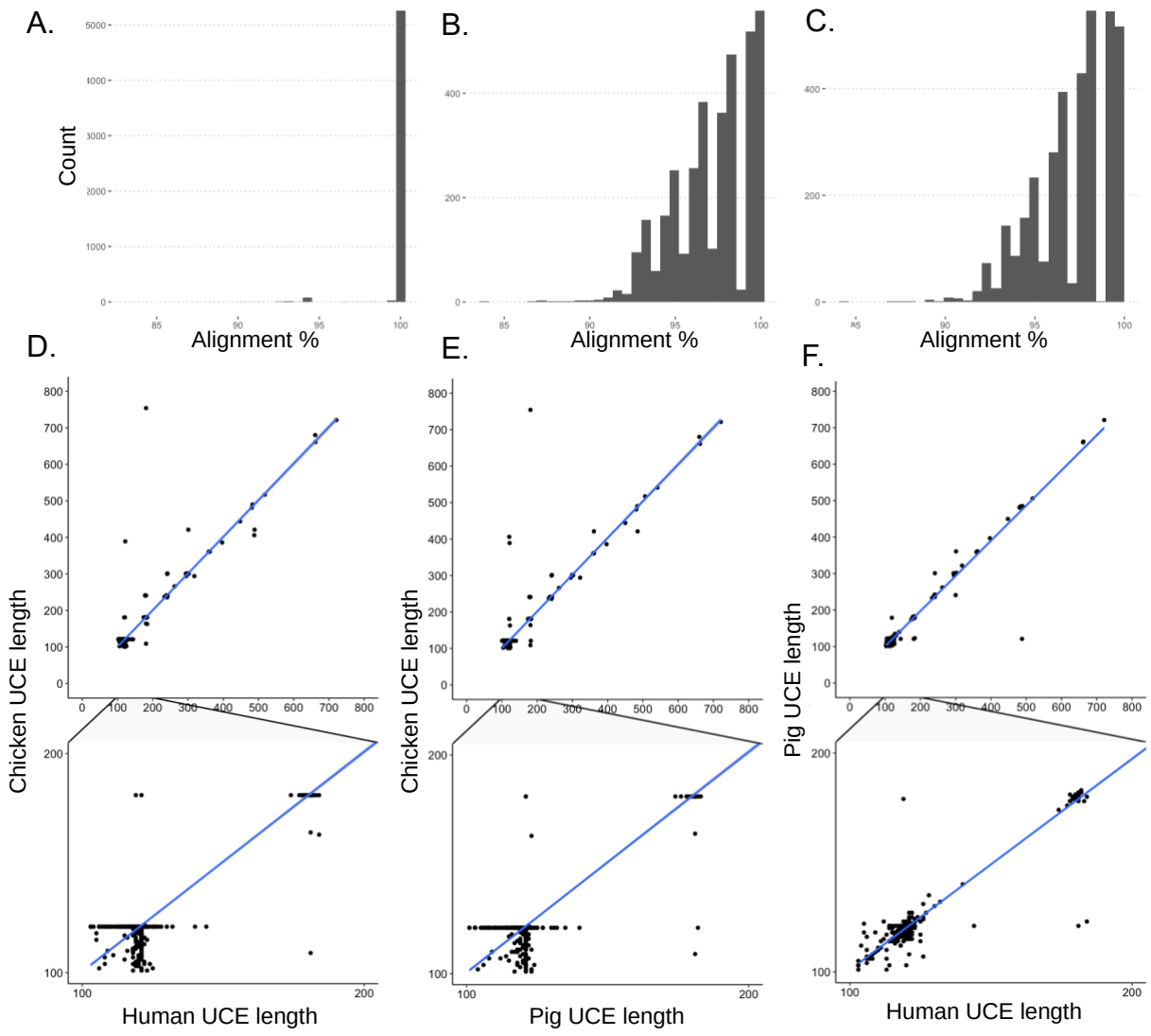


Figure 9 – A comparison of the alignment and length of UCEs to the genomes of the three species. Histograms showing the BLAST alignment of the UCE probes to the reference genomes of: **A)** chicken **B)** pig and **C)** human, **D)** correlations comparing chicken UCE length to human UCE length, **E)** chicken UCE length to pig UCE length and **F)** pig UCE length to human UCE length. For each comparison, a cut out showing the comparisons between 100 and 200 bp are provided.

Of the original 5472 sequences in the Faircloth UCE probe set (Faircloth et al., 2012), 5301 UCEs were aligned to the chicken reference genome (Figure 9A), 3510 to the pig reference genome (Figure 9B) and 3536 to the human reference genome (Figure 9C). After filtering to

remove UCE sequences below 90bp and 95% alignment to the probe sequence, 2537 UCE loci were found in all 3 species. The total length of filtered sequences for only the UCE regions was compared in a pairwise manner across the three species, showing that there was a significant correlation between the human and chicken (Figure 9D) (Regression analysis: $R^2 = 0.8896$, $p < 2.2e\text{-}16$) and pig and chicken (Figure 9E) (Regression analysis: $R^2 = 0.8798$, $p < 2.2e\text{-}16$). The strongest correlation in length was between the pig and human UCEs (Figure 9F) (Regression analysis: $R^2 = 0.9581$, $p < 2.2e\text{-}16$).

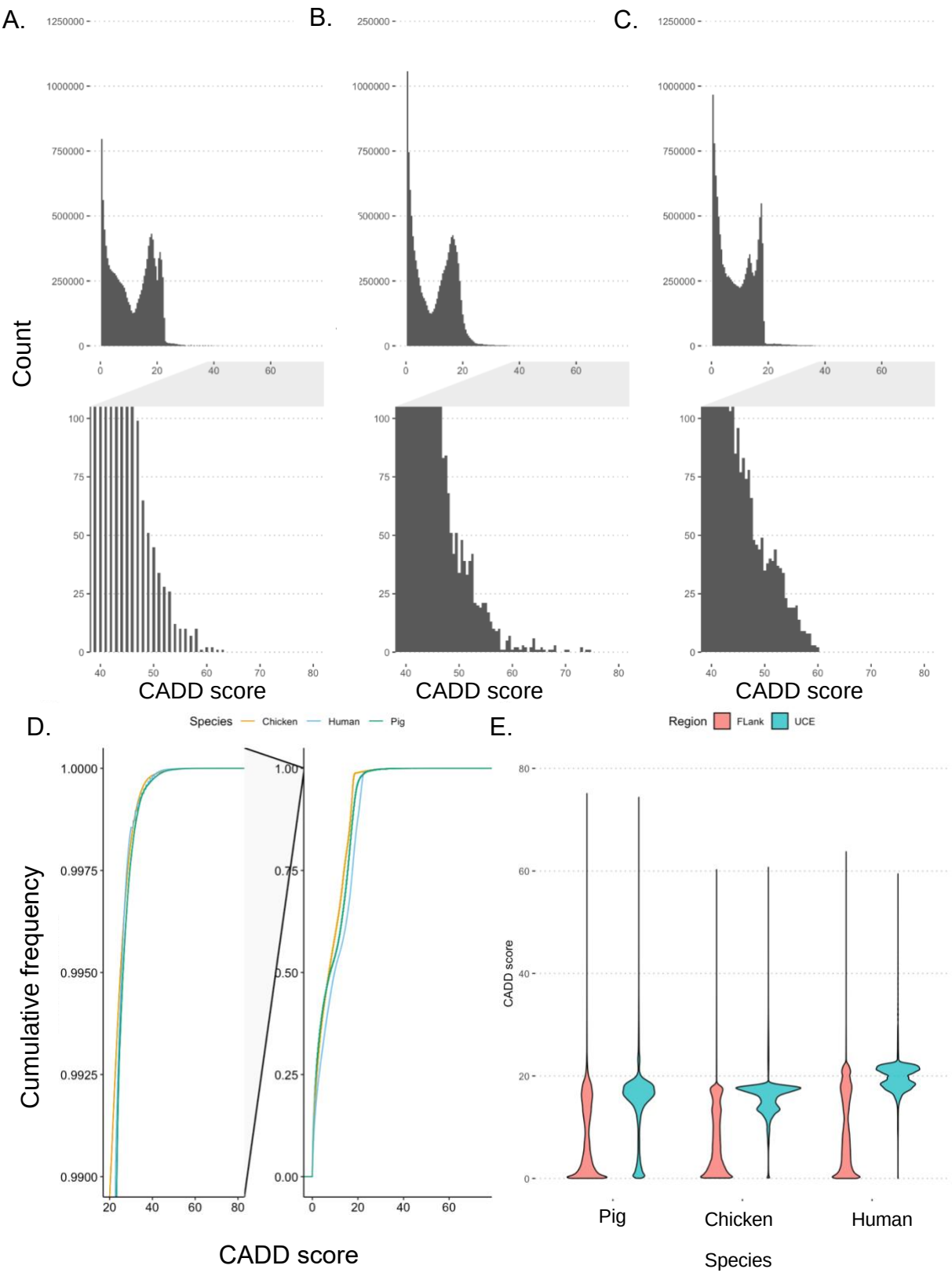


Figure 10 – Distribution of all CADD scores within the UCEs and their 1000bp up and downstream flanking region for each of the three model species. A) human (hCADD), **B)** pig (pCADD) and **C)** chicken (chCADD), with an insert for each species for CADD scores

of over 40. All three species have a similar distribution of CADD scores. Maximum CADD scores for human, pig and chicken were at 63, 74 and 60 respectively. **D)** The cumulative frequency plots for the CADD scores for the chicken (Orange), human (Green) and pig (Blue). **E)** The distribution of CADD scores within the UCEs (Blue) and the 100bp up and downstream flanking regions (Red) across the three species, pig, chicken and humans.

The shape of the distributions of the three CADD scores for the UCEs of the three species (humans, pigs and chickens) follow a similar shape (Figure 10A), with approximately 25% of all scores between 0-5, and 50% of all scores between 0-10 (Figure 10B). The maximum CADD scores within the UCEs were comparable: 63 in humans; 60 in chickens and 74 in pigs. The median CADD scores were significantly different between all three species: 9.74 in humans; 7.29 in chickens and 7.83 in pigs (Kruskal-Wallis test, H = 746039, df = 2, p-value < 2.2e-16). The shape of CADD score distributions are broadly similar, with the majority of sites between 0-10 and a second peak between 15-20 for each species. However, this suggests that the medians of the CADD distributions are distinct. The distribution of CADD scores within the UCEs and flanking regions of the three model species follows the same trend as that seen in Chapter 2 (Figure 5), with UCEs having higher on average CADD scores than sites within the flanking regions. The UCE regions for all three species mostly contain CADD scores between 15 and 25, which was higher than those in the flanking regions. The pCADD scores showed the highest scores, with CADD scores > 70 for sites in both the UCE and their flanking regions. The hCADD scores possessed very few CADD scores of between 5 to 10 (Figure 10C) with 99.75337 % of these scores being in the flanking regions (Figure 10E).

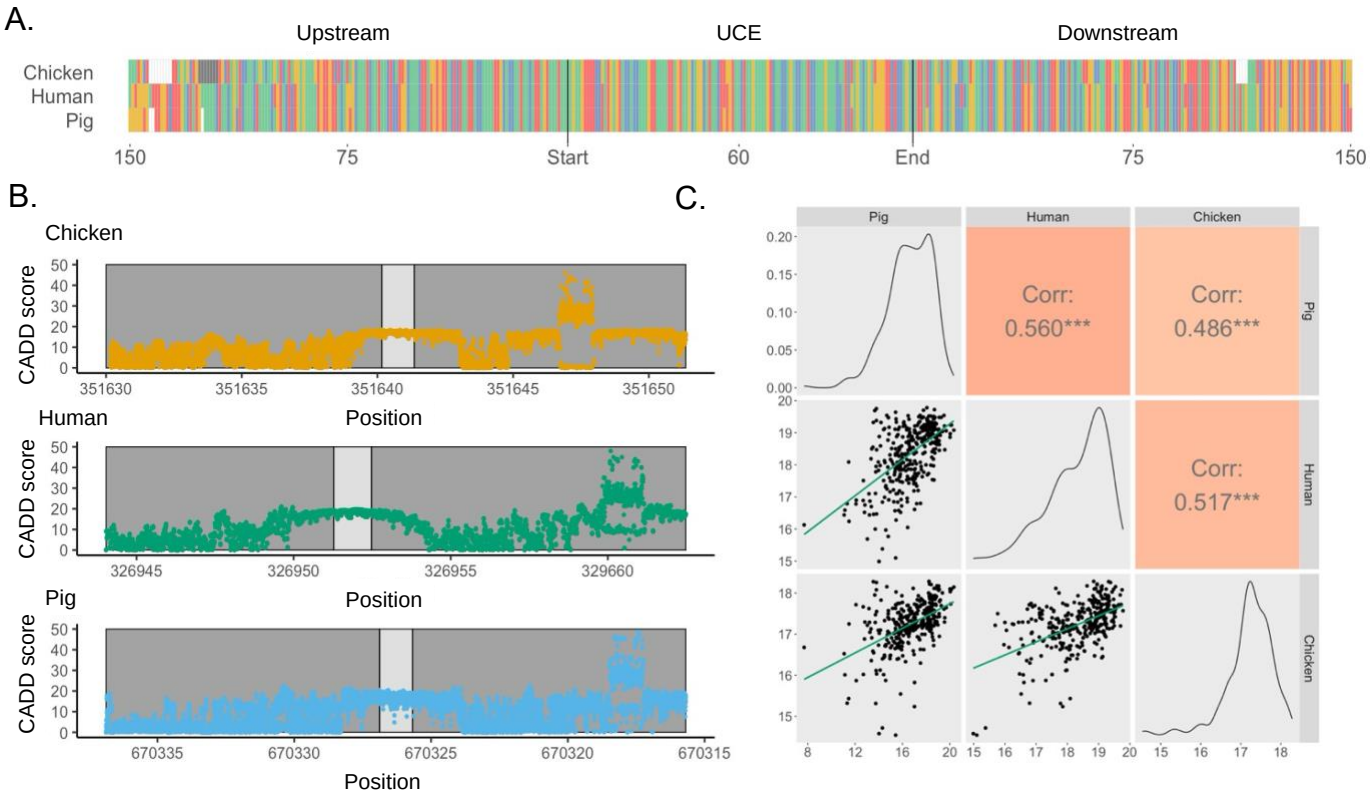


Figure 11 - An example of the alignment of all three species' CADD scores across one single UCE locus and flanking region. A) A multiple sequence alignment for a single UCE, UCE 1004 showing the 120 bp of the UCE in the centre and 150 bp of up and downstream flanking region. **B)** Per nucleotide CADD score for all bases within UCE 1004 of the three model species pig (Blue), chicken (Orange) and human (Green). All x-axis values are the nucleotide position divided by 100. **C)** Correlation plot for UCE 1004 showing the correlation of the CADD score per site between the species in the upper half of the plot and the distribution of the per-site CADD scores along the diagonal and an X-Y comparison of the per-site CADD score for all sites in UCE 1004 showing the pairwise comparison between the species in the lower half.

All three species show comparable major peaks of CADD scores at a CADD score of circa 20, in the centre of the UCE which would represent the region of the UCE itself and a small amount of the flanking regions (Figure 11A). All three model species show a similar second-high peak in CADD score towards the end of the flanking region. This could potentially indicate the presence of an intron or in the human genome (Green) the presence of Kinase anchoring protein 6. The chicken and human sequences are in the forward orientation whilst

the pig is in the reverse orientation and therefore the peak is still in the same region across all three species.

There is high continuity between the sequences of all three species at the UCE elements compared to within the flanking regions up and downstream of the UCE loci (Figure 11A). When comparing the CADD scores across a single UCE and its flanking regions, there were low CADD scores within the flanking regions (Figure 11B). There was an increase in CADD scores to around 20 at the UCE loci (light grey). Increased CADD scores then show a suspected gene with high CADD scores and low potential scores, indicating that these relate to the third codon position. There is a strong pairwise correlation between the CADD scores at every site within UCE 1004 (Figure 11C) showing that the CADD scores within the UCEs are comparable between the three species.

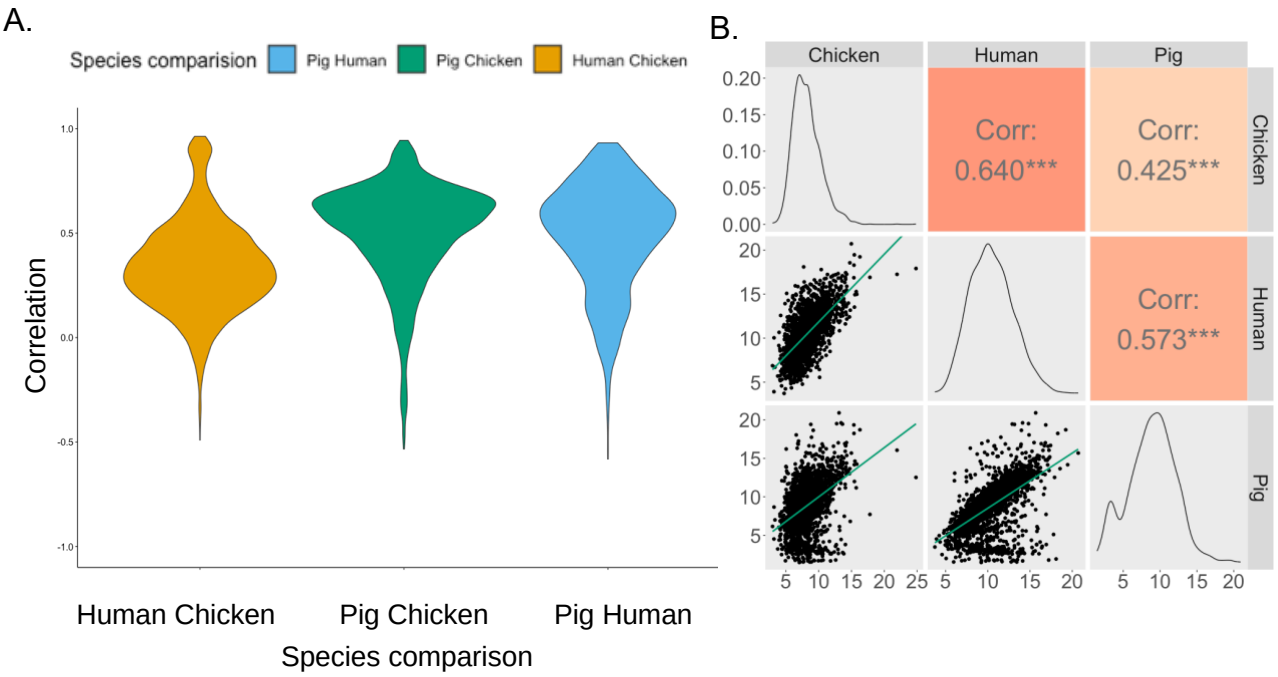


Figure 12 – A comparison of all UCEs across three model species. A) The distribution of correlations of each of the 2537 UCE loci, comparing human and chicken CADD scores (Orange), chicken and pig CADD scores (Green) and human and pig CADD scores (Blue). **B)** Correlation plot showing the correlation of the average score per UCE between the species in the upper half of the plot and the distribution of the UCE averages along the diagonal and an X-Y comparison of the average score of each UCE showing the comparison between the two species in the lower half.

The pairwise comparisons of CADD scores at every site within the UCEs show that there are a large range of correlations, with some UCEs showing no correlation, but the majority for all three species comparisons (pig to human, human to chicken and chicken to pig) all showing high average correlation values (Figure 12A). All three species correlated significantly across the average CADD score across all the UCEs (Figure 12B), with the strongest correlation being between the human and chicken, followed by those between human and pig. This suggests that although there is a large evolutionary time between species, the CADD scores at the UCEs are similar and are suitable for use on more closely related species. However, it is prudent to use a model species that is the most closely related phylogenetically, as variation between scores were observed, which is likely to increase with phylogenetic distance.

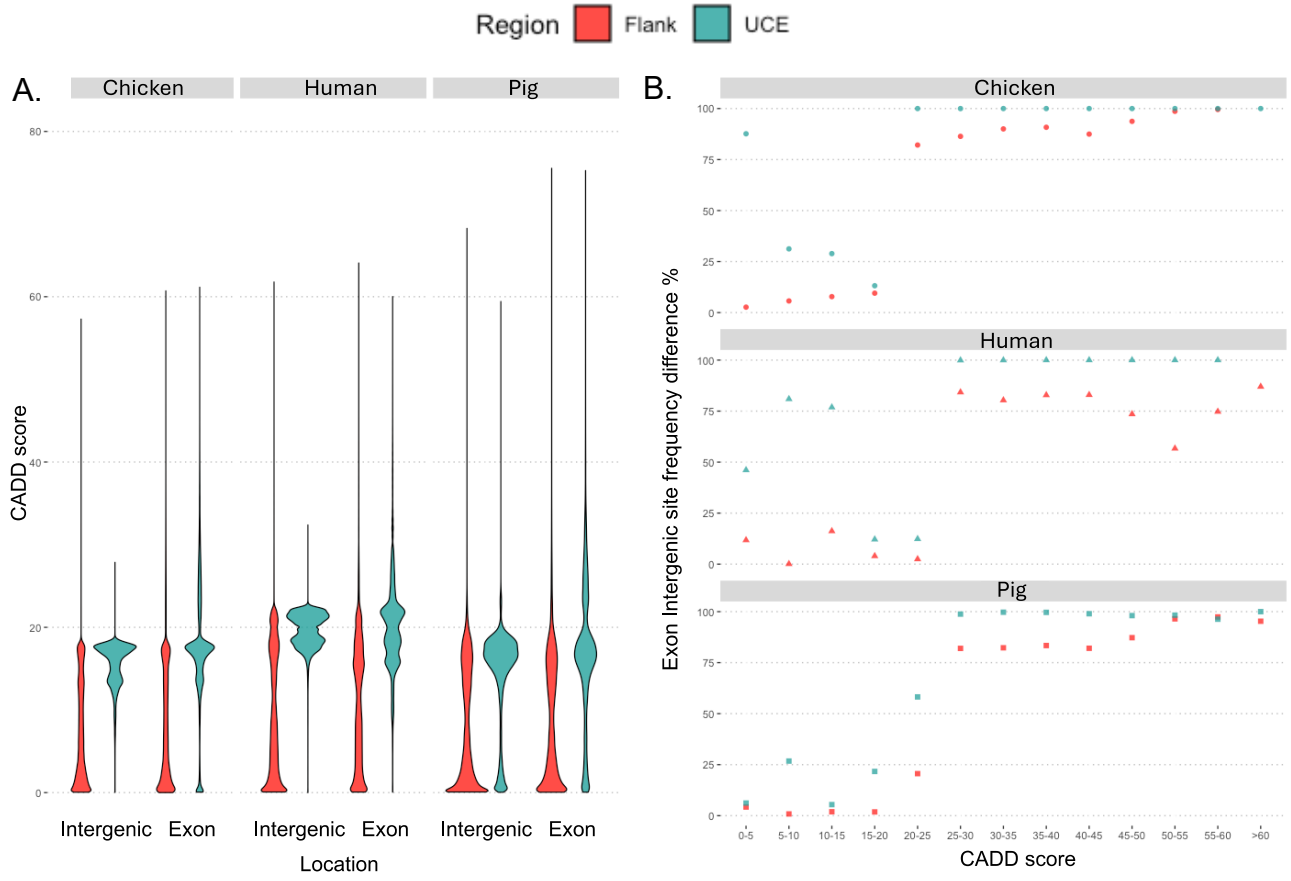


Figure 13 – A comparison of the CADD scores across three model species, humans (hCADD), chicken (chCADD) and pig (pCADD). **A**) Comparing the CADD scores within the exons and intergenic regions within both the UCE (Teal) and the up and downstream flanking regions (Red) for all three model species. **B)** Comparing the difference in the frequency of sites within CADD score ranges between the exon and intergenic regions of the UCEs (Teal) and the up and downstream flanking regions (Red). These are shown for the chicken (Circles), human (Triangles) and pig (Squares). The difference in frequency was calculated using Equation 5, as the difference in count between the exons and introns within each CADD range. Raw counts within each region for each species are available in Supplementary Tables 6 & 7.

Due to the high number of sites within all three species (chCADD n = 13,975,044, hCADD n = 13,709,307 and pCADD n = 13,885,554) (Supplementary Table 6) all the pairwise comparisons between introns and exons were statistically highly significant (ANOVA: $F_{2,41569893} = 3840.0$, $p < 2.2e\text{-}16$). It is therefore more informative to calculate and report the medians, and their relative differences.

The median CADD scores of the three model species within the UCEs (median: hCADD = 19.8, pCADD = 16.3, chCADD = 16.6) and the exons (median: hCADD = 19.8, pCADD = 16.8, chCADD = 17.3) were similar to the intergenic UCE sites (median: hCADD = 19.8, pCADD = 16.3, chCADD = 16.5) (Figure 13). This reinforces that within the UCE, the median CADD scores across all species were similar in both exon and intergenic regions. There was a visual increase in CADD scores within exons compared to introns for all species (Figure 13A), but this was also paired with a decrease in CADD scores due to the third codon position (Figure 11A), resulting in small changes in median CADD score. The exons possess significantly more UCE sites than intergenic regions for all three species.

There was also a strong shared similarity in that higher CADD scores were found in the exons (Figure 13), with 100% of hCADD and chCADD scores and 98% of pCADD scores over 25 being found in exon regions (Figure 13B). In comparison, there was less difference between the CADD scores in the flanking regions between the exons and intergenic regions (Figure 13B). This suggests that for all species, CADD scores within the exon and UCE regions are more likely to be deleterious than in other regions of the genome.

Of the 2201 UCEs with annotation and CADD scores, 17.25% were located within exons for humans, 13.31% for chicken and 13.35% for pig. When accounting for the relative size of both regions (chicken genome: exons = 25%, pig genome: exons = 13%, human genome: exons = 23%), the exons were therefore not significantly enriched for UCEs (Supplementary Table 7). The highest CADD scores in the intergenic and UCE regions are lower in human and chicken than in intergenic flanking regions (Figure 13) suggesting that, although CADD scores are still on average higher than in the flanking regions, the most deleterious sites with extremely high CADD scores are located within exons in UCE regions.

3.9 Discussion

3.10 Multi-species comparison of CADD scores

CADD scores can be lifted over from model species to the genome of threatened species, thereby providing an insightful assessment of the genetic load (Speak et al., 2024) (Chapter 2). The genetic load can thus be compared between individuals, and optimal mate pairings can be determined for captive breeding programmes (Speak et al., 2024). Here in Chapter 3, I have shown that CADD scores in UCEs and flanking regions correlate between three model species (human, pig and chicken), demonstrating that the lift-over of CADD scores between model and non-model species is feasible and appropriate. There is a high similarity in the length of UCEs across all three pairwise comparisons between chickens, pigs and humans (Figure 9). This confirms that the extraction of UCEs ensures that we are comparing similar regions across all three species. Exact CADD score values for specific mutations differ between the three model species, but this is likely due to the CADD scores being a rank relative to the size of the species genome. As the three genomes differ in size (human (hg38) genome length = 3.1 Gb, pig (susScr11) genome length = 2.5Gb and chicken (galGal6) genome length = 1.1 Gb), the exact CADD score values for specific mutations will not be identical. However, as we have shown the trend in CADD score values are comparable, with high CADD scores correlating to high CADD scores across these species (Figures 11 & 12). I have furthermore demonstrated that the CADD scores in UCEs are higher than those in their flanking regions (Figure 10C), and that most UCE CADD scores fall between 15 and 25. This is equivalent to mutations that are between the 3.16% and 0.316% most deleterious mutations within the entire genomes. This supports the hypothesis that despite the relatively small number of sites that comprise the UCEs, mutations within these regions comprise some of the most deleterious variants in the genome. This is shared across all three model species, despite their high divergence.

3.11 Contextualising this work in the wider literature

CADD scores across all three model species show strong significant correlations when comparing the CADD score per site for individual UCEs, such as UCE-1004 (Figure 11). All three species had low CADD scores in the flanking region upstream, rising to CADD scores

of 20 for potential mutations within the UCE. All three species also showed a similar distribution of CADD scores in the downstream flaking region, suggesting the presence of a gene, presenting both high CADD scores of greater than 20 and low CADD scores implying that these are scores for the third codon positions. In the pCADD scores, third codon positions were found to be lower than first codon positions in 8,830 genes, and second codon scores in 8,901 genes of 10,942 genes assessed (Groß, Derks, et al., 2020). This strong shared pattern across all three species suggests that the value of CADD scores does indicate deleteriousness.

Gene annotations of the three species show that UCE sites in the exome possess higher CADD scores than intergenic UCE sites (Figure 13B). The initial publications of the human, chicken and pig CADD scores did not focus on the difference in scores in coding and non-coding regions. However, studies into genetic diseases in humans have utilised CADD scores in their detection of potentially deleterious variants (Gelinas et al., 2020; Leung et al., 2024). Studies investigating the genetic basis of childhood-onset Pulmonary Arterial Hypertension identified variants in the gene platelet-derived growth factor D which had a high CADD score of 28.2 (Gelinas et al., 2020). Similarly, an investigation into whole-exome sequencing for Alzheimer's disease reported that 15.5% of their variants had CADD scores over 20, and that 1.8% of their exome variants had CADD scores over 30 (Leung et al., 2024). Taken together, these results suggest that deleterious coding variants possess high CADD scores. Hence, substitutions in exome UCEs are amongst the most deleterious in the genome. Identifying individuals with equivalent mutations through conservation-genomic techniques has the potential to limit their deleterious impacts, thereby improving fitness and population viability in captive breeding and reintroduction programmes.

3.12 Future investigations into CADD scores

Although we can confidently transfer CADD scores between closely related species, future work should examine the relationship between CADD scores and observed reduction in fitness. CADD scores were correlated between all three model species (Figure 12) showing that a deleterious mutation with high CADD scores in one species is likely to also receive a high CADD score in the other two species. Future investigations into what phenotypic

changes result from these mutations could help to verify if high CADD scores translate into deleterious mutations. In species where a large amount of fitness data is available, it would be interesting to investigate if a decrease in fitness is observed in individuals who possess high CADD score mutations. This would help to provide a wider understanding of how quantifications of genetic load using CADD scores relate to individual fitness. Although CADD scores calculated on model species can be transferred to threatened species, raw CADD scores are not an appropriate measure of genetic load as they are log-converted scores (Rentzsch et al., 2019). Therefore, CADD scores should be converted to selection coefficients which can be added to calculate the lethal equivalents an individual possesses.

3.13 Conclusion

To conclude, CADD scores calculated in model species can be applied to threatened species using LoadLift to quantify the genetic load of individuals within the UCEs. Although not produced for the specific species, these scores show similarity across three model species: humans, pigs and chickens (Figure 10). Therefore, we can be confident in applying them to evaluate threatened species that are less diverged. CADD scores are, on average, higher in the UCEs than their flanking regions in all three species, and most UCEs are within non-coding regions. However, CADD scores are higher within coding regions than non-coding regions (Figure 13). In all three species, CADD scores appear to show a similar pattern at a specific gene locus, showing a split of high and low CADD scores, suggesting third codon positions (Figure 11). The strong correlation of CADD scores for all UCEs across all pairwise comparisons suggests that although exact values for individual scores may differ between species, high CADD scores in the UCES of one species equate to a high CADD score in another. This thereby supports the use of LoadLift as a tool to apply CADD scores for the use of assessing genetic load within threatened species. This will enable conservation managers to make informed captive breeding decisions to conserve the “genetic health” of the population for the future.

4. Conversion of CADD scores to selection coefficients for load calculations

4.1 Abstract

CADD scores transferred from model species to endangered species using LoadLift allow for the assessment of an individual's genetic health. The resultant mutation impact scores can be summed and compared between individuals. However, CADD scores are calculated as the log transformation of the rank of how deleterious a mutation is predicted to be, compared to all other mutations in the genome. This means that the deleterious impact of having two mutations is not the same as that of a single mutation with the value of the two CADD scores summed together. Therefore, for load comparisons, selection coefficients for each mutation need to be calculated. Here, I display two methods to convert CADD scores into values that can be used to calculate load components. Firstly, I will convert the CADD scores to the rank score of the mutation within all possible variants in the genome. These ranks can then be summed to calculate selection coefficients while retaining the more deleterious impact of higher CADD scores. Secondly, I will use selection coefficients from SLiM simulations. SLiM was used to simulate neutral and deleterious mutations under five Distribution of Fitness Effects (DFE) alternatives. I correlate the simulated selection coefficients to the raw CADD scores, thereby creating a function with which future investigations using CADD scores can infer selection coefficients. These selection coefficients are then used to calculate "Lethal Equivalents" for sites within the UCEs of six pink pigeons. This analysis shows that selection and dominance coefficients for deleterious mutations can be inferred from CADD scores, allowing for the calculation of genetic load, realised load and masked load in threatened species.

4.2 Introduction

Using mutation impact scores such as CADD to quantify the genetic load within individuals of threatened species can provide a valuable resource for captive breeding managers (Chapter 2). CADD scores can provide per-base values for potential deleterious mutations that, when compared between individuals, can indicate the potential genetic health of

individuals. This can provide a great resource and allow for a comparison of deleterious mutations within UCEs and flanking regions (Chapter 2). However, the genetic load components cannot be estimated simply by summing CADD scores (Speak et al., 2024). CADD scores are the logarithmically converted rank of how deleterious each mutation is relative to all possible mutations in the entire genome (Kircher et al., 2014; Rentzsch et al., 2019). For example, CADD scores between 0 and 10 represent 90% of all potential mutations. Meanwhile, CADD scores greater than 20 represent the most deleterious 1% of all possible mutations, and a CADD score of 99 is the single most deleterious mutation. Therefore, if an individual possesses two separate mutations which both have a CADD score of 5, its genetic load is not equivalent to an individual that possesses a single mutation with a CADD score of 10. It is therefore pertinent when using CADD scores to accurately quantify the genetic load for conservation applications to ensure that the data and suggestions that are shared with population managers are not misleading.

For accurate calculations of genetic load, realised load and masked load based on CADD scores, we first require the conversion of CADD scores into coefficients for selection (s) and dominance (h) for each variant. These can be expressed in Lethal Equivalents (LE) (Morton et al., 1956) (see section 1.8 of this thesis). Here we use the definition that a lethal equivalent is a group of mutant alleles with a summed selection coefficient equal to one (Bertorelle et al., 2022). Therefore, the load of a population was expressed by grouping genes or mutations that together caused, on average, the death of a given number of individuals (see Bertorelle et al., 2022; Dussex et al., 2023). LEs are mainly a theoretical value of the predicted total genome-wide genetic load within a population. Previous studies on the pink pigeon have predicted that the species possesses a genetic load of 12 LEs (Jackson et al., 2022). The challenge is to convert mutation-impact scores such as CADD (and GERP) into selection and LEs.

In this chapter, I will explore two possible solutions to convert CADD scores to selection coefficients. Firstly, I will examine whether before summing these values, it would be more accurate to first calculate the actual rank score of a variant based on its CADD score. Given that CADD scores are log-transformed values, it is likely that calculating the rank score of

variants before summing these values will give a more accurate reflection of the relative genetic load between individuals. Second, we will use the distribution of simulated selection coefficients (s) from SLiM simulations, again giving these rank scores (i.e., s-rank). We will then align the rank of the CADD score CADD-rank) to the s-rank of the s coefficient, thereby converting CADD into s coefficients. These two approaches will then be compared to visualise the impact of different conversion methods on the estimated load of individuals. The rationale is that these summed mutation-rank scores are better comparable because they are at a linear scale. I will test this by calculating the genetic load components in simulated selfing crosses of the six individuals analysed in Chapter 2. We know that selfing should increase the inbreeding coefficient F by 50% (F=0.5). This level of inbreeding results in a quantifiable increase in the realised load (see Bertorelle et al., 2022) We can thus calculate the increase in realised load in simulated crosses based on summing the rank-scores. We can compare that to the increase in realised load based on summing the CADD scores to examine which values fit the theoretical increase in realised load best.

This chapter aims to demonstrate that CADD scores lifted over from model species can be converted to selection and dominance coefficients simulated in SLiM (Haller & Messer, 2019). This will allow for more accurate calculations of the genetic, realised and masked load. These mutation impact scores can then be used to also calculate the load of potential offspring between mate pairings as demonstrated in Chapter 2. This allows for the comparison between individuals and to identify optimal mate pairings.

4.3 Material and methods

4.4 LoadLift and filtering

CADD scores were lifted over to the UCEs extracted from the genomes of six pink pigeons from the captive-bred population, using the LoadLift pipeline (Chapter 2.) (Speak et al., 2024). The scores were then filtered using R to remove any non-scoring sites (sites where there were no differences from the chicken wild-type allele in all six pigeon individuals) and

sites that were fixed as homozygotes within all six individuals. UCEs with a SNP frequency over 0.35 were removed. After filtering 680 unique CADD scores were identified ranging between 0.0013 and 31.4595. This represents mutations that are the 0.03015932% and 99.92854148% most deleterious mutations respectively of the 3,073,805,640 potential CADD scores.

4.5 CADD score rank

The 680 unique CADD scores were converted to their relative rank within all potential CADD scores in the whole chicken genome. chCADD scores are defined in Groß et al., 2020 as:

$$chCADD_i = -10 \log_{10} \frac{n_i}{N} \quad [6]$$

Where n_i is the rank of the potential mutation and N is the 3,073,805,640 potential mutations across all chromosomes. chCADD scores were converted to determine proportion ($\frac{n_i}{N}$), for example, a chCADD score of 20 equates to a $\frac{n_i}{N}$ of 0.01. A scaled conversion was applied $\frac{1}{n_i}$ to act as a proxy for the selection coefficient.

4.6 Selection coefficient simulation

Selection coefficient simulations were performed by Hernán E. Morales. We performed individual-based forward simulations with SLiM 3.1 (Haller & Messer, 2019). We modelled neutral genetic variation and a genetic load as in Jackson et al. (2022). The simulations are based on a non-Wright-Fisher implementation, which considers overlapping generations, age structure, and customizable offspring generation and migration patterns. During the simulation, each time step consists of three stages: reproduction, dispersal (between captive

and wild populations, if any), and mortality. Absolute fitness (i.e., probability of survival) was regulated by the carrying capacity and the known aged-based probability of mortality for pink pigeons (see supplementary information in Jackson et al., 2022). We simulated an ancestral population of 16,000 individuals (Ryan, 2021) that had a slow population collapse followed by a severe recent bottleneck. The simulated demographic trajectory was informed by the inferred (pre-1980s) population size and recorded census trajectories from 1980 to 2020. Reproductive age, fecundity, and mortality were drawn from a distribution that reflected the productivity of pink pigeons in the wild (see supplementary information in Jackson et al. 2022). This resulted in an average generation time of 3.5 simulation steps, similar to the generation time of the pink pigeon.

We simulated 4000 genes of 3400 bp each distributed proportionally to the number of genes contained across 28 autosomal chromosomes in the collared flycatcher genome (Kawakami et al., 2014). We thus simulated a total genomic length of 13.6 Mb. We used a recombination rate of $1x10^{-4}$ per base position per generation, with no recombination within genes. We simulated neutral and deleterious mutations at relative proportion of 1:2.3 with selection coefficients taken from different Distribution of Fitness Effects (DFE) alternatives as explained in the supplementary information of (Dussex et al., 2023).

4.7 CADD score to selection coefficient conversions

The 680 unique CADD scores were converted to their relative rank within all potential CADD scores in the whole chicken genome. Selection coefficients and their respective dominance coefficients were similarly ranked, and the corresponding rank selection coefficients were applied to the CADD scores. CADD scores for the SNPs within the UCEs of the six pink pigeon individuals were then converted to the relative selection and dominance coefficients.

4.8 Potential offspring crosses

Potential offspring crosses for all crosses of the six pink pigeons including selfing were simulated using the Hardy-Weinberg equilibrium and the genotypes of both potential parents to inform the ratios of potential offspring genotypes. This was then used to calculate the homozygote and heterozygote load scores for each SNP. The genetic, realised, and masked load were then calculated using the following formulas (Bertorelle et al., 2022):

$$Genetic\ load\ (individual\ k) = \sum_{i=1}^{L(hom)} s_i + \sum_{j=1}^{L(het)} 0.5\, s_j \quad [7]$$

$$Realised\ load\ (individual\ k) = \sum_{i=1}^{L(hom)} s_i + \sum_{j=1}^{L(het)} h_j s_j \quad [8]$$

$$Masked\ load\ (individual\ k) = \sum_{j=1}^{L(het)} (0.5 - h_j) s_j \quad [9]$$

4.9 Realised load under selfing

Under selfing, there should be an increase in the inbreeding coefficient F by 50% (F=0.5). A 50% increase in the inbreeding coefficient will theoretically lead to the realised load being doubled (Bertorelle et al., 2022). The realised load for the six pink pigeons and simulated selfing crosses were calculated using the methods defined above (Equation 8). These values were then compared to determine the increase in realised load relative to the increase in the inbreeding coefficient between the six individuals and selfing crosses (Supplementary Table 10).

4.10 Results

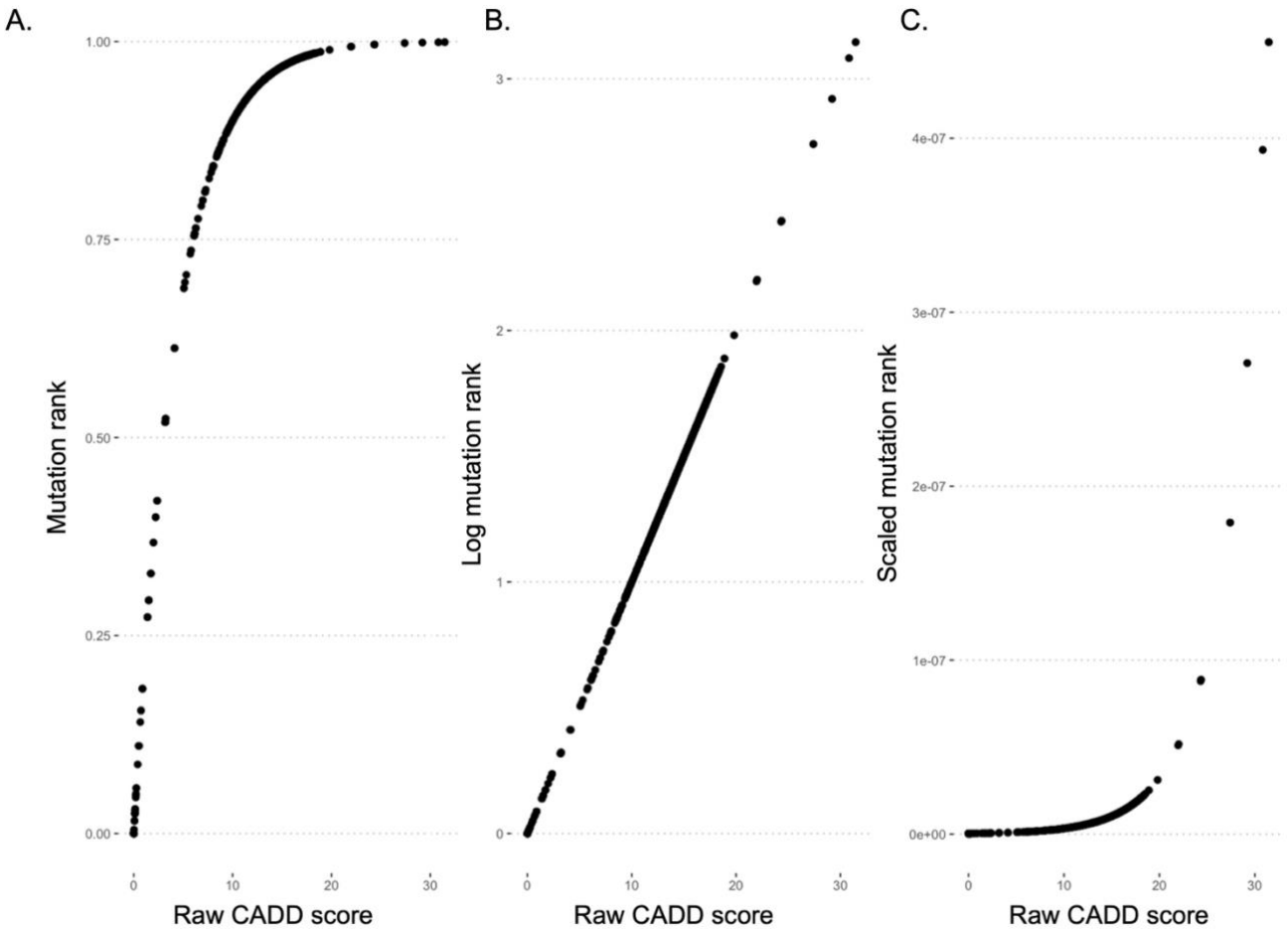


Figure 14 – Raw CADD score conversions to determine proxies for the selection coefficients. A) The CADD scores were converted to the rank of the mutation within all 3,073,805,640 possible substitutions. **B)** the log conversion of the rank of the mutation. **C)** a scaled conversion $\frac{1}{n_i}$ for each mutation.

CADD scores can be converted to their original rank within the genome using Equation 1 (Figure 14A), these values are therefore able to be used to create scaled mutation ranks that can be summed whilst retaining how deleterious each variant is to each other (Figure 14C). These values are scaled between 0 and 1, whereby 1 is the most deleterious mutation within the entire genome, one which would possess a CADD score of 99.

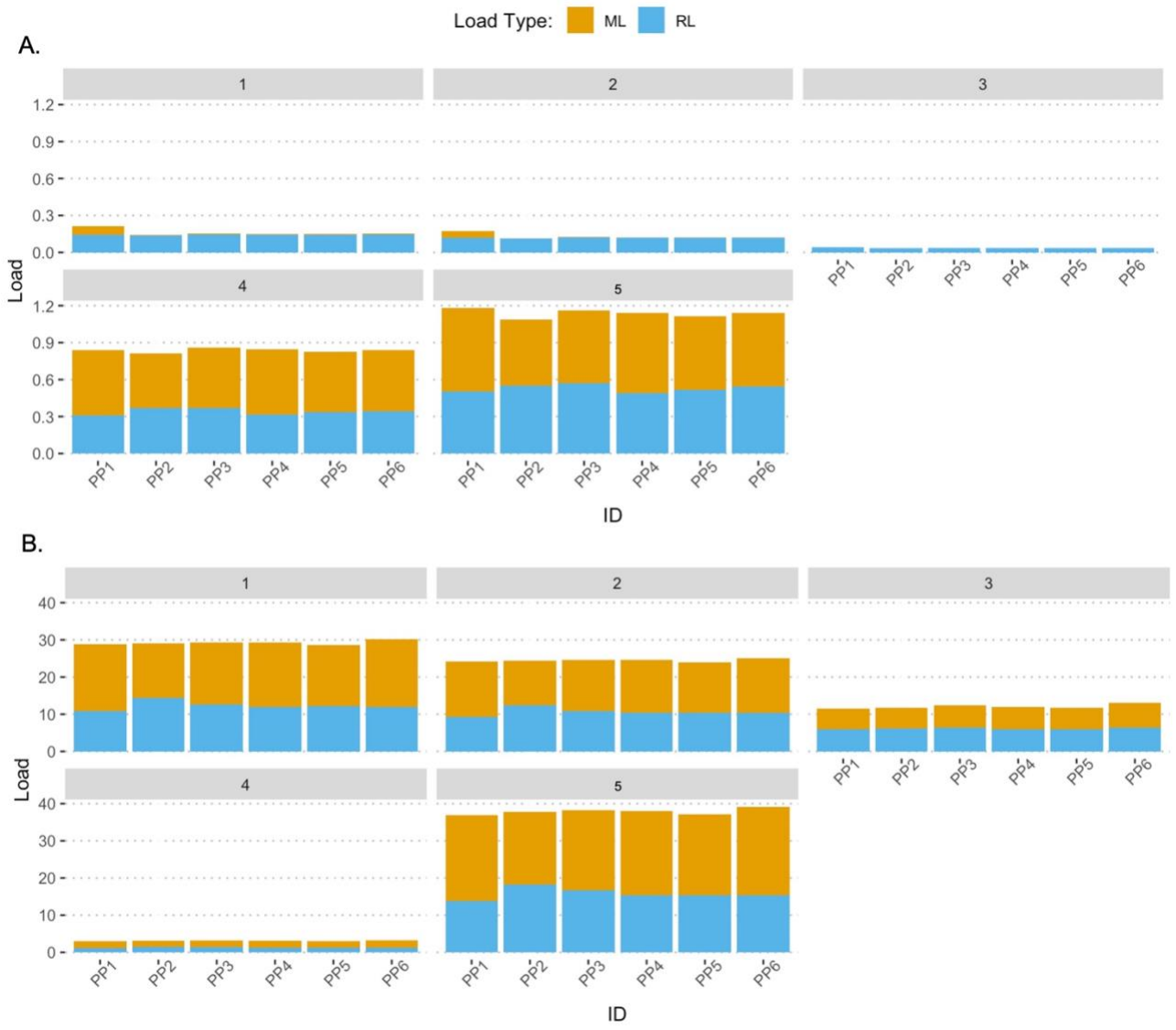


Figure 15 – A comparison of load components calculated for individuals across five different Distribution Fitness Effect scenarios (DFEs). The masked load (ML) (Orange) and realised load (RL) (Blue) in the five different DFEs for the six pink pigeon individuals, using simulated selection and dominance coefficients. **A)** Whereby neutral mutations as well as harmful mutations were simulated. **B)** Whereby only harmful mutations were simulated. Load was calculated as the sum of the selection coefficients for each individual. Selection coefficients were simulated for 680 sites using SLiM (Haller & Messer, 2019) following the historical demography of the pink pigeon (Jackson et al. 2022).

Five separate DFE were simulated using the demographic trajectory of the pink pigeon (Supplementary Figures 3 & 4). When only the harmful variants from the simulations were ranked and assigned to the same rank CADD score SNP position, the total amount of genetic load within the UCEs was over ten for all individuals in four DFE. This would equate to 10 lethal equivalents in all individuals which would be unrealistic. In comparison, when also including neutral variants, three DFE (DFE 1, 2 & 3) predicted no realised load in most of the individuals (Figure 15A), and hence, these three scenarios were deemed to be unlikely. DFE 5 is the most realistic scenario as it showed consistent load calculations when using both neutral and only harmful selection coefficients. This DFE results in a genetic load of between 1.37 and 1.88 LE across the six individuals (Supplementary Tables 8 and 9).

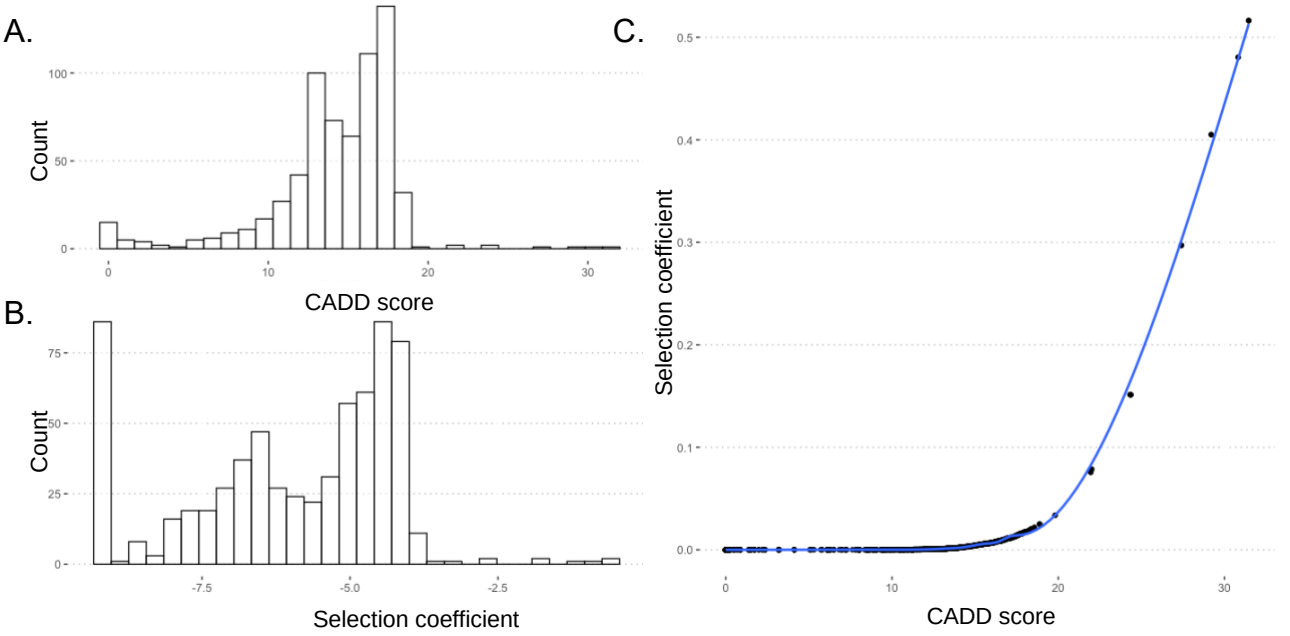


Figure 16 – A comparison of the CADD and simulated selection coefficients for 680 SNPs within the UCEs of the pink pigeon. A) Distribution of CADD scores within the UCE for 680 SNPs not fixed within all six pigeons. **B)** The distribution of 680 simulated selection coefficients. **C)** Selection coefficients and CADD scores were ranked due to how deleterious they were, and the corresponding ranked CADD and selection coefficients were assigned to each other. A regression line was applied using the gam function (blue). CADD scores were lifted over from the model chicken (chCADD) onto the subject pink pigeon species using the LoadLift (Speak et al., 2024) pipeline. Selection coefficients were simulated for

680 sites using SLiM following the historical demography of the pink pigeon (Jackson et al. 2022).

The distribution of CADD scores and selection coefficients are similar with a comparable peak around a CADD score of 15-20 and a similar drop in the number of scores over 20 (Figure 16A & B). There is a strong correlation between the CADD score and the selection coefficient (Figure 16C) (GAM, $R^2 = 0.999$, $p < 2.2e\text{-}16$, $n = 4036$). This suggests that CADD scores can be successfully converted to simulated selection coefficients.

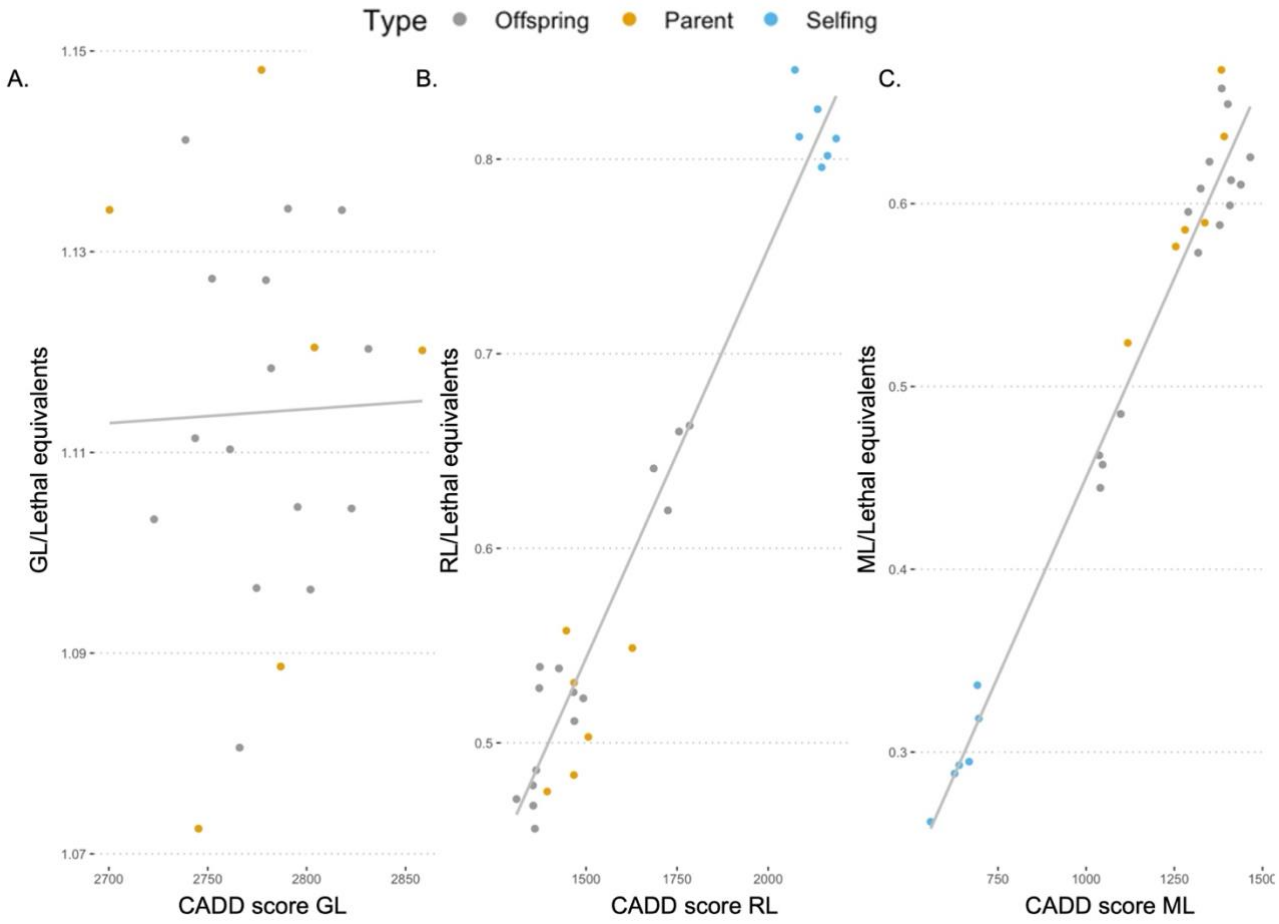


Figure 17 – The potential offspring load for 650 sites were calculated using CADD scores lifted from the model chicken to six pink pigeon genomes UCEs using the LoadLift pipeline (Speak et al., 2024), these were ranked and converted to equal ranked selection coefficients simulated in SLiM. A) The genetic load, **B)** realised load and **C)** masked load were calculated in lethal equivalents and as summed CADD scores. Simulated crosses between the same parent individuals "Selfing" (blue) and the load values for the original six "parent" (orange) are shown along with potential "offspring" of crosses between all six individuals (grey).

There is a strong correlation between the sum of CADD scores and the summed selection coefficients (Lethal equivalents) that individuals possessed (Figure 17) for the realised load (Regression analysis: $R^2 = 0.9462$, $F_{1,40} = 721.5$, $p < 2.2e\text{-}16$) and masked load (Regression analysis: $R^2 = 0.$ 0.9654, $F_{1,40} = 1145$, $p < 2.2e\text{-}16$). All predicted offspring's genetic load was between that of the maximum and minimum of the parents (Figure 17A). The realised

load was highest in the selfing crosses (Figure 17B) as with selfing there would be an increase in homozygosity and therefore more masked load would be converted to realised load. Similarly, the masked load was lowest for the selfing crosses (Figure 17C).

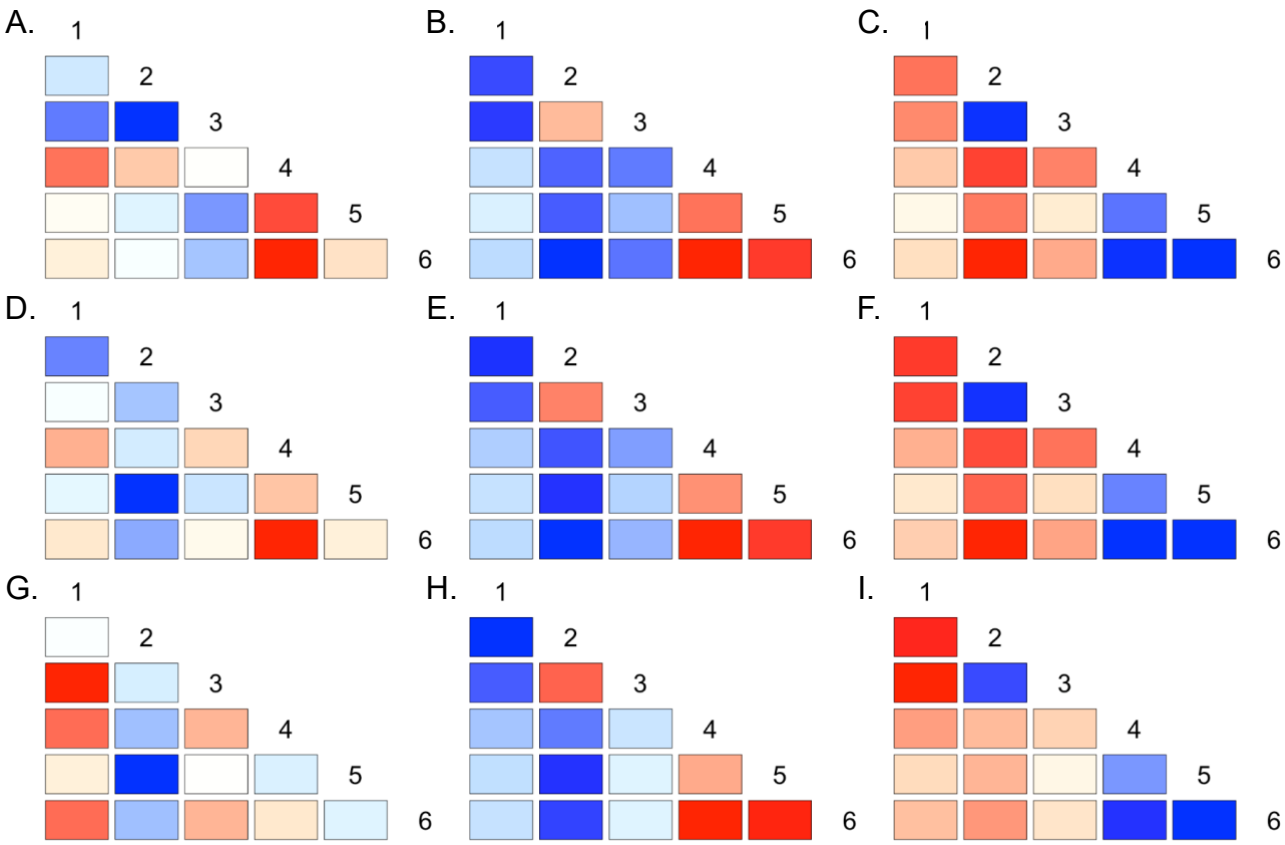


Figure 18 – Potential offspring load matrix calculated using CADD scores lifted from the model chicken to six pink pigeon genomes UCEs using the LoadLift pipeline (Speak et al., 2024). Comparing the values calculated using summed CADD scores for the **A)** total genetic load, **B)** realised load and **C)** masked load. In comparison values calculated using summed scales rank conversions values derived from CADD scores, for the **D)** total genetic load, **E)** realised load and **F)** masked load. Calculated using summed selection coefficient values simulated in SLiM and substituted for their ranked CADD score, **G)** for the genetic load**, H)** realised load and **I)** masked load.

The optimal mate pairings between individuals across all samples were made using all three methods for the calculation of the genetic load components, using summed CADD scores (Figure 18A, B & C), scales rank conversions (Figure 18D, E & F) and SLiM simulated selection coefficients (Figure 18G, H & I) allowing for hypothetical crosses between all six pink pigeon individuals.

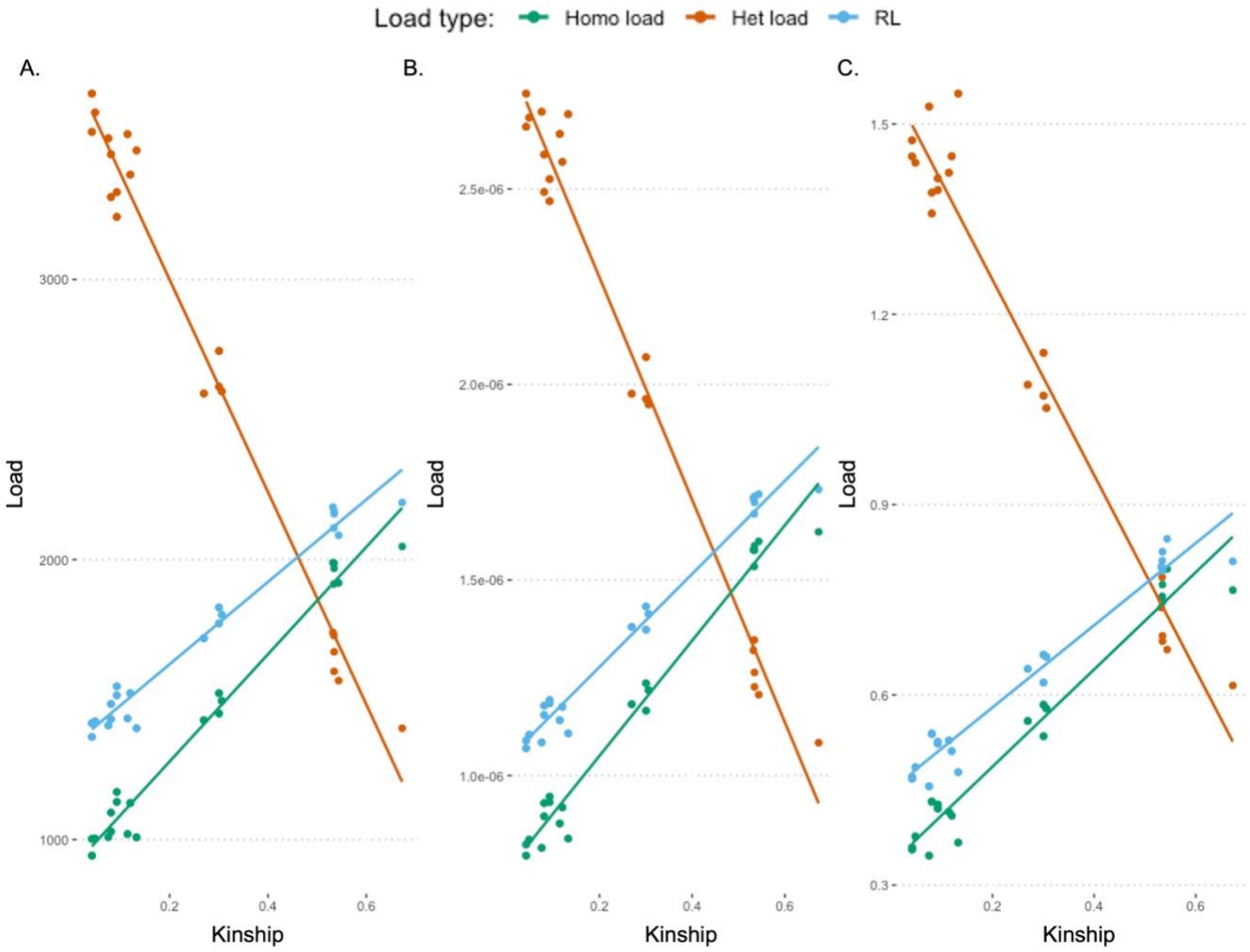


Figure 19 – A comparison of the Heterozygous load conversion within the potential offspring of crosses between the six captive-bred pink pigeons. A) The conversion of load when calculated using the raw CADD scores, **B)** the conversion of load calculated using a scaled conversion of the mutations rank and **C)** calculated using selection and dominance coefficients simulated in SLiM (Haller & Messer, 2019). Heterozygous load (Orange), Homozygous load (Green) and realised load (Blue).

All three methods of load calculation show increasing realised load and homozygous load with increasing kinship between parents (Figure 19), as kinship increases more loci become homozygous due to the higher chance that both individuals have the derived allele, leading to an increased homozygosity and realised load. The increase in realised load between the least related and highest related individuals, when calculated from the CADD scores, is 61.2% (Figure 19A). This increase is 61.8% when calculated with the scaled rank of the mutation (Figure 19B) and 85.6% when using the selection coefficients from the SLiM simulations (Figure 19C). The average ratio of load between the realised load for the six individuals and that of offspring from selfing crosses was highest for the SLiM simulated selection coefficients at 1.58. Meanwhile, the scaled rank of the mutation had a mean ratio of 1.34 and the CADD scores had a mean ratio of 1.41.

4.11 Discussion

CADD scores can be lifted over from non-model species to quantify the genetic load of individual members of a threatened species (Chapter 2). However, as the CADD scores are scaled ranks of the mutation within the entire genome, scores are not able to be summed to calculate load components. Here, I have ranked the CADD scores that are found within the UCEs of six pink pigeon genomes using five different DFE, and simulated selection and dominance coefficients in SLiM (Figure 15, Supplementary Figure 4). I have converted the CADD scores to selection and dominance coefficients (Figure 16). The selection coefficients were then summed to allow for the calculation of the realised, masked, and total genetic load present within the UCEs of each individual as Lethal Equivalents (LE). I correlated the load components calculated as LEs to the summed CADD scores (Figure 17) and showed that the summed values for realised load and masked load correlated highly, but that the total genetic load varied between the three methods (Figure 18). I also converted the CADD scores into their raw scores as the rank of the mutation within the genome and created a scaled rank score as a potential means for conservationists to mimic selection coefficients without SLiM simulations (Figure 19). I compared these three methods of calculating the realised and masked load (Figure 18) and showed that the SLiM simulated selection coefficients were the most accurate means of calculating the load components (Figure 19).

The difference between the calculation of realised load for the six individuals and their selfing crosses was the closest to the increase expected under selfing. This was likely due to the SLiM simulations providing both selection and dominance coefficients for all mutations, whilst the scaled rank conversion and CADD score values both used the same dominance coefficient for all mutations (h = 0.1). Therefore, future investigations should adjust the dominance coefficient with the increase in CADD score, with higher dominance coefficients for less deleterious mutations, and low dominance coefficients for highly deleterious mutations.

Selection coefficient calculations have been completed for a wide variety of species and populations (Thurman and Barrett, 2016). This calculation, however, requires a large amount of fitness data, allele frequencies and multiple generations of individuals (Tataru et al., 2017), which is only possible for a limited number of model species. Therefore, the use of SLiM simulations allows for selection coefficients to be derived for mutations within the genomes of threatened species where this data is not available. Being able to infer selection coefficients for individuals using CADD scores, allows for more accurate calculations of genetic load components which can be used by conservation managers to inform their breeding programmes. Using this methodology, population managers will be able to reduce the predicted realised load of future generations.

The total amount of genetic load calculated for each individual's UCE region ranged between 1.37 and 1.88 LE (Supplementary Table 8). With the UCEs equating to 0.5% of the genome, this may initially be seen as a potentially high estimate of the number of LE as previous studies in the pink pigeon suggest that they possess 12 LE across the entire genome (Jackson et al., 2022). However, mutations within the UCEs do equate to higher CADD scores in both the pink pigeon (Chapter 2, Figure 5) and all three model species that have available CADD scores, humans, pigs and chickens (Chapter 3, Figure 10E) and due to this are more likely to represent a greater quantity of load relative to the rest of the genome. Therefore, the estimate of 1.88 LE within the UCEs of an individual when summing the selection coefficients may be accurate as this equates to an assessment of some of the regions of the genome whereby a mutation would be extremely deleterious. Using the

conversion curve of selection coefficients and CADD scores (Figure 16C), CADD scores over 20 would equate to a selection coefficient of greater than 0.037. SLiM simulations of the Seychelles paradise flycatcher, a species which also underwent a population decline and recovery, suggested that large ancestral population sizes influence the genetic load of the populations during and post-recovery from a population bottleneck (Femerling et al., 2023). As the pink pigeon had a large ancestral population of around 16,000 individuals (Ryan, 2021) and suffered a severe population decline, the high levels of genetic load calculated are possible.

In this chapter I have used simulated selection coefficients extracted from SLiM simulations that were performed, imitating the demographic history and population bottleneck of the pink pigeon. However, the simulated data was based on the 13.6 Mb chromosomal assembly of the collared flycatcher. Future investigations could be performed using the newly published pink pigeon reference genome (Morales et al., 2024). In this way, the simulated selection coefficients would be more accurate and any bias that would be introduced to the investigation due to the divergence between the collared flycatcher and the pink pigeon could be avoided. Simulating using the whole pink pigeon genome would also allow for the selection coefficients for the precise mutations to be extracted rather than the ranking of the selection coefficients and the CADD score rank within the chicken genome. This would allow for a higher confidence in the selection coefficient conversion. For other study species where a large quantity of sequence and pedigree data is available across multiple generations of the captive-bred population, it would provide an opportunity to perform a gene drop analysis (Honda et al., 2002; Doekes et al., 2020). In this way, specific mutations could be tracked across generations and provide more accurate selection and dominance coefficients that could be used to inform and check the simulated coefficient values.

4.12 Conclusion

In conclusion, CADD scores can be lifted over from model species to threatened species allowing for comparison of genetic load between individuals (Speak et al., 2024). However, as CADD scores are log-converted ranks of how deleterious a specific mutation is in the

genome, summing CADD scores will not give a true interpretation of genetic load. For more accurate calculations of genetic load, realised load and masked load using the equations in Bertorelle et al. (2022), selection and dominance coefficients should be used. I have shown in this chapter how selection and dominance coefficients can be inferred from SLiM simulations and applied to CADD scores that have been transferred to threatened species. I compared two potential methods for creating selection coefficients those derived from slim simulated data and those derived from the scaling the rank of CADD score within all 3,073,805,640 potential mutations within the chicken genome. Calculations of the genetic load components within the genomes of the six pink pigeons showed that they possessed between 1.37 and 1.88 LE, previous work on the pink pigeons from wild samples calculated that the population possessed 12 LE across the genome. This therefore suggests that although the UCEs are only 0.5% of the genome, due to the highly deleterious nature of mutations present within these regions (Chapter 3), they equate to a large proportion of the genetic load within individuals. Within individuals, most of the genetic load is masked load and therefore not expressed (Figure 15). Therefore, quantifying the genetic load within individuals in captive breeding populations and calculating the optimal mate pairings within the population can reduce the realised load of individuals in future generations.

5. Conservation genomics of the previously endangered whooping crane (*Grus americana*)

5.1 Abstract

In this Chapter, I apply the techniques developed in Chapters 2, 3 and 4 to assess the genetic load in a similarly threatened species, the whooping crane (*Grus americana*), using the LoadLift pipeline. The whooping crane, native to North America, suffered an intense population bottleneck in the 20th century declining to sixteen individuals. Through intense conservation efforts including habitat conservation, establishing captive breeding populations and reintroducing individuals into the wild, the population has now recovered to 800 individuals. Here, the LoadLift pipeline is used to assess the genetic load within 37 whooping crane individuals across wild and captive-bred populations by applying CADD scores derived from model species. I also simulate mate pairings between all individuals to determine the optimal crosses. I demonstrate that the whooping crane possesses a relatively large realised load (RL) compared to its masked load (ML). The RL/ML ratio exceeds unity, which is unusually high. It likely reflects that besides the high levels of inbreeding during the population bottleneck, the realised load has been elevated by drift during prolonged periods of population size decline. A comparison of methodologies for calculating genetic load between SNPeff and CADD scores transferred using LoadLift was conducted to determine if the methodologies rank mutations similarly within the UCEs of the whooping crane. I show that mutations classified by SNPeff as highly and moderately deleterious have significantly higher average CADD scores than those classified as less deleterious by SNPeff, i.e., highly (mean CADD = 23.466), moderately (mean CADD = 20.718), and low (mean CADD = 2.604). Finally, I discuss the value of converting the mutation categories of SNPeff and CADD scores into selection coefficients, and how that would help in the assessment of the genetic load, the fitness of individuals, and the extinction risk of populations.

5.2 Introduction

Of the 163,040 species assessed by the International Union for Conservation of Nature (IUCN) Red List, 28% are threatened with extinction globally (IUCN 2024). Currently, over 400 species are part of EAZA *Ex situ* Programmes (EEPs) (EAZA, 2024b) with their captive breeding populations acting as insurance populations against extinction caused by the ongoing threats to their environment. Of these captive populations, very few possess whole genome sequenced individuals.

Conservation genomics has principally focussed on populations of species in the natural environment and in reserves. For example, the kākāpō (*Strigops habroptilus*) is a ground-dwelling parrot endemic to New Zealand and is considered the genomic gold standard for conservationists as nearly all of the extant individuals have been sequenced (Guhlin et al., 2023). Other bird species that have been genome sequenced to assess genetic diversity and population viability including the echo parakeet (*Psittacua echo*) (Tollington et al., 2013), pink pigeon (*Nesoenas mayeri*) (Jackson et al., 2022; Speak et al., 2024), burrowing owl (*Athene cunicularia*) (Mueller et al., 2018), hihi (*Notiomystis cincta*) (Nichols et al., 2024), Seychelles warbler *(Acrocephalus sechellensis)* (Gilroy et al., 2017; Wright et al., 2014), Seychelles magpie robin (*Copsychus sechellarum)* (Femerling et al., 2023), Swinhoe's white-eye (*Zosterops simplex*) (Wu et al., 2024), black grouse (*Lyrurus tetrix*) (Wang et al., 2014; Chen et al., 2023), rock dove (*Columba livia*) (Smith, 2023; Smith & Clegg, 2023) and many more as part of the Bird 10000 genomes project (B10K) (Feng et al., 2020). Many of these species have suffered from a population bottleneck, and some have recovered with conservation assistance, resulting in their down-listing on the IUCN Red List (IUCN, 2024).

The whooping crane (*Grus americana)* is endemic to North America, where the species historically existed in a large population across much of modern-day Canada and the United States of America. During the 20th century, due to a combination of over-hunting and the loss of native habitat and nesting grounds, the population suffered a large demographic collapse to around 16 wild individuals in the 1940s (Butler et al., 2013; Smith, 2019). Unlike most other bird species, the population crash and recovery happened around 80 years ago, giving us a longer-term perspective of the processes involved (see section 1.25 for further information). The population of whooping cranes has now recovered to over 800 individuals

due to intense conservation efforts. These actions rescued the last surviving wild population at Aransas-Wood Buffalo, and they established a captive bred population. Despite this recovery, the species is still threatened with extinction due to man-made threats. The most severe threats to the whooping crane are collisions with powerlines and turbines whilst migrating, predation on chicks and eggs, and environmental changes caused by droughts, which threaten their vital overwintering habitats (Golden et al., 2022). In addition to these external threats to species survival, there is also a potential risk of inbreeding depression. For example, research on the pink pigeon has shown that even during population recovery, previously bottlenecked populations continue to lose genetic diversity due to a drift debt (Jackson et al., 2022; Pinto et al., 2024).

The captive population of the whooping crane provides a rare opportunity to analyse the longer-term consequences of a demographic recovery after a population bottleneck. The species also has a range of genomic samples readily available to apply the LoadLift pipeline to. This includes whole genome sequenced data for individuals sampled from: historical museum taxidermy samples killed before the population bottleneck; individuals from the original captive founder generations and each generation in captivity thereafter; today's captive population and today's wild population. This wide range of data allows for the assessment of the genetic load of individuals in the population, a comparison of the load between the wild and captive populations, and an assessment of how much of the load is realised versus masked. By analysing temporal samples, it is also possible to assess how this ratio between the realised and masked load has changed over time. This will therefore allow conservation managers to assess the risks of the species to inbreeding depression now that the demographic rescue appears to have been successful. This will also provide an opportunity to simulate crossings between individuals from the wild and captive populations and compare potential crosses to select those that would have the lowest realised load in future generations. This would provide an example for conservation managers to visualise how they could compare and select individuals for mate pairings as well as for reintroduction programmes.

Over recent years, there have been many advances in bioinformatic techniques to study the large quantity of genetic data that is more easily produced due to decreases in the costs of

sequencing technology. There are multiple recently developed techniques to calculate and compare the genetic load of species including: PhyloP (Pollard et al., 2010), GERP (Davydov et al., 2010), SNPeff (Cingolani et al., 2012) and CADD (Kircher et al., 2014; Rentzsch et al., 2019). GERP scores are calculated as the number of substitutions observed minus the number of substitutions expected under a neutral model (Davydov et al., 2010). In this way, a GERP score can be seen as the measure of how conserved a sequence is across the species investigated. Positive GERP scores represent regions where substitutions are observed at a lower rate than expected under a neutral model (Huber et al., 2020). SNPeff does not provide scores but predicts the effects of genetic variants and categorises them based on their likely fitness consequences: high, moderate and low impact mutations. SNPeff does this by using the reference genome and annotation information provided by a user to identify when a variant is present within a region with annotated data and predicts the likely result of this mutation (Cingolani et al., 2012). SNPeff is a widely used approach with over 10,000 citations, primarily due to its speed, flexibility and integration with other widely used tools (Cingolani et al., 2012). Therefore, it is appropriate that SNPeff and CADD methodologies are compared in this study, with consideration of the application of both approaches when looking to assess the genetic load of endangered species.

This Chapter investigates the genetic load present within 37 whooping crane samples using CADD scores transferred using the LoadLift pipeline. Using this dataset, we demonstrate that the pipeline is applicable to species besides the pink pigeon. We also compare the genetic load of individuals of the wild and captive-bred populations. We compare the load components at different generations in the captive-bred population. We furthermore identify optimal mate pairings across all individuals to minimise inbreeding depression in the offspring generation (Speak et al. 2024; Chapter 2). Finally, we compare the methods of calculating the genetic load using SNPeff and CADD scores from LoadLift and discuss the value of converting the mutation categories of SNPeff and CADD scores into selection coefficients.

5.3 Material and Methods

5.4 Study species

Blood samples were taken from 37 modern whooping cranes (*Grus americana*), of which 19 were from “Wild” individuals. The remaining 18 were from the captive population across three separate time points six from the founder individuals of the captive bred population, six from the F1 captive generation “Early” and six from amongst the F2 and F3 generations “Late”.

5.5 Modern whooping crane sample preparation.

Library preparation and sequencing were carried out by Molly Cassatt-Johnstone of the University of California, Santa Cruz, they extracted modern whooping crane samples from blood stored on FTA cards using the DNEasy Blood & Tissue kit (cat #69506) as described in the protocol for nucleated blood. Briefly, they punched two 2 mm holes from the blood spot on each FTA card using a hole punch. They cleaned the hole punch between samples with 2% NaClO and then with 70% EtOH. The punched-out blood spots were added to one 1.5 mL tube per sample with 20 µL proteinase-K and buffer PBS was added until the final volume was 220 µL. Finally, they performed the DNA isolation as described in the protocol.

They converted the modern extractions into libraries using the NEB Ultra II FS kit (cat #E7805L) with enzymatic shearing. We used 100ng of DNA as input, shearing extracts for 14 minutes at 37°C, and followed the protocol as described. They size selected for final library products of 270-370 bp and amplified the libraries for 5 cycles. They pooled the libraries at an equimolar ratio and sequenced the libraries on one lane of 2x150 S4 Illumina NovaSeq 6000 run at Duke University.

5.6 FASTQ trimming, mapping and quality control

FASTQ trimming, mapping and quality control was carried out by Hernán Morales of the Globe Institute, Copenhagen University. Reads were trimmed and sequencing adapters

were removed from Raw FASTQs with SeqPrep2 (https://github.com/jeizenga/SeqPrep2), retaining reads longer than 30 base pairs and with mapping quality (MQ) higher than 20. Sequences were aligned to the chromosome-level assembly of the whooping crane (GCF_028858705.1) (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_028858705.1/) using BWA 0.7.17 *mem* (default parameters) for modern samples and *aln* (-l 1024 -n 0.03 -o 2) for historical samples (Li and Durbin, 2009). PCR duplicates were removed with picard MarkDuplicates 2.27.5 (http:// broadinstitute.github.io/picard/)

5.7 Variant calling and filtering

Variant Calling was carried out by Claudia Fontsere of the Globe Institute, Copenhagen University. We used snpAD v0.3.1.10 (Prüfer, 2018) to call genotypes in all samples on each autosomal chromosome. Next, we merged individual VCFs per autosome in a single file with bcftools v1.20 (Danecek et al., 2021) *merge* option. We kept only those variants with a read depth between 4 and 50 (included) and a minimum genotype quality of 30 with bcftools (Danecek et al., 2021) *filter*. Finally, we concatenated all VCFs with bcftools (Danecek et al., 2021) *concat*.

5.8 LoadLift CADD score analysis

Using the LoadLift Snakemake pipeline (Speak et al., 2024) (Chapter 2, Figure 4) a chain file was created for the chicken (*Gallus gallus*) and whooping crane reference genomes, for which the read data had been aligned. The CADD scores calculated for the chicken (chCADD scores) were then crossmapped to the annotation of the whooping crane reference genome using crossmap.py (Zhao et al., 2014).

Ultraconserved Elements (UCEs) and 2000bp flanking regions were extracted from the whooping crane reference genome using the Phyluce pipeline (Faircloth, 2016). The BCF files were subsampled to include only sites within the UCEs and their flanking regions using BCFtools (Danecek et al., 2021). UCEs in the reverse orientation were corrected using a custom awk script (Speak et al., 2024), both the forwards and orientation corrected files

were then filtered to remove Insertion and Deletions (indel) using Picard tools and subsequently converted to BED format using vcf2bed (Quinlan & Hall, 2010). CADD scores were then extracted using BEDtools intersect to produce whooping crane CADD scores for the UCEs and their 2000 bp flanking regions (Chapter 2, Figure 4).

5.9 SNPeff

The genetic load was estimated using snpEff by Claudia Fontsere of the Globe Institute, Copenhagen University. For genetic load analysis, we added three ancestral notes and one sister species (*Grus nigricollis*) into the VCF to polalirze variants (derived/ancestral). We chose the first three closest ancestral notes to the whooping crane (birdAnc314, birdAnc315, birdAnc316). We obtained the ancestral nodes fasta sequence from (Feng et al., 2020) in the hal file using hal2fasta (Armstrong et al., 2020) the *G. nigricollis* reference genome (GCA_004360235.1) (Zhou et al., 2019). Then we fragmented each fasta sequence into 150bp long sequences using bedtools windowMaker v2.30.0 (Quinlan & Hall, 2010) and subsequently map them to the whooping crane reference genome with BWA (Li and Durbin, 2009) *mem* v0.7.17. Finally, we obtained variants using bcftools *mpileup* and bcftools *call* v1.15. (Danecek et al., 2021), and we combined them to the whooping crane VCF with (Danecek et al., 2021) *merge*. We filtered the resulting VCF to remove fixed positions and keep variants with a genotyping rate of 40%. We also restricted the analysis to those variants with a 100% of genotyping rate in the four outgroups. We annotated the resulting VCF with snpEff (Cingolani et al., 2012) using the gff annotation file from the whooping crane reference genome.

From the annotated VCF we extracted those variants classified as high, moderate and low. High-impact variants are assumed to have a high (disruptive) impact on the protein, probably causing protein truncation, loss of function (LoF) or triggering nonsense-mediated decay (i.e., stop codons, splice donor variant and splice acceptor, start codon lost, etc.). Moderate impact variants are non-disruptive variants that might change protein effectiveness (i.e., missense variants), whereas low impact variants are mostly harmless or unlikely to change protein behaviour (i.e., synonymous variants). We considered only variants with a minimum read depth of five and samples with at least 10x coverage.

We separated total genetic load into heterozygous and homozygous load, and counted the number of derived alleles with low, moderate and high predicted levels for homozygous (multiplied by two as they are represented twice) and heterozygous alleles.

5.10 Genetic load calculations

The whooping crane CADD score data was filtered to remove sites that were non-scoring within all 37 individuals (sites at which the chicken reference genome and whooping crane reference genomes were the same and no alternate allele was present) and fixed sites where all 37 individuals were homozygous for the same allele. Selection (s) coefficients were calculated using a generalised associative model (GAM) function based on ranked CADD and selection coefficients simulated using SLiM (Haller & Messer, 2019) (Chapter 4). The sites were grouped into CADD scores less than 10, between 10 and 20, greater than 20 and greater than 30 and given dominance coefficients of 0.3, 0.15, 0.02 and 0 respectively.

For every locus within the UCEs and 2000bp flanking regions the individuals' genetic load, realised load and masked load were calculated (Speak et al., 2024) per the following formulas (Bertorelle et al., 2022) (see Chapter 2):

$$Genetic\ load\ (individual\ k) = \sum_{i=1}^{L(hom)} s_i + \sum_{j=1}^{L(het)} 0.5\, s_j \quad [10]$$

$$Realised\ load\ (individual\ k) = \sum_{i=1}^{L(hom)} s_i + \sum_{j=1}^{L(het)} h_j s_j \quad [11]$$

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

[12]

Here, s_i (and s_j) are the values of the selection coefficient inferred from the whooping crane CADD score at locus *i* (and *j*), and they are summed across all homozygous (or heterozygous) loci at the UCEs of individual k. The dominance coefficient (h_j) was categorised depending on the CADD scores of the mutation and was determined as follows. For mutations with a CADD score less than 10, h = 0.3; between 10 and 20, h = 0.15; between 20 and 30, h = 0.02; and over 30, h = 0.

5.11 Sample pedigree, inbreeding coefficients and relatedness.

A pedigree of 1836 individuals of the captive-bred population was produced using the R package FamAgg (Rainer et al., 2016). Inbreeding coefficients for all individuals were calculated using the function inbreeding from the ribd package (Vigeland, 2020) and the relatedness of the sampled individuals was calculated using the kinship function of the FamAgg package (Rainer et al., 2016). This calculated relatedness as the chance a locus on the genome would be identical between the two individuals, whereby in the absence of inbreeding a parent and child would have a calculated kinship of 0.25 and a selfing cross equates to a kinship of 0.5 (Rainer et al., 2016). The genetic relatedness (θ) of the samples was calculated by Claudia Fontsere Alemany of the Globe Institute, Copenhagen University using ANGSD NgsRelate v2 (Korneliussen and Moltke, 2015).

5.12 Results

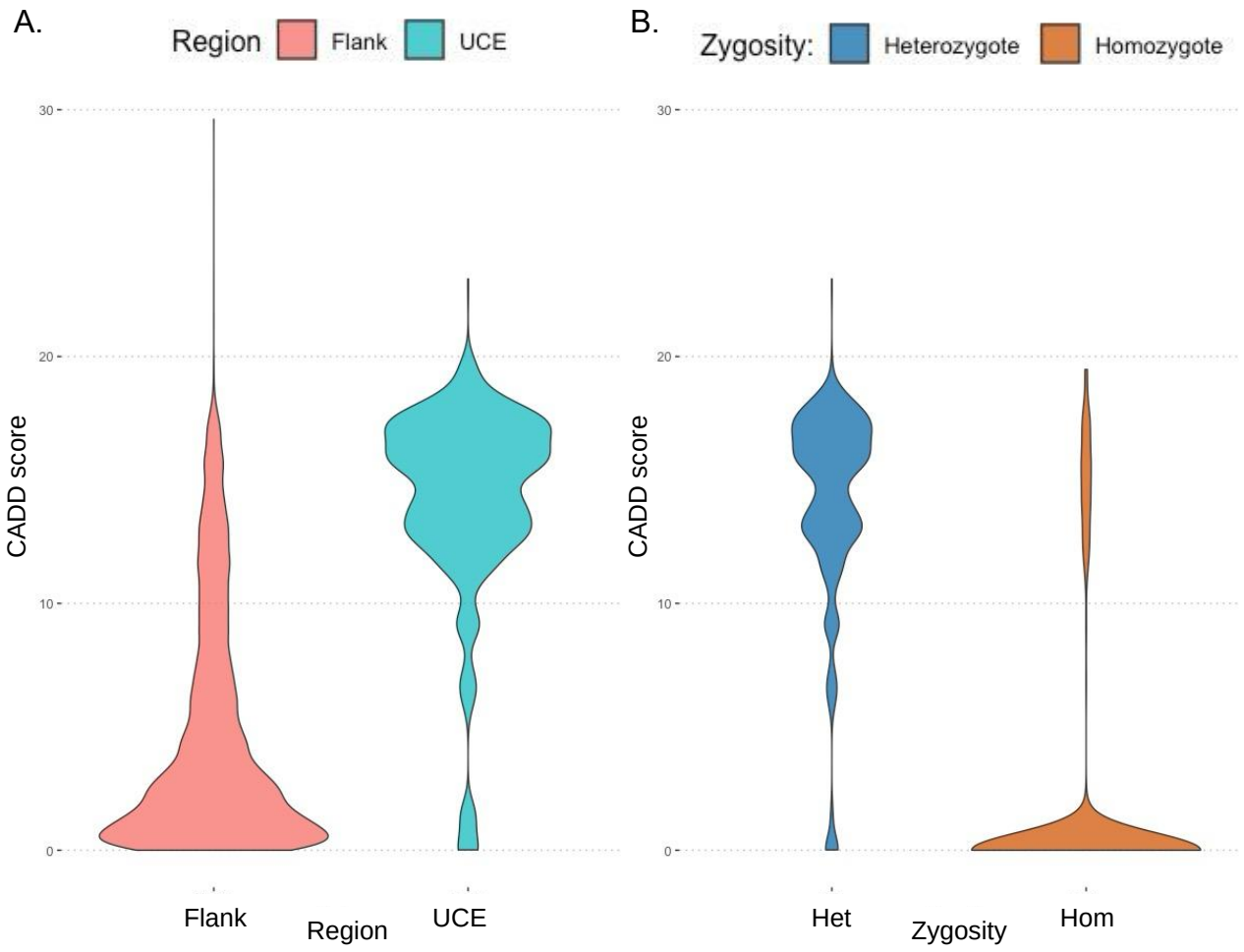
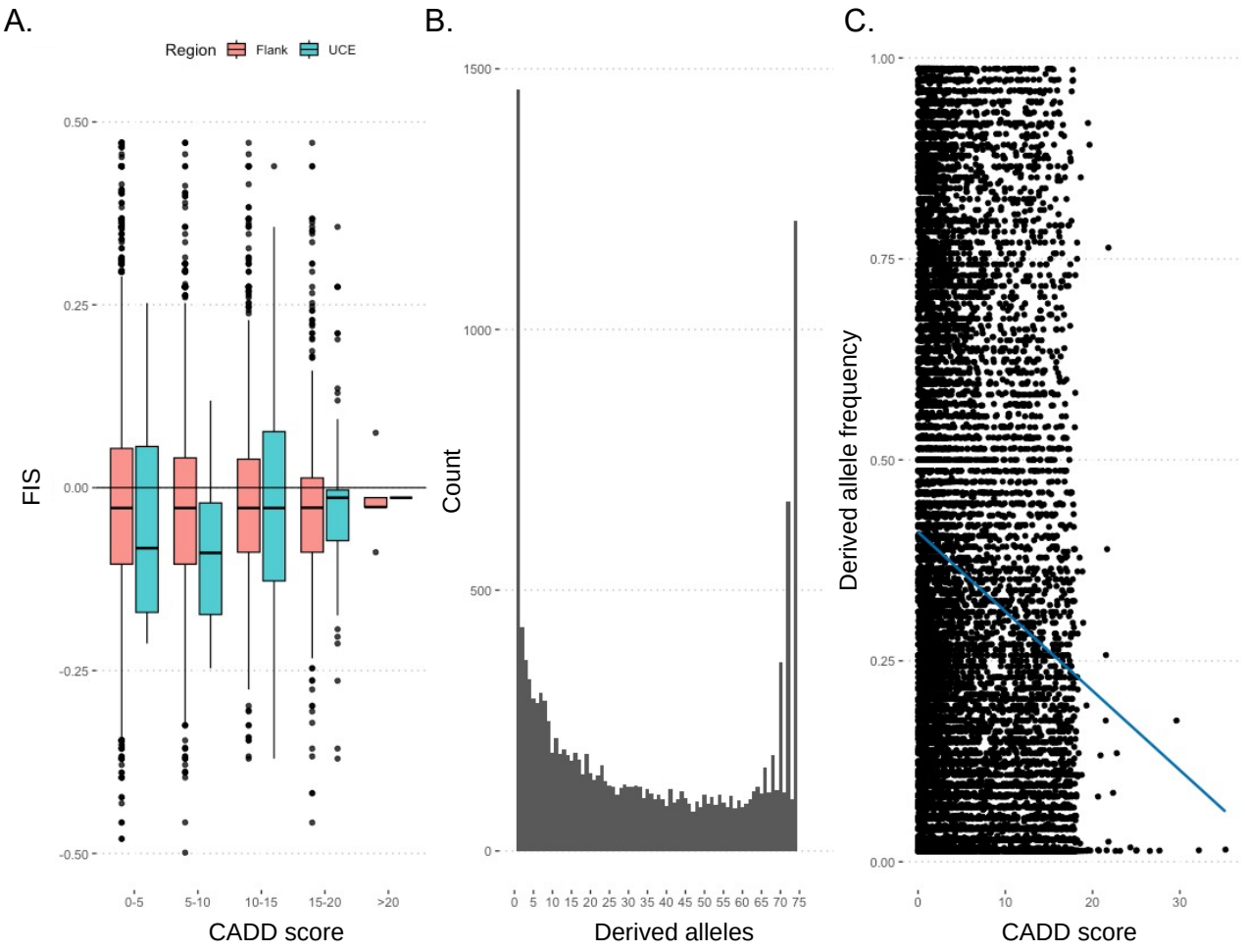


Figure 20 – The distribution of CADD scores lifted over from chicken CADD scores (chCADD scores) using the LoadLift pipeline (Speak et al., 2024) for 37 modern whooping crane samples. A) The distribution of CADD scores within the Ultraconserved Elements (UCE) (Teal) and the 2000bp up and downstream flanking regions to the UCEs (Red). **B)** The distribution of CADD scores within 144 sites within the UCE regions for sites that are Homozygous (Orange) and Heterozygous (Blue).

There is a significant difference between the CADD scores in the UCE and their 2000bp up and downstream flanking regions (Kruskal Wallis test, df = 1, $p < 2e\text{-}16$). Although the largest (most deleterious) CADD scores were present in the flanking regions, the majority of CADD scores with values between of 10-20 were found within the UCEs (Figure 20A). There was a significant difference in CADD scores of heterozygous and homozygous sites within the

UCEs (Figure 20B) (Kruskal Wallis test, df = 1, p < 2e-16), with higher CADD scores being found at predominantly heterozygous sites.



<u>Figure 21 – CADD scores lifted over to the whooping crane.</u> A) The inbreeding coefficient (F_{IS}) of sites relative to the CADD score, comparing the UCEs (Teal) and the Flanking regions (Red). Deleterious mutations were grouped as: <5; 5-10; 10-15; 15-20 and >20. **B)** A histogram showing the count of the derived allele found across the UCEs and flanking regions (data prior to the removal of sites not found in all 37 individuals and the removal of fixed sites). **C)** Correlation of the derived allele frequency and the CADD score of the deleterious allele (regression line shown in blue) (Regression analysis: R2 = 0.02715, $F_{1,11482}$ = 321.5, p < 2.2e-16), shows that mutations with high CADD scores tend to have a lower frequency.

The median F_{IS} was below zero for all CADD score categories, indicating a relative heterozygous excess (Figure 21A). This is consistent with purifying selection against deleterious mutations, but it could also be an effect of captive population management and the avoidance of inbreeding. The latter explanation is likely, given that variants with the lowest CADD scores (CADD-score < 5) also show a heterozygote excess. In fact, the mean F_{IS} was lower for the low CADD scores, but there was no significant difference in F_{IS} between CADD score range or location within UCEs or flanking regions (ANOVA: $F_{4,6178} = 0.850$, $p = 0.493$). This suggests that breeding practice caused the excess of heterozygosity rather than purifying selection.

Across the 37 samples, the number of derived alleles was summed for each site before filtering of data for the removal of fixed sites and sites with insufficient depth for all samples (Figure 21B). This is the “Count” in Figure 21B, and it tallies the number of loci with a given number of derived alleles. The most common derived allele count was for loci that were heterozygotes for one individual and homozygous for the other 36 individuals. In total, there were 1460 of such loci. Such rare variants may represent genotyping or sequencing errors, but this is unlikely because the sequencing depth of single heterozygote sites (mean depth = 12) was not much lower than the depth of other variants (mean depth = 13). Moreover, all loci analysed were filtered to a minimum depth of five reads. Hence, it seems more likely that they are deleterious mutations that are kept at a low frequency by a combination of purifying selection and good breeding practice to increase the observed heterozygosity. In addition, many of these variants may represent *de novo* mutations that arise and are lost over generations. We tested this by examining the temporal distribution of these “singleton” variants, which revealed a more-or-less uniform distribution (Supplementary Figure 7). The next highest peak was at 74 alleles (n = 1209) representing loci at which all individuals were homozygous for the derived allele. These fixed derived mutations could represent the drift load, or they may be species-specific adaptations.

The derived allele frequency was significantly correlated to the CADD score (Figure 21C) (Regression analysis: $R^2 = 0.02715$, $F_{1,11482} = 321.5$, $p < 2.2e\text{-}16$), with high derived (deleterious) allele frequency correlating with lower CADD scores. This supports the

hypothesis that they represent deleterious mutations that are kept at a low frequency by purifying selection.

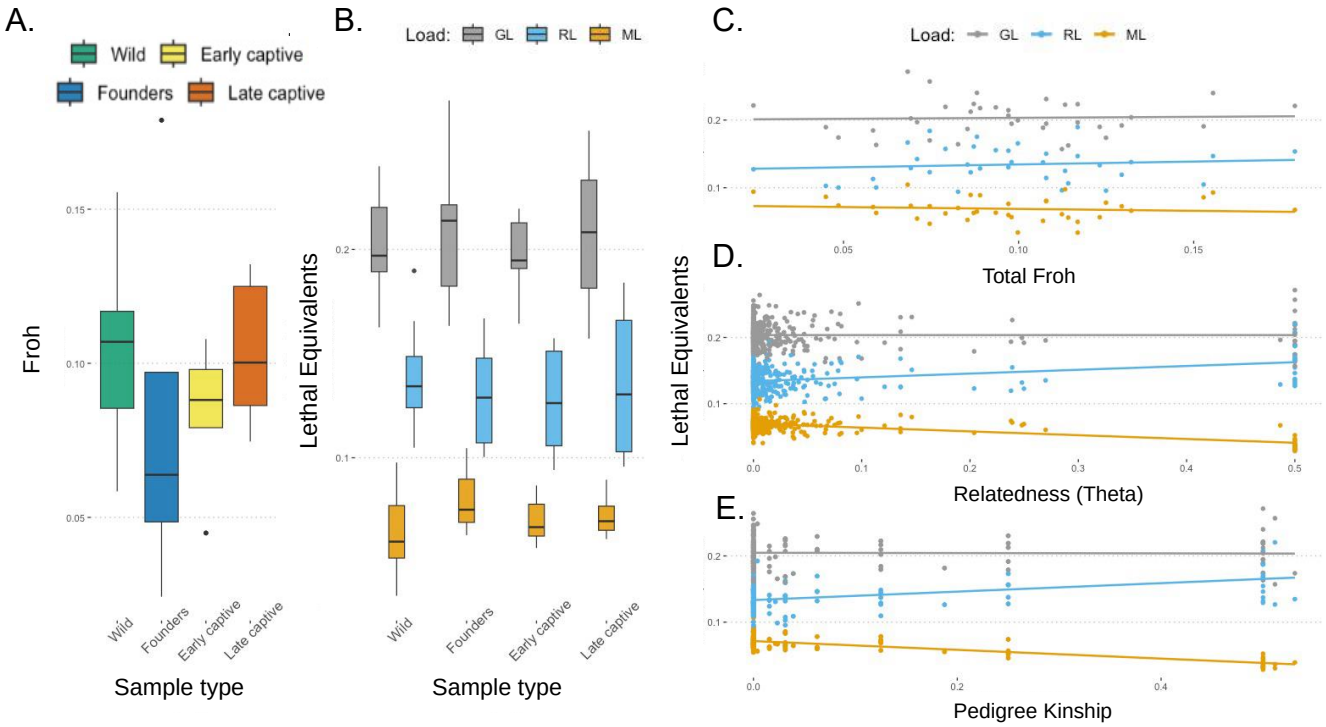


Figure 22 – The summed genetic load in lethal equivalents within the Ultra conserved Elements (UCE), for 37 whooping crane samples taken from wild individuals (Wild, n = 19), samples from founders of the captive population (Founders, n = 6), samples from early members of the captive population (Early captive, n = 6) and samples from later members of the captive population (Late captive, n = 6). A) The total F_{ROH} of the 37 whooping crane samples, Wild (Green), Founder individuals (Dark Blue), Early captive (Yellow) and Late captive (Orange). **B)** The genetic load of each individual was calculated for sites within the UCE the realised load (RL) (Orange), masked load (ML) (Blue) and the genetic load (GL) (Grey). **C)** The GL, RL and ML of each of the 37 individuals as a function of their F_{ROH}, **D)** the GL, RL and ML of potential crosses compared to the genetic relatedness of the two parents, and **E)** the GL, RL and ML of potential offspring of crosses as a function of the relatedness of their parents based on the pedigree.

There was a significant difference between the F_{ROH} and the sample type (Figure 22A) (Kruskal Wallis test: chi-squared = 15.516, df = 3, p = 0.001425), with wild individuals being significantly higher than founder (Wilcoxon rank sum, p = 0.0089) and early captive

individuals (Wilcoxon rank sum, p = 0.0212), and founder individuals being significantly fewer runs of homozygosity than cate captive members (Wilcoxon rank sum, p = 0.0212).

There was no statistically significant difference between the genetic load components between the wild, founder, early or late captive individuals (Figure 22B) (ANOVA: $F_{6,99}$ =0.471, p=0.829). The realise load was consistently higher than the masked load, which is unusual, even in recently bottlenecked species such as the pink pigeon. Such elevated RL/ML ratio is rarely (if ever) observed in non-threatened species.

Surprisingly, the load components and F_{ROH} of the individuals showed no significant correlation (Figure 22C) (Regression analysis: Genetic load, R^2 = 0.001217, p = 0.8376, Realised load, R^2 = 0.01113, p = 0.5342, Masked load, R^2 = 0.0111, p = 0.5349). Possibly, the Runs of Homozygosity (ROH) are comprised of mostly ancient ROH that have already been purged of the most highly deleterious mutations (Bosse and van Loon, 2022). In contrast, the realised and masked load in the potential offspring of the crosses was correlated to both the genetic relatedness of the parents (Figure 22D) (Regression analysis: Genetic load, R^2 = 5.854e-05, p =0.9348, Realised load, R^2 = 0.2388, p = 2.272e-08, Masked load, R^2 = 0.6277, p < 2.2e-16) and the pedigree relatedness of the parents (Figure 22E) (Regression analysis: Genetic load, R^2 = 0.003523, p = 2.272e-08 0.5872, Realised load, R^2 = 0.2559, p = 6.745e-07, Masked load, R^2 = 0.6711, p < 2.2e-16). This shows that contemporary inbreeding does increase the realised load, despite the masked load being relatively low, even though historic inbreeding (based on F_{ROH}) is not correlated to the realised load.

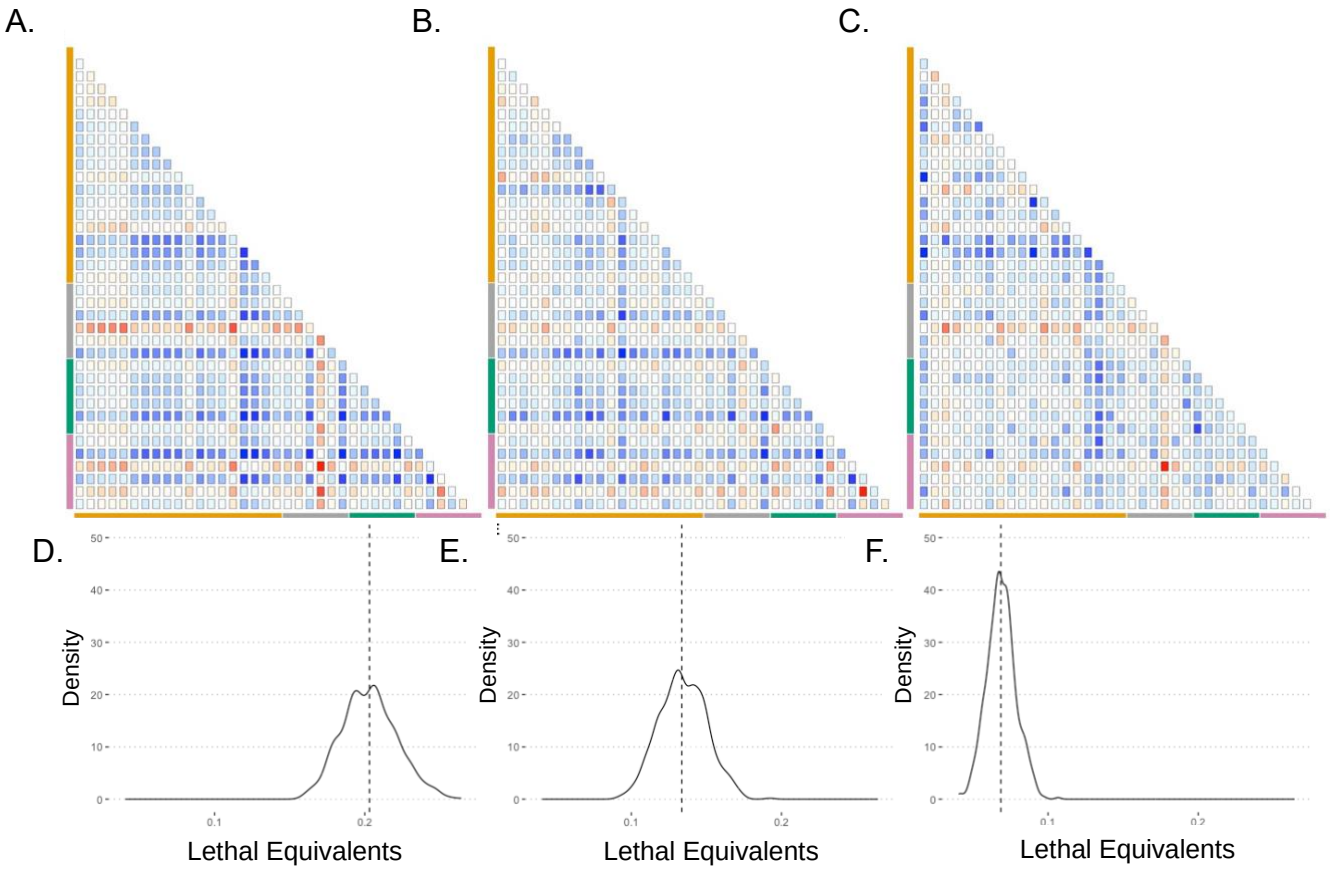


Figure 23 – Mating matrix showing all the possible potential theoretical mate combinations between whooping crane individuals for samples from wild individuals (n = 19), founder members of the captive population (n = 6), early members of the captive population (n. = 6) and late captive-bred population members (n = 6). A) The genetic load expressed in lethal equivalents (LE) for all potential crossings. **B)** The realised load in LE of all potential crossings and **C)** the masked load in LE of all potential crossings. Scaled from the highest load (Red) to the lowest load (Blue) of all potential crossings, excluding selfing crosses. All crosses were made irrespective of the sex of individuals. The categories of the parents are denoted along the x and y axis: wild individuals (Yellow), founder individuals (Grey), early captive (Green) and late captive (Pink). The distribution of summed LE for all potential offspring excluding selfing crosses for the **D)** genetic load, **E)** realised load and **F)** masked load. For all distributions, the median value for the potential offspring is shown with dashed lines.

The optimal mate pairings between individuals across all samples were made (Figure 23), allowing for hypothetical crosses between individuals within the captive and wild populations, as well as between generations. Some individuals provide potential crossings

that, relative to all other potential crosses, have low realised load (Figure 23 B) and low masked load (Figure 23C). There is a large range in the distribution of realised load between the potential offspring, indicating that the optimal mate pairing of potential crosses could effectively reduce the level of inbreeding depression in the offspring. In particular, the realised load of the best (0.091 LEs) crosses is only 47.56% of the worst crosses (0.192 LEs) (excluding self-matings).

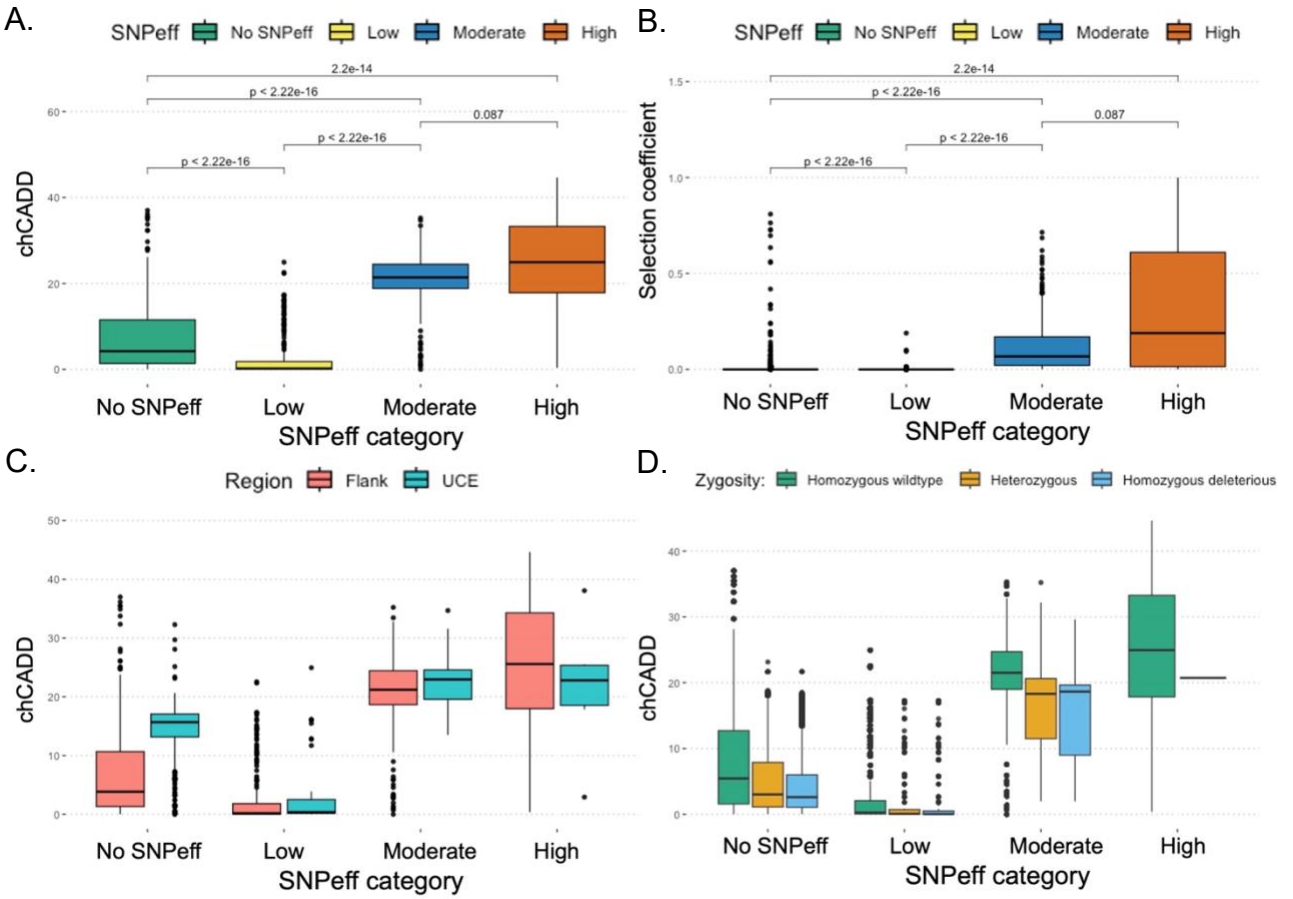


Figure 24 – A comparison of chCADD scores lifted over to the whooping crane annotation using the LoadLift pipeline and SNPeff categories across 37 whooping crane samples. A) SNPeff categories and the chCADD score of the same mutation were compared to determine whether both methods are comparable. SNPeff categories are colour-coded: low (Yellow), moderate (Blue), high (Orange) and sites with no SNPeff category (Green). **B)** The SNPeff categories and distribution of the selection coefficient simulated using SLiM (Chapter 4) of the same mutation to compare the average selection coefficient (s) across low SNPeff (Yellow), moderate (Blue), high (Orange) and sites with no SNPeff category (Green). **C)** The chCADD scores for the whooping crane samples were then compared to determine if the categories were different depending on if the site was in the 120bp UCE (Teal) or the 2000bp up and downstream flanking regions (Red). **D)** The zygosity of the 37 individuals was compared across the different SNPeff categories, with individuals homozygous for the wildtype chicken reference allele (non-scoring) (Green), heterozygous for the deleterious allele with a chCADD score (Orange) and homozygous for the deleterious allele with a chCADD score (Blue).

The SNPeff Category and chCADD score were comparable in their ranking of the severity of deleterious mutations (Figure 24A) with the average chCADD scores significantly different (Kruskall-Wallis, df = 3, p-value < 2.2e-16) between sites categorised as low (mean chCADD = 2.60, SE = 0.257), moderate (mean chCADD = 20.7, SE = 0.309), high (mean = 23.5, SE = 2.05) and sites not categorised by SNPeff (mean chCADD = 5.96, SE = 0.0252) (Supplementary Table 11). Sites categorised as low by SNPeff had significantly lower chCADD scores than sites categorised as moderate (Wilcoxon rank sum, p < 2.2e-16) and high (Wilcoxon rank sum, p < 2.2e16). Sites categorised as moderate and high by SNPeff were not significantly different in chCADD score (Wilcoxon rank sum, p < 0.087), potentially due to the small number of sites with the high SNPeff category (n = 34). There were many sites (n = 30064) that were not identified as deleterious by SNPeff. These sites do have chCADD scores (Supplementary Table 11). Moreover, SNPeff categorised a total of 623 loci across the UCEs and flanking regions as deleterious, which compares to 30687 when using CADD.

Across all categories classified as deleterious by SNPeff there was no significant difference in chCADD score between loci in the UCE or flanking regions (Figure 24C). Whilst loci that were not identified by SNPeff as deleterious were significantly different in chCADD score between UCE and flanking region (Wilcoxon rank sum, p < 2.2e-16) and followed the trend of higher CADD scores in the UCEs (Figure 20A).

Highly deleterious mutations with higher chCADD scores were rarely found in a homozygous state across the 37 samples, with loci identified as highly deleterious by SNPeff were almost all homozygous for the wild-type (non-scoring) allele (Figure 24C) (homozygous n = 33, heterozygous n = 1). Loci classified as moderate by SNPeff showed a significant difference between the zygosity and CADD score, with loci homozygous for the non-scoring allele having higher CADD scores than heterozygous sites (Wilcoxon rank sum, p < 2.2e-16) and loci homozygous for the deleterious CADD scoring allele (Wilcoxon rank sum, p < 2.2e-16). This suggests that purifying selection was reducing the frequency of highly deleterious alleles within the population.

5.13 Discussion

5.14 High CADD scores are present at low frequency

In this Chapter, I lifted over the CADD scores from the model species (chicken) to the whooping crane. I showed that CADD scores in the 144 segregating sites within the UCEs were higher on average than those within the flanking regions (Figure 20A), similar to the results found across all three model species (humans, pigs and chickens) in Chapter 3. I identified that CADD scores for mutations found as heterozygotes were higher than CADD scores for sites that were homozygotes (Figure 20B). This suggests that when high scores become homozygotes, the reduction in fitness is great enough that it leads to selection against the homozygous deleterious allele (or genotype). I found that across all sites with CADD scores within the UCE and flanking regions, the F_{IS} values were on average negative (Figure 21A), indicating that there was an excess of heterozygotes in the population. Further to this, 1460 sites across the UCE and flanking regions had only one derived allele (Figure 21B) out of all 74 alleles present within the 37 individuals, supporting the excess of heterozygotes. CADD scores correlated with the derived allele frequency (Figure 21C), suggesting that highly deleterious mutations (high CADD scores) were rare within the population. In comparison, less deleterious mutations (low CADD scores) were found in more individuals and as homozygotes more frequently. This suggested that the deleterious mutations within the genome were being selected against once they became homozygotes, resulting in the observed excess of heterozygotes in the population. These highly deleterious mutations were removed from the population by purifying selection resulting in purging (García-Dorado, 2012; Hedrick, 1994; Dussex et al., 2023). In addition, the large number of singletons (i.e., mutations occurring in one copy across all 37 individuals) would also have increased the F_{IS}, given that these mutations can only occur as heterozygotes. These singletons are likely to represent *de novo* mutations that continuously pop in and out of existence, as evidenced by their homogeneous temporal distribution.

Purifying selection has been shown to be effective at removing highly deleterious variants during population bottlenecks, as demonstrated in the case of the Alpine ibex (*Capra ibex*) (Grossen et al., 2020) and in small, isolated populations of Bengal tigers (*Panthera tigris tigris*) in India (Khan et al., 2021). However, many slightly less deleterious mutations with

less effect on fitness can increase in frequency within the population and are in many cases fixed within the population. The majority of the mutations with high CADD scores that were categorised as highly deleterious by SNPeff were homozygous for the wildtype (non-soring) allele (Figure 24C). Similarly, the average CADD score for wildtype (non-scoring) homozygotes was higher than heterozygotes or homozygotes for the scoring allele for all SNPeff categories (Figure 24C). This supports the finding that highly deleterious mutations (with high CADD scores) are very rarely found in homozygote conditions within the UCEs (Figure 20B & 21C). This could be a result of the rarity of these mutations within the population, as only 0.0006% of the loci identified with mutations present in the population had a CADD score over 30 (which is equivalent to the most deleterious 0.01% of mutations within the genome). Combined with this, the highly deleterious mutations are more likely to be removed by purifying selection, with homozygotes for these deleterious mutations potentially having lethal effects and therefore meaning that they were not able to be observed.

5.15 The impact and applications of captive breeding management

One of the main applications for the LoadLift pipeline is to be used to assess the optimal potential mate pairings within captive breeding populations. Interestingly, when I investigated the genetic load within the potential offspring from all of the 1369 crosses available with the 37 individuals that I had sequencing data for (Figure 23), the crosses between members of the original founding population with other founder individuals for five of the six individuals produced offspring that were on average lower in realised load than other crosses made in the population (Figure 23B). Further to this, I selected all of the crossing combinations between wild individuals and founder individuals to simulate the individuals who would have been available to be selected to form the captive breeding population. For the purpose of this work, I assumed the wild individual and wild-borne founders came from the same population. I found that two of the crossing combinations between the original founders resulted in a significantly lower realised load in their predicted offspring than would be expected by random mating between any combination of founder and wild individuals (Supplementary Table 12) (Supplementary Figure 6). With only one of the 30 potential mate pairings being significantly worse than random. This therefore

suggests that the original members of the founding population were good candidates to use for the founding of the captive population. The good selection of founder individuals can also be seen in the low inbreeding coefficients, calculated as the Runs of Homozygosity (F_{ROH}), of the founder individuals (Figure 22A). This shows that five of the six founder individuals were not highly inbred, and the population was genetically diverse. This is exhibited in the high number of heterozygous sites that were found in the sampled individuals, with 1460 sites across the UCE and flanking regions where only a single derived allele was present in the population (Figure 21B). This indicates that the founders of the captive population were genetically diverse and not highly inbred. This is critical for captive breeding populations as the small number of founders and often little information on their original relatedness can, especially during the early years of many captive breeding populations, make it hard to avoid inbreeding. The distinct genetic diversity present in the population will also be vital for reintroduction programmes to attempt the genetic rescue of the wild population and to provide the adaptive potential to survive future environmental or climatic changes. LoadLift is a useful tool in conservation management to assess the genetic load in candidates for captive breeding programmes, and to assess the genetic diversity relatedness of potential founders.

The crossing matrix for all of the individuals within a captive population can provide valuable breeding information for captive-breeding population managers for the whooping crane. It displays the relative impact of mate pairings on load components of their predicted offspring, and it can thereby reduce the severity of inbreeding depression in the offspring generation. Using LoadLift, I predicted the optimal and suboptimal mate pairings across all 1332 mate pairings that are hypothetically possible within the population, irrespective of the sex of the individuals and avoiding selfing scenarios. Here, I have shown that the best crosses can produce offspring with a total realised load of 0.091 LE, in comparison the worst crossings would produce a realised load of 0.192 LE. This means that screening the population can result in a 52% reduction in the realise load in offspring relative to the worst potential crosses. This approach can also be applied to provide information on which current captive breeding individuals have predicted offspring with low realised load when crossed with wild individuals. This information could be applied to select the optimal candidates for reintroduction programmes allowing for the wild and captive-bred populations to be managed as one large population fulfilling the goal of the One Plan Approach.

5.16 Evaluating the impact of inbreeding on genetic load

Although there was no significant difference between the genetic load components of the wild and captive-bred populations of the whooping cranes (Figure 22), the realised load across the wild, founders, early captive and late captive populations was higher than that of the masked load. This is indicative of the extremely high inbreeding that the population went through due to the population decline and bottleneck, as well as the small number of founders that were available to start the captive population. Due to this, a large proportion of the deleterious mutations that previously contributed to the masked load were converted to realised load. This is because they were present in the population at low frequency and therefore rarely became homozygotes. This likely happened due to drift and inbreeding during prolonged population decline, before the bottleneck and early generations in captivity. It increased in the frequency of some harmful mutations, raising the realised load within both the wild and captive populations. Similar investigations in both the Montezuma quail (*Cyrtonyx montezumae*) (Mathur and DeWoody, 2021) and the Alpine (*Capra ibex*) and Iberian (*Capra pyrenaica*) ibex (Grossen et al., 2020) which investigated the genetic load within highly inbred populations of threatened species, found elevated levels of realised load compared to less inbred populations. Both investigations conducted analyses using SNPeff to quantify genetic load.

The investigation into the Montezuma quail found that the number of sites that were homozygotes for moderate, low and no impact mutations were significantly higher in populations that were highly inbred compared to those with a lower inbreeding coefficient (Mathur and DeWoody, 2021). Further, the mutations that were classified as highly deleterious were rarely homozygous, and not significantly higher in frequency. Both the Alpine and Iberian ibex, species that have been through a severe population bottleneck, were shown to have the largest accumulation of homozygous deleterious mutations compared to closely related species (Grossen et al., 2020). Grossen et al., in their investigations, found that the species with small population sizes and/or having recently suffered large population bottlenecks had accumulated more genetic load when comparing related species (Grossen et al., 2020). Simulations of the Alpine ibex also showed a

significant increase in realised load post the population bottleneck. This investigation into ibex species also highlighted that across all of the populations of the Alpine ibex, the most highly inbred population founded by just six captive released individuals had the most homozygotic mutations (realised load) for moderate and low categorised mutations, while having a low number of highly deleterious mutations (Grossen et al., 2020). Both investigations demonstrated extensively that highly inbred species will, after a population bottleneck such as the one the whooping crane suffered, show elevated proportions of realised load. These investigations also both showed that these severe population bottlenecks can result in the purging of the most highly deleterious mutations within the population, which is conducive to the results from the analysis in the whooping crane (Figure 24D), whereby highly deleterious mutations were rare across the 37 samples (n = 33) and none of the individuals were homozygotes for these mutations.

5.17 A comparison of the methodologies LoadLift and SNPeff

I compared the scoring for deleterious mutations within the UCEs quantified by LoadLift (Speak et al., 2024) with the results of the commonly used tool SNPeff (Cingolani et al., 2012). This analysis suggested that the methodologies of LoadLift and SNPeff were comparable in the scoring of mutations and that the two methods are congruent with each having complementary strengths. CADD scores provide the ability to quantify the genetic load in terms of mutation impact scores, thus allowing for the calculation of the realised and masked load (Figures 22 & 23). However, it is recommended that CADD scores lifted over with LoadLift are only applied to the UCEs, where it can be assumed that the mutations will likely be deleterious (Speak et al., 2024). In comparison, SNPeff can be applied to the whole genome meaning that they can provide a wider assessment of the number of mutations that are predicted to be deleterious and categorise them. However, as I have shown in this Chapter, 30064 sites were lifted over by LoadLift with CADD scores for potentially deleterious mutations for sites which were not identified as deleterious by SNPeff. This is due to the way that SNPeff scores are categorised, as SNPeff is based on the reference allele for the species which can lead to biases in the results. This can occur if the deleterious mutation is present within the sample provided to make the reference genome, these variants will be assumed to be harmless irrespective of the mutation's potential impact. In

comparison, CADD scores are created using a wide range of data from model species allowing for higher confidence in the categorisation of a mutation as deleterious. Across the UCEs, SNPeff had not categorised 1290 loci (93.2%) for which there were potential CADD scores available, 247 (15.1%) of which were found to be heterozygotes and 123 (7.5%) were homozygotes of the deleterious mutation in at least one of the 37 samples. When summing the selection coefficients for these mutations this represents a possible 1.65 LE not identified as deleterious. This could lead to a high underestimation of the genetic load of the threatened population as well as biasing against individuals possessing the wildtype allele at any of these sites. Due to the potential benefits that can be gained from using both techniques, it is recommended that analyses of endangered species apply both methods to assess the quantity of genetic load present in the genomes of individuals.

As demonstrated in this chapter, CADD scores were concurrent with the SNPeff category predicted. Mutations predicted to be highly (mean CADD = 23.466) and moderately (mean CADD = 20.718) deleterious had on average higher CADD scores than mutations predicted by SNPeff to be nearly benign (low) (mean CADD = 2.604) (Figure 24A) (Supplementary Table 11). This result helps to confirm that SNPeff categories are correct in their ranking of mutations within species, as I have shown previously (Chapter 3). High CADD scores within the UCEs are shared across multiple model species indicating that these mutations are likely to be deleterious if found as homozygotes. CADD scores were able to be converted to selection coefficients (Chapter 4) and showed higher selection coefficients for mutations classified by SNPeff as highly deleterious (mean s = 0.306) or moderately deleterious (mean s = 0.115) than those classified as benign (low) (mean s = 0.002). This provides a unique possibility for future investigations as SNPeff is a popular methodology within the bioinformatics and conservation genomics community with coverage of over 320 genomes (Cingolani et al., 2012; Sukumar et al., 2021), and over 10000 citations. Therefore, there currently is a large quantity of data across a wide range of species that have been classified by SNPeff. This data could be used to calculate genome-wide assessments of the genetic load components of threatened species. This could be achieved by using the average selection coefficients for SNPeff categories from within the UCE and applying these to the loci categorised as deleterious by SNPeff. Whilst this method and approach would rely on the average selection and dominance coefficients, and therefore the results would not be precise measures of the whole genome-wide genetic load, they would be quick and easy

assessments that could retroactively be calculated for a large number of species, providing conservationists with calculations of realised and masked load to inform captive breeding and reintroduction programmes.

Conclusion

To conclude, over Chapters 2, 3 and 4 I have developed the LoadLift pipeline and demonstrated that CADD scores can be applied to threatened species and how to convert these CADD scores into selection and dominance coefficients. In this Chapter, I have applied these techniques to score another previously threatened bird species, the whooping crane, to investigate the impact of a severe genetic bottleneck on the population's genetic load. I also highlight how assessments of genetic load within threatened populations can be applied for species conservation. The whooping cranes showed no significant difference in genetic, realised or masked load between the captive-bred individuals and the wild population (Figure 22). However, the wild, founder, early captive and late captive samples possessed, all showed on average, higher realised load than masked load due to the high inbreeding that occurred following the severe population bottleneck suffered. The consequences of the population bottlenecks observed here are mirrored in other assessed species in the wider literature such as the Montezuma quail (Mathur and DeWoody, 2021) and the Alpine ibex (Grossen et al., 2020).

I was able to produce a breeding matrix identifying the optimal and suboptimal crosses for all 1369 potential pairings (Figure 23). Using these predictions, I was able to show that the original founding members were, by good fortune, good candidates to form this founding population, highlighting the strengths of the captive breeding management program. This also provides a means for identifying potential individuals that would be suited to be used in future reintroduction programmes with low realised load (Figure 23B). I have shown that CADD scores validate the categories that SNPeff predicts for mutations within the UCE (Figure 24) and that SNPeff identified a small number of mutations due to being based on the whooping crane reference genome. I conclude that CADD and SNPeff analyses can be highly complementary approaches to investigating the genetic load within threatened species, with the possibility to apply the selection coefficient retroactively quickly and easily to the large quantity of SNPeff investigations previously performed on threatened species.

Therefore, using LoadLift and SNPeff in conjunction can provide a wealth of information for conservationists and population managers. As well as to gain insights into the threats their species face, assess the risks of future inbreeding depression and provide information on the optimal captive individuals to be used for reintroduction programmes.

6. A discussion of the development of LoadLift and its future uses for conservation genomics

6.1 Abstract

Intensive efforts by conservationists have helped the demographic recovery of many species after they suffered a severe population decline or bottleneck. However, these species are still at risk of extinction from the effects of inbreeding depression and high genetic load. Therefore, genetic and genomic analyses are vital to supporting future conservation efforts, to ensure that conservation success, such as that of the pink pigeon and whooping crane, are sustainable at both a population and genomic level. With the aim to reduce the threat of imminent extinction from inbreeding depression during and following their recovery from severe population bottlenecks. Here, I summarise the development and applications of LoadLift in the pink pigeon and whooping crane. I highlight how LoadLift and conservation genomics can be applied for captive breeding and reintroduction programmes. I suggest future analyses that could be conducted to compare LoadLift and SNPeff across the whole genome, as well as how it could be applied to the study of the fitness effects of CADD scores in captive-bred and wild populations.

6.2 Discussion

6.3 Development of LoadLift

In Chapter 2 I performed a proof-of-concept study to provide a method for the assessment of the genetic load within individuals using CADD scores lifted-over from model species. I showed that using LoadLift CADD scores could be transferred to sites within the UCEs of sequenced individuals, enabling conservation genomicists to quantify the genetic load of individuals. Using the genomes of six pink pigeon individuals, I demonstrated how *in silico* crosses could be simulated to identify the optimal mate pairings, providing a tool to manage the genetic health of threatened populations for captive breeding managers.

In Chapter 3 I showed that CADD scores that have been previously calculated on model species, chickens (chCADD) (Groß, Bortoluzzi, et al., 2020), humans (hCADD) (Kircher et al., 2014; Rentzsch et al., 2019) and pigs (pCADD)(Groß, Derks, et al., 2020), are significantly and positively correlated and that they can therefore be used to assess the genetic load of species threatened with extinction. CADD scores are calculated as a rank of the score within the whole genome and therefore their absolute values cannot be shared across the species. However, I showed that CADD scores were comparable between all three species despite the large evolutionary distance between them. This is likely because ultra-conserved elements (UCEs) are, by their nature, highly conserved across species. UCEs play an important role in processes vital for survival, such as development (Visel et al., 2008; Warnefors et al., 2016), and consequently, their sequence variation evolves at a much slower pace than other genomic regions. I also highlight that, across all three species, the highest CADD scores (i.e. those most likely to be deleterious) were found within coding regions of the UCEs.

In Chapter 4, I demonstrated how CADD scores lifted over from model species could be converted into selection coefficients to calculate lethal equivalents (LE) within individuals. In this way, I showed that despite UCEs accounting only for 0.5% of the genome, this tiny region harboured an estimated load of between 1.37 and 1.88 LE. Moreover, of the six individuals that this work assessed, half of the lethal equivalents identified were realised load, and half as masked load and therefore not expressed. This also means that the pink pigeon is still vulnerable to inbreeding depression in the future.

In Chapter 5, I expanded the application of the LoadLift pipeline to the whooping crane, another bird species threatened with extinction from significant inbreeding following a population bottleneck to a minimum population of between 16 to 20 individuals in 1944 (Butler et al., 2013; Smith, 2019). This allowed for the comparison of genetic load between wild and captive individuals, showing that there was no statistically significant difference in total load between wild and captive individuals. Interestingly, the majority of this load was expressed as realised load, far higher than seen in the pink pigeon (Chapter 2), the Montezuma quail (*Cyrtonyx montezumae*) (Mathur and DeWoody, 2021), the Alpine (*Capra*

ibex) (Grossen et al., 2020) and Bengal tigers (*Panthera tigris tigris*) (Khan et al., 2021). I was also able to compare CADD scores to SNPeff within the UCEs and showed that SNPeff categories were significantly different in CADD score values for the mutations. Variants that were assessed by SNPeff as low impact have a mean(SE) selection coefficient of s = 0.002(0.00005), moderate variants an s = 0.115(0.007), and highly deleterious variants s = 0.306(0.603). However, CADD scores identified a large number of potentially deleterious sites that were not detected using SNPeff, and the total genetic load across the UCEs was underestimated by 1.65 LE by SNPeff. Therefore, this work presents a proof of concept for how conservationists can assess genetic load in threatened species globally. With real-world applications that can help to improve species survival globally.

6.4 Applications for reintroductions and genetic rescue

Across this thesis, I have shown the development and various applications of LoadLift (Speak et al., 2024) to the assessment of genetic load within species threatened with extinction. This work also highlights the value of quantifying the genetic load within these species to better understand the risks of sudden population decline that they may face, which conservationists may have been unaware of in the past. Previously, many practical assessments and studies focused only on the genetic diversity present within threatened populations (DeWoody et al., 2021; García-Dorado and Caballero, 2021; Kardos et al., 2021; White et al., 2023; Bossu et al., 2023). This is reflected in global conservation policies such as the Kunming-Montreal Global Biodiversity Framework, which states “The genetic diversity within populations of wild and domesticated species, is maintained, safeguarding their adaptive potential”(CBD, 2022). Conservationists should not overlook the importance of maintaining genetic diversity within declining or previously threatened species as it can be a critical measure of threat to species extinction (Kardos et al., 2021). However, when setting up an insurance population or considering reintroduction programmes following a genetic bottleneck, genetic diversity should not be the only metric for which species are evaluated as genetic load is also a concern (van Oosterhout, 2020). This is because, if the ancestral population of a species was historically large enough, the population may have accrued a large historical genetic load which was previously masked due to its large population size (Bertorelle et al., 2022; Femerling et al., 2023). If the population bottleneck

was both short and sharp (Bortoluzzi et al., 2020) much of this load could, by random chance, have persisted into the current population without being removed by purifying selection (Dussex et al., 2023). Therefore, with a large number of species globally having demographically recovered, or in conservation breeding programmes, there is currently a great opportunity to apply these techniques. This can be achieved by managing the populations to limit the number of deleterious alleles that are present as realised load in subsequent generations, helping to breed genetically healthy individuals. In this way, we can help to ensure the long-term survival of demographically recovered species that have recently been down-listed in the IUCN Red List (IUCN, 2024).

Much of the conservation genetics and genomics community has highlighted the benefits of using reintroduction programmes for their potential to support the successful genetic rescue of threatened populations. Despite this, to date, only a very limited number of reintroductions and reinforcement from captive-bred populations have taken place. A fear that may be limiting the widespread application of translocations from captive populations, is that after successful demographic rescue work, conservationists want to avoid causing any harm by reintroducing deleterious alleles not in the current wild populations (Frankham et al., 2011; White et al., 2023). Therefore, understanding the most appropriate populations and individuals to use for reintroduction programmes will provide a mechanism for screening translocation candidates to reduce the chances of introducing deleterious alleles. Studies differ in the ideal population to use for reintroduction programmes, whether that is to use historically small populations that have through inbreeding purged many deleterious mutations (Kyriazis et al., 2021; Robinson et al., 2019). However, this has been contested by others who argue that the benefits of maximising genetic diversity outweigh the risks of reintroducing deleterious mutations (Ralls et al., 2020). The approaches developed and piloted in my thesis will help assess the risks posed by harmful mutations in genetic rescue management programmes. Further studies should be conducted to assess the potential genetic load that has been inadvertently introduced through these translocations. Computer simulations can be performed to conduct such retrospective and counterfactual analysis, addressing the question: “What would have happened to the population without rescue?”. When the genetic load can be managed, the genetic diversity of the whole metapopulation can be maximised, increasing its adaptive potential to future changes in the environment or threats.

6.5 Whole genome application of LoadLift

In Chapter 3 I investigated how comparable the CADD scores of the three model species (humans, pigs and chickens) were, and demonstrated that the UCEs were highly comparable. UCE sites are highly important and harbour some of the most deleterious mutations in the genome with higher average CADD scores than regions flanking them. This results in a large quantity of genetic load being present in these regions, despite them only accounting for between 0.5% to 5% of the total genome (Bejerano et al., 2004). It would therefore be interesting to assess the application of CADD scores in non-UCE regions, allowing for whole genome calculations of genetic load. However, many non-UCE regions will likely contain lineage-specific adaptations that have evolved since the last common ancestor of chickens and the target species. This could lead to sites being identified as deleterious as they possess an allele not found in the chicken reference, leading to overestimations of the genetic load. To avoid this, ancestral alleles for the lineage being investigated could be determined (Feng et al., 2020), alleles that are different in the ancestor and chicken reference genome could therefore be identified and removed from the investigations. Another way that this could be achieved would be by combining the analyses of SNPeff with those of LoadLift.

The quantification of genetic load within individuals has been achieved with methods such as GERP (Davydov et al., 2010; Huber et al., 2020), PhyloP (Pollard et al., 2010) and SNPeff (Cingolani et al., 2012), which have been gaining traction in the genomics community due to their ability to be applied to non-model species. Approaches using SNPeff provide a quick and reliable method to categorise and assess the number of deleterious mutations within samples (Cingolani et al., 2012). As I have shown in this thesis investigations with CADD scores can provide additional benefits by providing impact scores of deleterious mutations, which can be converted to selection coefficients to calculate genetic load components (Chapter 4). As such information is not able to be calculated using SNPeff alone combining the application of both techniques could lead to whole genome quantifications of genetic load components in threatened species.

In Chapter 5, I demonstrated that CADD scores and SNPeff categories for mutations within the UCEs of the whooping cranes are comparable. To further ensure that lineage-specific mutations which account for 0.2% to 5.5% of avian order genomes (Feng et al., 2020) are not being wrongly determined as deleterious. To solve this, methodologies such as SNPeff (Cingolani et al., 2012), could be used to certify assumptions of deleteriousness. In this way mutations that are not categorised as deleterious by SNPeff and have the same lineage-specific adaptation as their ancestors would not be scored. Thereby, ensuring that only mutations that are likely to be deleterious are evaluated. Future investigation into genetic load within species using CADD and SNPeff across the whole genome would also enable a greater comparison of the two methodologies. In this way, we could expand the application of the LoadLift pipeline to allow for the calculation of genetic load components across the whole genome rather than being limited to the UCEs, which only account for 5% of the genome (Bejerano et al., 2004), where we can be confident that the mutations are likely to be deleterious (Silla et al., 2014). In this way, we can also validate that CADD scores and SNPeff categories are similar by investigating how they jointly score mutations across a greater number of sites, increasing the power of the analysis.

After filtering this investigation into the CADD scores within the UCEs of whooping crane individuals analysed only 144 sites (Chapter 5). This is likely due to a combination of factors. This could be as a result of a low genetic diversity that the population now suffers as a result of its small founder population. This would influence the filtering of the data, for example, due to the large amount of inbreeding, many SNPs are homozygous for all of the individuals, and these mutations were not assessed. Within the UCE, 1209 sites were homozygote for the deleterious allele (non-chicken reference) in all of the 37 individuals all of which will have been filtered out. There is, especially in the wild and later generations of the captive-bred population, a high total F_{ROH} (Chapter 5, Figure 22B), which shows that there has been recent inbreeding within both populations due to the bottleneck.

There were many sites where at least one individual had insufficient read depth to confidently call the genotypes and I retained only sites that were present within all 37 individuals. In future investigations, more sites could be retained by increasing the sequencing effort to increase the read depth and confidence in genotype calling. The

continued advancement in Next-Generation Sequencing technology (NGS) means that high-quality data is becoming more easily accessible (Levy and Boone, 2019; Satam et al., 2023). Alternatively, in analyses of historical using aDNA, samples are more degraded, and the read depth will be lower, this can make differentiation between heterozygotes and homozygotes harder (Parks and Lambert, 2015; Dalal et al., 2023). Therefore, genotypes could be called using genotype likelihoods, using ANGSD (Korneliussen et al., 2014). For potential future investigations especially in those with many historical samples, it may be pertinent to set a lower threshold for the number of individuals that have data for each site, for example, a 95% threshold and that all individuals needed to have data for 95% of sites analysed. This would mean that the investigation would be able to retain more sites to improve confidence in the results whilst also ensuring that a few lower-quality samples did not hinder the analysis.

It has been suggested previously that although CADD scores do range from 0 to 99 and can allow for the comparison of how deleterious SNPs are, there is no official guideline as to a value for a CADD score of a mutation for it to be considered deleterious or benign (Kircher et al., 2014). Despite this, investigations have attempted to provide their own arbitrary cut-offs for what can be considered deleterious, in an attempt to identify potentially deleterious mutations within the genome. An investigation to attempt to identify mutations linked to Amyotrophic lateral sclerosis (ALS) reported that filtering for CADD scores greater than 20 resulted in too many potential sites and that limiting the investigation to CADD scores of greater than 30 provided too few (Fiorini et al., 2023). Through my comparison of CADD scores and SNPeff categories, I found that although those identified by SNPeff as highly and moderately deleterious did have significantly higher CADD scores than those identified as low (Figure 24A), for all of the categories in the range of CADD scores did overlap. Therefore, investigations conducted using CADD scores should not apply an arbitrary cut-off for the selection of SNPs to investigate as this could be misleading. In this way, the categorisation of alleles by SNPeff can be used in conjunction with CADD scores to estimate the genetic load of populations. SNPeff assesses the genetic load across the entire genome (not just UCEs), but it does not categorise a large proportion (93.2%) of the variants. Still, if the SNPeff categories can be converted into s coefficients (see Chapter 4), this would be a big advantage, especially if the proportion of unassessed SNPeff variants is known and accounted for.

6.6 Evaluating the future applications of LoadLift

I have shown that CADD scores can be converted to selection and dominance coefficients and then used for the calculation of load components (Chapter 4). These can then be used to predict the optimal mate pairings of individuals (Chapters 2, 4 & 5), allowing for the assessment of crosses between wild and captive individuals simulating future mating matrices that could be used by conservation managers to select candidates for reintroduction programmes. Whilst the primary aim of LoadLift is to allow for captive-bred populations to be employed for reintroductions to provide genetic diversity not found in the wild populations without introducing deleterious alleles, other applications of LoadLift could be used for the management of threatened species.

Firstly, in captive breeding populations, where crossings can be more actively controlled, once mate pairings have been made samples could be taken from the offspring generations and LoadLift could be used to assess the load present in these individuals to compare to the predicted values of the crossing. This would allow conservation genomicists to identify if there are any alleles with high CADD scores that are not found as homozygotes in the offspring. This would indicate that these alleles are being removed by selection before hatching, this information could then be used in future mate choice selections. An analysis of this kind would also provide a wealth of information on how high CADD scores relate to phenotypic effects.

Secondly, for populations where reintroduction or population reinforcement is being planned LoadLift can be used to select the optimal individuals for the reintroduction programme that would increase the genetic diversity whilst not increasing the realised load of the population. Once crosses have been made between the wild and captive populations the offspring could be sampled and compared to those of crossings in captive populations. This could highlight how effective the genetic rescue is as well as monitor the levels of realised load within the population. In this way, with long-term monitoring conservationists would be able to have a continued assessment of the threats from genetic load to the population as well as identify the optimal individuals to take place in the next round of introductions. In an ideal situation

where the wild and captive populations have all been sequenced and analysed and mating between the wild and reintroduced individuals are not controlled, a comparison between the optimal theoretical crosses for the lowest genetic load and the observed crosses could be made. This will help to investigate if any potential behaviours are acting to limit inbreeding or genetic load naturally within the wild populations. Monitoring the offspring from these releases would also ensure that no single individual is being overrepresented in the wild populations, this would help conservationists to avoid future inbreeding depression such as that seen in the Isle Royal wolves (Robinson et al., 2019). In this way, conservationists would be able to protect against any potential harm as a result of the reintroduction of individuals and alleles.

Thirdly, sampling historical samples of threatened species such as museum samples would allow for a comparison of the species' genetic load over a longer period of time. Using samples that were preserved prior to the species declines would also allow for investigations to quantify the genetic load present in the populations before the bottleneck. It has been shown that ancestral population size (Femerling et al., 2023) as well as the quantity of load that the ancestral population harboured prior to a bottleneck influences the load present in the post-bottleneck populations (Dussex et al., 2023). Therefore, using historical samples of currently endangered species may allow for a comparison of how the load present in the current populations has changed over time. In this way, it will also provide greater context surrounding the historical population to be used in simulations, such as those used for predicting the future population (Chapter 2, Figure 8) or to simulate selection coefficients (Chapter 4).

LoadLift also has great potential to investigate the load present within historical populations that have gone extinct, such as the passenger pigeon (*Ectopistes migratorius*). The passenger pigeon population was thought to have numbered between 3 to 5 billion individuals (Murray et al., 2017) and were once present in such great numbers across North America that the sky would blacken as great flocks flew over (Hung et al., 2014). The passenger pigeon population collapsed suddenly and catastrophically in the early 1900s, leading to their eventual extinction in 1914 (Herman William C, 1948). Whilst this is commonly attributed to overhunting by humans, previous studies into the passenger pigeon

suggest that it had suffered frequent and dramatic population declines (Hung et al., 2014) and possessed a high genetic diversity prior to its extinction (Murray et al., 2017). Due to the high genetic diversity in a population of billions, they likely possessed a high genetic load that may have led to such a rapid decline. In a large population, a high masked load can be accumulated due to the low frequency at which they are found as homozygotes. This means that selection could not act to remove them from the population. When the population began to decline this masked load would have become realised and expressed, leading to the eventual extinction of the species. This, paired with conspicuous behavioural adaptations that made hunting easier (Hung et al., 2014), could have resulted in the catastrophic collapse the population suffered. Tools like LoadLift could be applied to species like the passenger pigeon to help establish why they became extinct and to help protect current species from suffering the same fate.

Finally, many zoos globally not only possess living individuals of endangered species, but it is also common that when individuals die in captivity some biological material is preserved in biobanks such as the EAZA biobank (EAZA, 2024). This biobank material has the potential to be extremely valuable for conservation efforts (Bolton et al., 2022; Breithoff & Harrison, 2020; Howell et al., 2021). This material is often frozen very soon after the animal has died and is often preserved in either ethanol or a buffer such as RNA-later. Therefore, this material will be less degraded than samples taken from museum specimens which often can face issues due to the preservatives used in taxidermy, as well as the age of the samples and risk of contamination (Singh & Bahuguna, 2014; Tsai et al., 2020; Zimmermann et al., 2008). Therefore, samples preserved in a biobank will provide a means of comparing genetic load across long periods within a captive-bred population. These samples if combined with pedigree information can also provide an opportunity to use gene drop analysis (Honda et al., 2002) to track how selection acts on deleterious traits and to help infer selection and dominance coefficients within captivity. A recent application for biobank material is to use it to perform reintroductions of individuals and lost genetic variation (Bolton et al., 2022; Howell et al., 2021; Novak et al., 2024). In the black-footed ferret, stomatic cell lines from two individuals were cryopreserved, which allowed for the individuals to be cloned (Novak et al., 2024). The individuals cloned had no living descendants. Therefore, the aim is to integrate the clones into the breeding programme to act as a source of “new” genetic diversity to the population (Novak et al., 2024). This approach is groundbreaking and would act as a means

of reducing the effects of population bottlenecks by breeding the current population with the pre-bottleneck population providing geneflow across generations. However, reintroductions of individuals in this manner could introduce deleterious mutations that were removed from the population during the bottleneck through purifying selection (Dussex et al., 2023; Grossen et al., 2020; Hedrick, 1994; van der Valk et al., 2021). While this method therefore has great advantages, assessments should also be carried out to determine the genetic load within the biobank samples to determine the best samples to use for future reintroduction programmes. This is especially important as the costs related to this method are high (around USD 40,000) (Fritts, 2022) and therefore the number of individuals that this can be performed on is low. This means selecting the optimal individuals prior to cloning would reduce the risks of wasting both potential samples and funds that could be applied elsewhere.

6.7 Conclusion

To conclude, 45,300 species globally are currently threatened with extinction (IUCN, 2024), whilst many more species have managed to recover demographically, due to a concerted effort by conservationists globally. The pink pigeon and whooping crane are two such examples, declining to populations of 12 and 21 respectively. Both populations have since recovered due to effective population management, including captive breeding populations designed to act as insurance populations against extinction. However, due to the large ancestral population sizes and subsequent significant population bottleneck converting masked load to realised load, these species are now, at high risk of extinction in the future, despite the population's apparent recovery. Both species have benefited from intense conservation efforts including captive breeding populations, from which reintroductions can be performed. While many in the conservation community call for genetic rescue attempts (Frankham, 2015), the number of genetic attempts from captive populations is limited. One of the reasons for this is due to concerns about reintroducing deleterious alleles that are not found in the wild populations.

With the growing availability of whole genome sequenced data and the advances in bioinformatics and genomic techniques, tools can be applied to assess the number of

deleterious alleles present within sequenced data. I developed a pipeline to apply CADD scores calculated in model species to assess the genetic load of individuals as well as to identify optimal mate pairings (Chapter 2). I investigated the CADD scores across the UCEs of three model species and showed that they are comparable and that they are therefore appropriate to be lifted over to assess threatened species (Chapter 3). I converted the CADD scores into selection and dominance coefficients to allow for calculations of the genetic load within individuals that can be summed to produce values in the form of lethal equivalents (Chapter 4). I applied all of this to the whooping crane showing that the population possessed more realised load than masked load. As well as validating that SNPeff categories were comparable to CADD scores but that SNPeff only identified half of the sites as deleterious as those with available CADD scores (Chapter 5).

LoadLift has the potential to provide conservationists with calculations of genetic load components within individuals in both captive breeding and wild populations (Speak et al., 2024). This tool can be applied by captive-breeding population managers to assess the optimal mate pairings within populations, ensuring the long-term survival of the captive population. This approach can then be applied to assess the optimal individuals to breed for reintroduction programmes to select the healthiest individuals to be released into the wild. As well as to screen the individuals to select those that would be able to introduce genetic variation that will improve the species' adaptive potential without introducing harmful genetic variation. Investigations into the historical load in extinct species or the ancestral species of currently threatened populations would help to understand how genetic load can impact species recovery and survival. Zoo specimens and biobank samples can provide vital information on the founder members of these critical insurance populations to ensure they are managed sustainably. These assessments may also enable future conservation applications such as inseminations with biobank samples or cloning of individuals for reintroduction programmes to be screened to ensure that the optimal material is used in these expensive processes (Al Hikmani et al., 2024; Fritts, 2022; Jackson et al., 2022; Lott et al., 2020). LoadLift has wide-ranging applications to help maximise the potential of captive breeding populations for endangered species conservation globally. This will ensure that populations avoid inbreeding depression and increase the recovery potential of threatened species.

Appendix

7. Development and application of LoadLift for genomics-informed captive breeding

7.1 Supplementary Tables

Supplementary Table 1 – The Inbreeding coefficient (F_{ped}), the molecular estimate of inbreeding (F_{ROH}) and the kinship coefficient between the six pink pigeon individuals of the study. The inbreeding coefficient was calculated using purgeR and the full captive breeding pedigree. The F_{ROH} is calculated across the whole genome of each individual using Equation 4. The kinship coefficient is calculated as the probability that a randomly selected allele from a locus will be identical by descent. Therefore, in the absence of prior inbreeding an individual crossed with itself would generate a kinship coefficient of 0.5, and a parent and child would have a kinship coefficient of 0.25.

			Kinship coefficient					
Individual	Fped	FROH	1	2	3	4	5	6
1	0.07	0.183	0.534	0.075	0.132	0.12	0.092	0.092
2	0.348	0.232	0.075	0.673	0.269	0.048	0.042	0.042
3	0.093	0.096	0.132	0.269	0.544	0.114	0.081	0.081
4	0.064	0.106	0.12	0.048	0.114	0.532	0.3	0.3
5	0.069	0.118	0.092	0.042	0.081	0.3	0.534	0.306
6	0.069	0.133	0.092	0.042	0.081	0.3	0.306	0.534

Supplementary Table 2 – The number of UCEs found on each of the 34 Chromosomes of the galgal6 chicken reference genome. Observed vs expected number of UCEs relative to the length of each chromosome. Significant depletion (Blue) and enrichment (Green) were determined by a binomial test ($p<0.05$).

Chromosome	Nucleotides	UCEs	Expected.UCE	UCE.depleted	UCE.enriched
1	197608386	620	917.721695	9.760000e-26	1.000000e+00
2	149682049	804	695.144910	9.999746e-01	2.540000e-05
3	110838418	561	514.749515	9.792127e-01	2.078735e-02
4	91315245	234	424.081099	4.340000e-24	1.000000e+00
5	59809098	453	277.762032	1.000000e+00	0.000000e+00
6	36374701	330	168.929330	1.000000e+00	0.000000e+00
7	36742308	288	170.636550	1.000000e+00	0.000000e+00
8	30219446	236	140.343443	1.000000e+00	6.970000e-14
9	24153086	110	112.170397	4.434139e-01	5.565861e-01
10	21119840	61	98.083567	3.890000e-05	9.999611e-01
11	20200042	220	93.811893	1.000000e+00	0.000000e+00
12	20387278	117	94.681444	9.885699e-01	1.143005e-02
13	19166714	92	89.012970	6.498065e-01	3.501935e-01
14	16219308	31	75.324793	6.040000e-09	1.000000e+00
15	13062184	31	60.662657	2.030000e-05	9.999797e-01
16	2844601	0	13.210735	1.830000e-06	9.999982e-01
17	10762512	44	49.982650	2.217560e-01	7.782440e-01
18	11373140	41	52.818494	5.538405e-02	9.446159e-01
19	10323212	38	47.942478	8.262848e-02	9.173715e-01
20	13897287	82	64.540995	9.846963e-01	1.530374e-02
21	6844979	27	31.789065	2.270891e-01	7.729109e-01
22	5459462	21	25.354525	2.260108e-01	7.739892e-01
23	6149580	18	28.559532	2.387451e-02	9.761255e-01
24	6491222	28	30.146166	3.928703e-01	6.071298e-01
25	3980610	0	18.486524	9.360000e-09	1.000000e+00
26	6055710	22	28.123586	1.431376e-01	8.568624e-01
27	8080432	4	37.526685	4.650000e-12	1.000000e+00
28	5116882	3	23.763534	1.220000e-07	9.999999e-01
30	1818525	0	8.445491	2.148630e-04	9.997851e-01
31	6153034	0	28.575573	3.890000e-13	1.000000e+00
32	725831	0	3.370863	3.435969e-02	9.656403e-01
33	7821666	5	36.324939	1.020000e-10	1.000000e+00
W	6813114	1	31.641079	5.920000e-13	1.000000e+00
Z	82529921	355	383.280794	7.663518e-02	9.233648e-01

Supplementary Table 3 – The genetic load of potential offspring from mate parings between all combinations of six individuals of the captive pink pigeon population. The sites within the ultraconserved elements of the genome were scored using CADD scores lifted from the chicken genome using the LoadLift pipeline. The genetic load of the potential offspring was calculated using equation [1]. All potential mate pairings were used irrespective of sex and including self-mating for illustrative purposes.

Parents	1	2	3	4	5	6
1	5878	5843	5814	5957	5887	5892
2	5843	5809	5779	5922	5852	5857
3	5814	5779	5749	5892	5822	5827
4	5957	5922	5892	6036	5966	5971
5	5887	5852	5822	5966	5895	5900
6	5892	5857	5827	5971	5900	5905

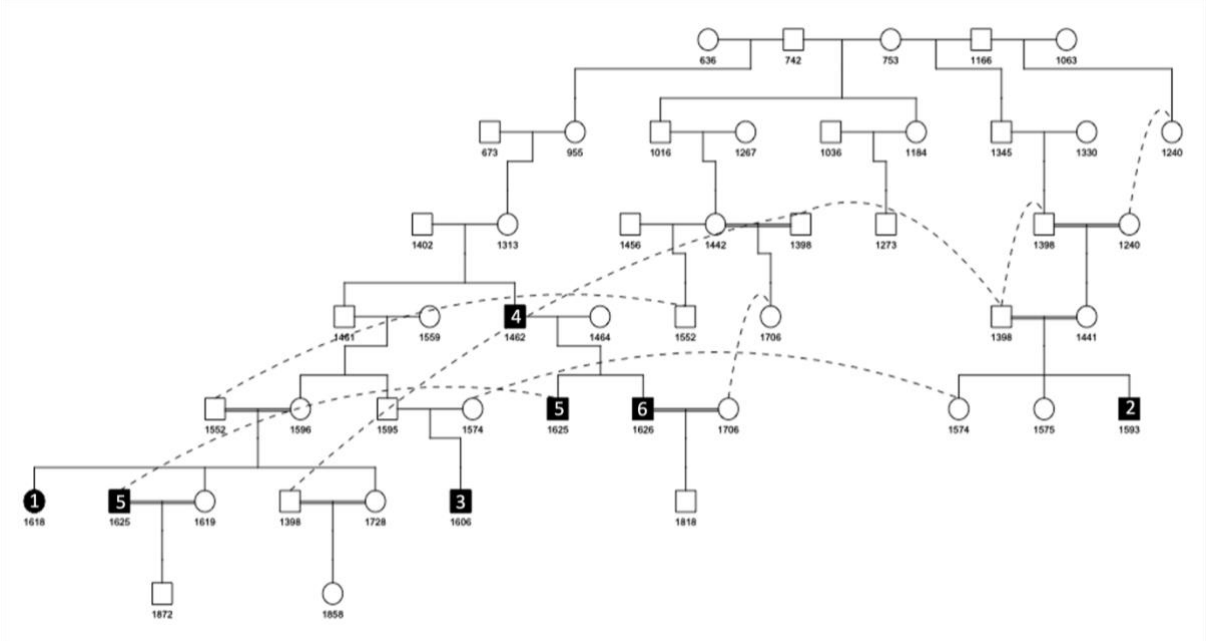
Supplementary Table 4 – The realised load of potential offspring from mate parings between all combinations of six individuals of the captive pink pigeon population. The sites within the ultraconserved elements of the genome were scored using CADD scores lifted from the chicken genome using the LoadLift pipeline. The realised load of the potential offspring was calculated using equation [2]. All potential mate pairings were used irrespective of sex and including self-mating for illustrative purposes.

Parents	1	2	3	4	5	6
1	4284	2600	2596	2838	2859	2798
2	2600	4494	3355	2602	2526	2425
3	2596	3355	4288	2628	2758	2581
4	2838	2602	2628	4434	3385	3514
5	2859	2526	2758	3385	4405	3439
6	2798	2425	2581	3514	3439	4319

Supplementary Table 5 – The masked load of potential offspring from mate parings between all combinations of six individuals of the captive pink pigeon population. The sites within the ultraconserved elements of the genome were scored using CADD scores lifted from the chicken genome using the LoadLift pipeline. The masked load of the potential offspring was calculated using equation [3]. All potential mate pairings were used irrespective of sex and including self-mating for illustrative purposes.

Parents	1	2	3	4	5	6
1	594	3243	3217	3119	3028	3094
2	3243	1315	2424	3321	3325	3432
3	3217	2424	1460	3264	3064	3246
4	3119	3321	3264	1602	2581	2457
5	3028	3325	3064	2581	1490	2462
6	3094	3432	3246	2457	2462	1586

7.2 Supplementary Figures



Supplementary Figure 1 – A partial pedigree of the captive bred pink pigeon population. Pedigree trimmed to focus on the six pink pigeon individuals for the study (black), males (square) and females (circles).

8. Investigating the similarity of CADD scores in three model species

8.1 Supplementary Tables

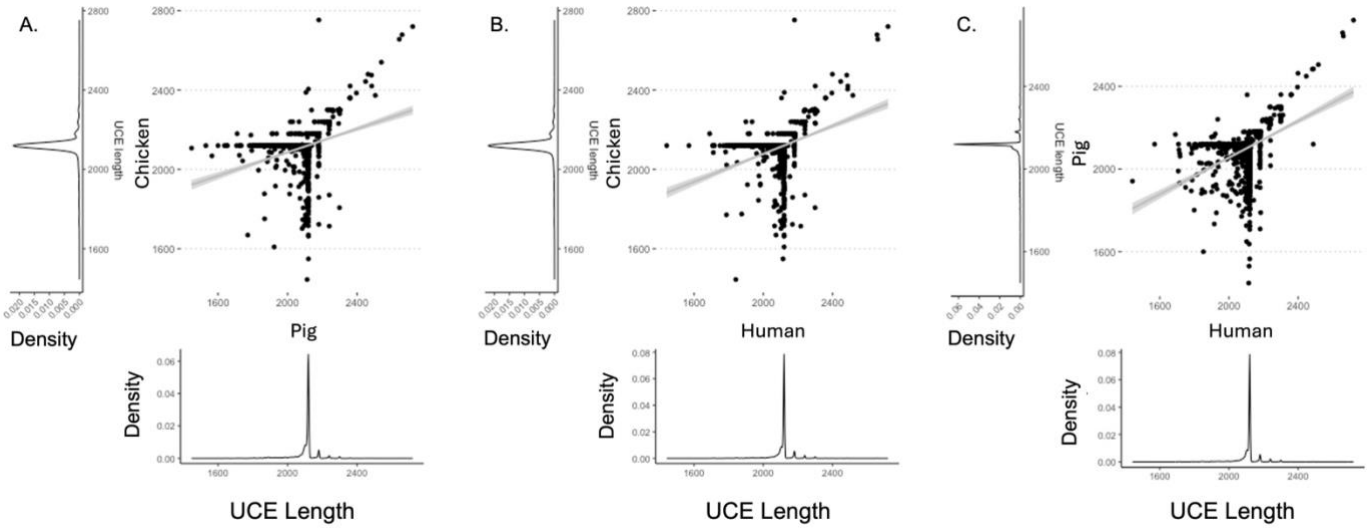
Supplementary Table 6 – A table to show of the number of CADD scores in the UCE and flanking regions of each of the three model species: humans, chicken and pig.

Region	Species	Total sites
Flank	Chicken	13101519
Flank	Human	12854136
Flank	Pig	13017477
UCE	Chicken	873525
UCE	Human	855171
UCE	Pig	868077

Supplementary Table 7 – A table to show of the number of CADD scores which fall within regions annotated as within exon or Intergenic regions within the UCEs and flanking regions of each of the three model species: humans, chicken and pig.

Region	Annotation	Species	Total sites
Flank	Exon	Chicken	1883940
Flank	Exon	Human	2219631
Flank	Exon	Pig	1872192
Flank	Intergenic	Chicken	11217579
Flank	Intergenic	Human	10634505
Flank	Intergenic	Pig	11145285
UCE	Exon	Chicken	119670
UCE	Exon	Human	144876
UCE	Exon	Pig	119943
UCE	Intergenic	Chicken	753855
UCE	Intergenic	Human	710295
UCE	Intergenic	Pig	748134

8.2 Supplementary Figures



Supplementary Figure 2 – Pairwise comparisons of UCE length between three model species A) Correlations comparing pig UCE length to chicken UCE length, **B)** chicken UCE length to human UCE length and **C)** pig UCE length to human UCE length. Distributions of the UCE lengths are supplied on for each species on the x and y axis. The total length of filtered sequences including UCE and flanking regions were then compared in a pairwise manner across the three species showing that there was a significant correlation between the pig and chicken ($R2 = 0.1031$, $p < 2.2e-16$), human and chicken ($R2 = 0.1219$, $p < 2.2e-16$) and pig and human ($R2 = 0.1659$, $p < 2.2e-16$).

9. Conversion of CADD scores to selection coefficients for load calculations

9.1 Supplementary Tables

Supplementary Table 8 – The load components as summed simulated selection coefficients for neutral and harmful mutations for all five DFEs. For all six pink pigeon individuals, the genetic load was calculated using Equation 2, the realised load was calculated using Equation 3 and the masked load was calculated using Equation 4.

ID	DFE1_GL_sum	DFE1_RL_sum	DFE1_ML_sum	DFE2_GL_sum	DFE2_RL_sum	DFE2_ML_sum	DFE3_GL_sum	DFE3_RL_sum	DFE3_ML_sum	DFE4_GL_sum	DFE4_RL_sum	DFE4_ML_sum	DFE6_GL_sum	DFE6_RL_sum	DFE6_ML_sum
PP1	0.3226430	0.1455227	0.1771203	0.2649790	0.1206849	0.1442941	0.1650950	0.06054832	0.1045467	0.9056750	0.3310956	0.5745794	1.366028	0.5318949	0.8341331
PP2	0.5473545	0.1412188	0.4061357	0.4153770	0.1168371	0.2985399	0.6741465	0.04438365	0.6297628	0.9707765	0.4020309	0.5687456	1.704835	0.5993721	1.1054624
PP3	0.4726620	0.1541279	0.3185341	0.3623090	0.1290926	0.2332164	0.4853880	0.05940113	0.4259869	0.9896610	0.3969613	0.5926997	1.648044	0.6118640	1.0361805
PP4	0.3808125	0.1481297	0.2326828	0.2993765	0.1226916	0.1766849	0.3512535	0.04363003	0.3076235	0.9470205	0.3433282	0.6036923	1.495888	0.5260311	0.9698574
PP5	0.3851740	0.1491782	0.2359958	0.3067390	0.1263329	0.1804061	0.2814770	0.04770134	0.2337757	0.9256025	0.3509413	0.5746612	1.458395	0.5413632	0.9170323
PP6	0.6680245	0.1519741	0.5160504	0.5171355	0.1284189	0.3887166	0.7296545	0.06998383	0.6596707	1.0227600	0.3515065	0.6712535	1.884396	0.5631018	1.3212942

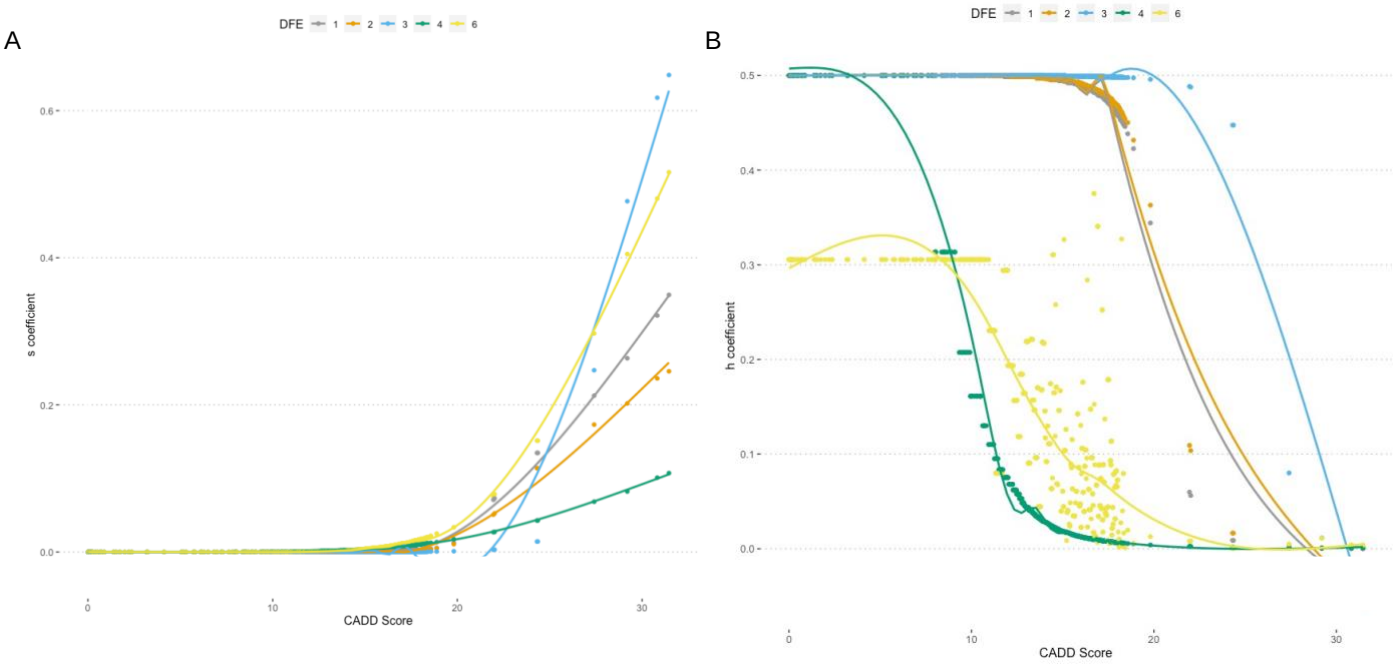
Supplementary Table 9 – The load components as summed simulated selection coefficients for only harmful mutations for all five DFEs. For all six pink pigeon individuals, the genetic load was calculated using Equation 2, the realised load was calculated using Equation 3 and the masked load was calculated using Equation 4.

ID	DFE1_GL_sum	DFE1_RL_sum	DFE1_ML_sum	DFE2_GL_sum	DFE2_RL_sum	DFE2_ML_sum	DFE3_GL_sum	DFE3_RL_sum	DFE3_ML_sum	DFE4_GL_sum	DFE4_RL_sum	DFE4_ML_sum	DFE6_GL_sum	DFE6_RL_sum	DFE6_ML_sum
PP1	28.73452	10.80866	17.92586	24.09939	9.302312	14.79708	11.49553	5.987410	5.508118	2.944571	1.061134	1.883437	36.84518	13.76020	23.08498
PP2	29.00417	14.27760	14.72656	24.32988	12.280004	12.04988	11.67459	6.245408	5.429182	3.070130	1.415808	1.654322	37.72176	18.26753	19.45423
PP3	29.24370	12.65441	16.58929	24.48664	10.825218	13.66142	12.29082	6.493691	5.797129	3.097059	1.297802	1.799257	38.33841	16.69925	21.63916
PP4	29.33550	11.99381	17.34169	24.57382	10.352853	14.22097	11.90705	5.923300	5.983753	3.048611	1.174899	1.873712	37.97864	15.28025	22.69839
PP5	28.68270	12.07587	16.60683	24.00163	10.381953	13.61968	11.71685	5.869129	5.847718	2.971729	1.182708	1.789021	37.02636	15.33615	21.69021
PP6	30.06202	11.94143	18.12059	25.03252	10.255073	14.77745	12.98146	6.428008	6.553452	3.201947	1.186074	2.015873	39.08107	15.35889	23.72217

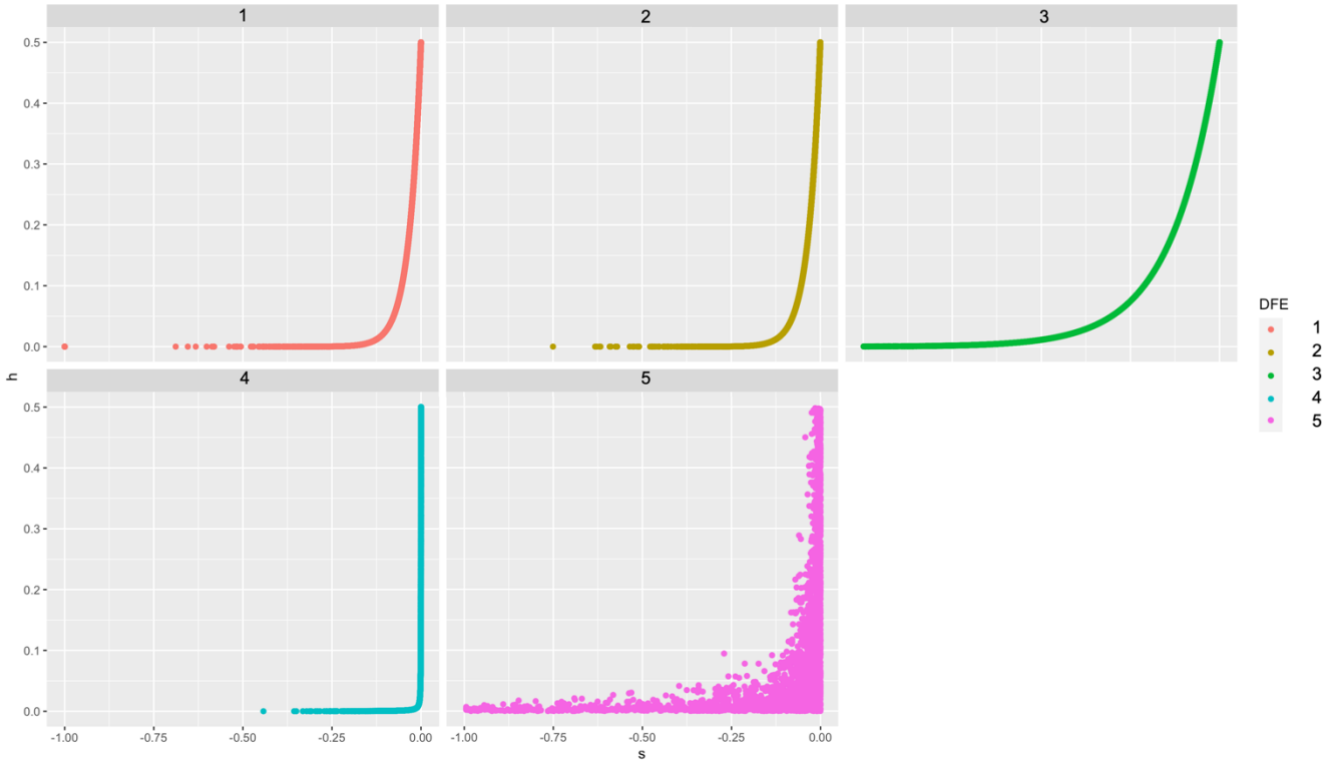
Supplementary Table 10 – The Inbreeding coefficient (F_{ped}) and the kinship coefficient for selfing crosses for each of the six pink pigeon individuals of the study, the increase in inbreeding coefficient and the predicted ratio of realised load. The inbreeding coefficient was calculated using purgeR and the full captive breeding pedigree. The kinship coefficient is calculated as the probability that a randomly selected allele from a locus will be identical by descent. Therefore, in the absence of prior inbreeding an individual crossed with itself would generate a kinship coefficient of 0.5, and a parent and child would have a kinship coefficient of 0.25. However, due to the inbreeding in the pink pigeon the increase in inbreeding coefficient is not 0.5.

Individual	Fped	Selfing	Increase	Expected ratio
1	0.07	0.534	0.464	1.86
2	0.348	0.673	0.325	1.3
3	0.093	0.544	0.451	1.8
4	0.064	0.532	0.468	1.87
5	0.069	0.534	0.465	1.86
6	0.069	0.534	0.465	1.86

9.2 Supplementary Figures



Supplementary Figure 3 – A comparison of all five DFEs simulated in SLiM (Haller & Messer, 2019) and the ranked CADD score. A) The simulated selection coefficients and **B)** the simulated dominance coefficients.



Supplementary Figure 4 – The distributions of selection and dominance coefficients for the five DFEs. Figure produced by Dr Hernán E Morales, Globe institute, University of Copenhagen.

10. Conservation genomics of the previously endangered whooping crane (*Grus americana*)

10.1 Supplementary Tables

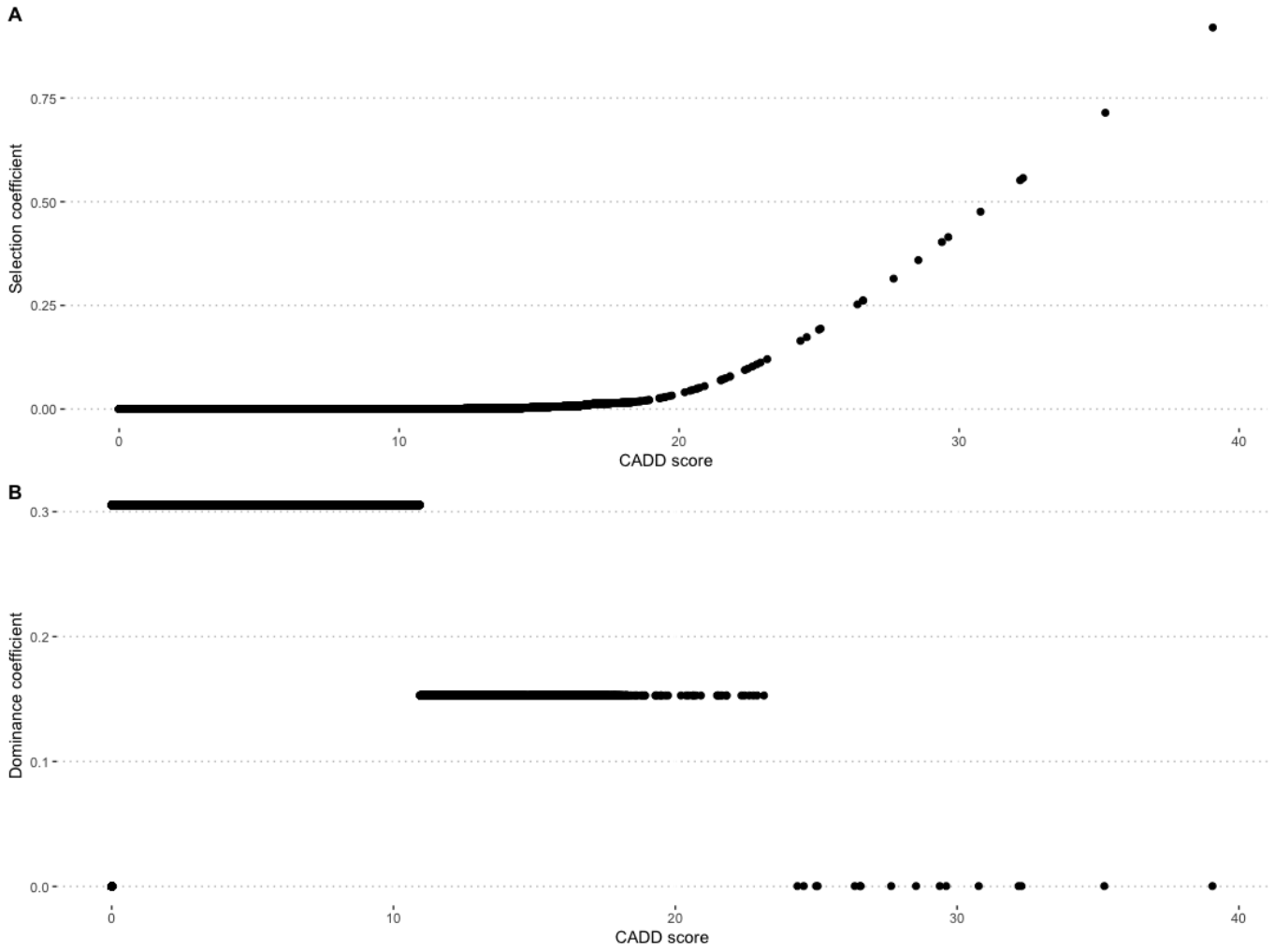
Supplementary Table 11 – The mean CADD scores and selection coefficient (s) and Standard Error (SE) for each of the four categories of deleterious mutations across the UCEs and their flanking regions. The count showing the number of each category of site is also shown.

SNPeff	mean CADD	SE	mean s	SE	Count
No SNPeff	6.419419	0.0330502	0.001745154	7.498769e-05	30064
Low	2.707494	0.3212577	0.001972806	9.093811e-04	258
Moderate	21.200541	0.3234661	0.122559374	7.706589e-03	332
High	23.549174	2.1097181	0.313749772	6.167435e-02	33

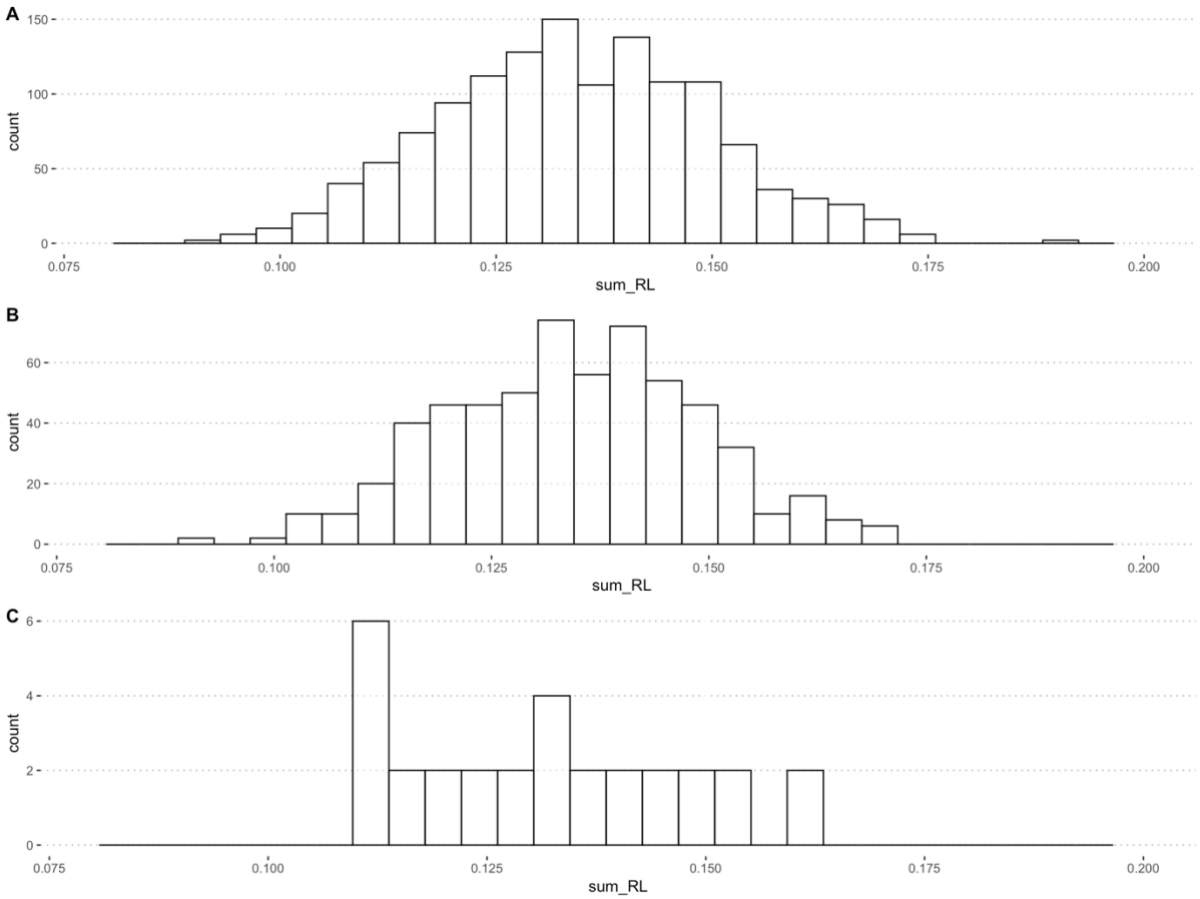
Supplementary Table 12 – The predicted realised load for all of the potential offspring of crossings between the six founder individuals sequenced. Showing the rank of the crossing in relation to all 600 potential crossings between members of the founder (n = 6) or wild population (n = 19), where 1 is the crossing with the lowest realised load and 600 is the crossing with the highest realised load, excluding any self-matings.

Parent 1 ID	Parent 1 Category	Parent 2 ID	Parent 2 Category	Realised Load	Rank	p value
EB25_S97	Founder	EB28_S100	Founder	0.1134909	43.5	0.07250000
EB25_S97	Founder	EB30_S102	Founder	0.1320935	257.5	0.42916667
EB25_S97	Founder	EB31_S103	Founder	0.1097984	25.5	0.04250000
EB25_S97	Founder	EB32_S104	Founder	0.1176418	83.5	0.13916667
EB25_S97	Founder	EB35_S107	Founder	0.1104050	27.5	0.04583333
EB28_S100	Founder	EB25_S97	Founder	0.1134909	43.5	0.07250000
EB28_S100	Founder	EB30_S102	Founder	0.1550288	555.5	0.92583333
EB28_S100	Founder	EB31_S103	Founder	0.1299011	215.5	0.35916667
EB28_S100	Founder	EB32_S104	Founder	0.1445264	445.5	0.74250000
EB28_S100	Founder	EB35_S107	Founder	0.1371752	335.5	0.55916667
EB30_S102	Founder	EB25_S97	Founder	0.1320935	257.5	0.42916667
EB30_S102	Founder	EB28_S100	Founder	0.1550288	555.5	0.92583333
EB30_S102	Founder	EB31_S103	Founder	0.1426378	423.5	0.70583333
EB30_S102	Founder	EB32_S104	Founder	0.1609218	573.5	0.95583333
EB30_S102	Founder	EB35_S107	Founder	0.1496866	519.5	0.86583333
EB31_S103	Founder	EB25_S97	Founder	0.1097984	25.5	0.04250000
EB31_S103	Founder	EB28_S100	Founder	0.1299011	215.5	0.35916667
EB31_S103	Founder	EB30_S102	Founder	0.1426378	423.5	0.70583333
EB31_S103	Founder	EB32_S104	Founder	0.1233398	143.5	0.23916667
EB31_S103	Founder	EB35_S107	Founder	0.1202559	113.5	0.18916667
EB32_S104	Founder	EB25_S97	Founder	0.1176418	83.5	0.13916667
EB32_S104	Founder	EB28_S100	Founder	0.1445264	445.5	0.74250000
EB32_S104	Founder	EB30_S102	Founder	0.1609218	573.5	0.95583333
EB32_S104	Founder	EB31_S103	Founder	0.1233398	143.5	0.23916667
EB32_S104	Founder	EB35_S107	Founder	0.1341356	293.5	0.48916667
EB35_S107	Founder	EB25_S97	Founder	0.1104050	27.5	0.04583333
EB35_S107	Founder	EB28_S100	Founder	0.1371752	335.5	0.55916667
EB35_S107	Founder	EB30_S102	Founder	0.1496866	519.5	0.86583333
EB35_S107	Founder	EB31_S103	Founder	0.1202559	113.5	0.18916667
EB35_S107	Founder	EB32_S104	Founder	0.1341356	293.5	0.48916667

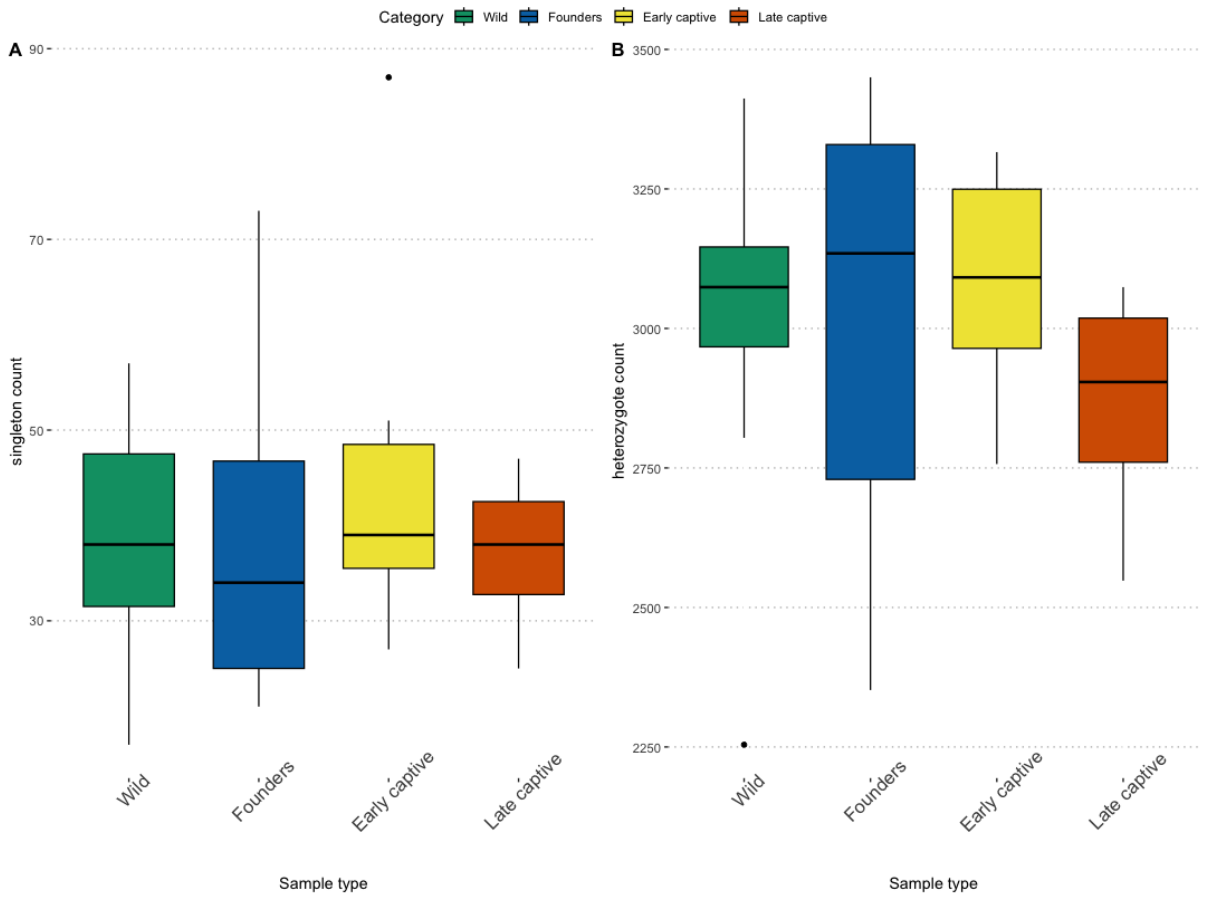
10.2 Supplementary Figures



Supplementary Figure 5 – CADD scores were converted to selection and dominance coefficients. A) The conversion of CADD scores for mutations and the selection coefficient converted using a generalised associative model (GAM) function based on ranked CADD and selection coefficients simulated using SLiM *(Haller & Messer, 2019)* (Chapter 4). **B)** The dominance coefficients of mutations for their respective CADD scores.



Supplementary Figure 6 – Histograms showing the predicted realised load within the offspring of potential crossings of the sampled whooping crane population. Showing the sum of the realised load between combinations of the wild (n = 19), founder individuals (n = 6), early captive individuals (n = 6) and late captive individual (n = 6) **A)** all potential crossing across all categories, **B)** all crossings using individuals of the wild and founder populations and **C)** founder individuals only. All selfing crossing excluded.



Supplementary Figure 7 – The number of A) singletons unique to a single individual within the UCEs and B) the count of the number of heterozygous loci per individual across the UCEs and flanking regions. A singleton was defined as a site found to be heterozygous for only one of the 37 individuals. Sites were counted across all categories of the captive breeding and wild population sampled, wild (n = 19) (Green), founder individuals (n = 6) (Blue), early members of the captive breeding programme (n = 6) (Yellow) and late members of the captive breeding programme (n = 6) (Orange).

Glossary

ALS – Amyotrophic lateral sclerosis

ART – Assisted reproductive technology

AZA - Association of Zoos and Aquariums

BAM – Binary sequence Alignment/Map file format

BCF – Binary variant Call Format file format

BED – Browser Extensible Data file format

BLAST – Basic Local Alignment Search Tool

CADD – Combined Annotation-Dependant Depletion

chCADD – Chicken Combined Annotation-Dependant Depletion

Deleterious mutation – A mutation to a gene that causes the protein product to not be produced, is produced but not functional or is produced and impacts regular function.

DFE – Distribution Fitness Effect scenarios

Dominance coefficient – The fitness of heterozygous relative to homozygous for a genotype.

Drift debt - The time lag of evolutionary genetic change during population size decline

EAZA – The European Association for Zoos and Aquaria

EEPs – EAZA *Ex situ* Programmes

Ex situ – A population outside or away from its natural location

Extinction vortex – A cyclic process resulting in the extinction of a population or species. Whereby, random genetic drift and increased inbreeding as a result of a decline in population size leads to a reduction in the fitness of individuals leading to a further reduction in population size.

FASTA – A text-based data format used to store nucleotide or amino acid sequences.

F_{IS} – Inbreeding coefficient of an individual relative to the local subpopulation

F_{ROH} – A measure of inbreeding as the proportion of the autosomal genome in runs of homozygosity.

Galgal6 – chicken reference genome

GAM – Generalised associative model

GATK – Genome Analysis Toolkit

GDEWS – Gerald Durrell Endemic Wildlife Sanctuary

Gene flow – The introduction of genetic material from one population to another

Genetic bottleneck– Sharp reductions in the effective population size (N_e) over one or multiple generations.

Genetic drift – The random fluctuations in allele frequency in a population due to finite sampling of individuals.

Genetic load – The sum of the realised load and masked load

Genetic Rescue – The reintroduction of new or previously rare genetic variation into a population with the aims of reducing inbreeding depression, increasing genetic variation and population viability.

Genetic swamping – The genetic variation of new immigrants that replaces the variation in the local gene pool, undermining local adaptation.

GERP scores – Genomic Evolutionary Rate Profiling scores

hCADD – Human Combined Annotation-Dependant Depletion

Inbreeding – Reproduction by two individuals that are closely related genetically.

Inbreeding depression – A reduction in the mean fitness of offspring arising from reproduction between related individuals.

Indel – Insertions and Deletions

In situ – A population within its natural location.

IUCN – International Union for Conservation of Nature

Karyotype – The appearance of a complete set of chromosomes.

Kinship coefficient – The probability that a randomly selected allele from a locus will be identical by descent between two individuals.

Lethal equivalent – A group of mutant alleles with a summed selection coefficient equal to one. For a population, it is expressed as a grouping of genes or mutations that together cause, on average, the death of one individual.

LoadLift Pipeline – A Snakemake pipeline for the transfer of CADD scores from model species to non-model species.

Masked load – The fitness effects of all (partially) recessive mutations that are in heterozygous conditions, and which are not completely expressed.

MHC – Major histocompatibility complexes.

MWF – Mauritius Wildlife Foundation

NCBI database – The National Centre for Biotechnology Information database. An online resource of biological information and data.

N_e – Effective population size

NPCS – National Parks and Conservation Service

OMIA – Online Mendelian Inheritance in Animals

One Plan approach – A method for species conservation and management to develop a conservation plan with input from all stakeholders involved with both *in situ* and *ex situ* populations of the species.

Outbreeding depression (c.f. Reproductive isolation) - a reduction in reproductive fitness of a population after gene flow due to either pre or postzygotic isolation, or a combination of both (Frankham et al., 2011)

pCADD – Pig Combined Annotation-Dependant Depletion

Pi – The neutral nucleotide diversity (π).

ppCADD – Pink pigeon Combined Annotation-Dependant Depletion

Purging – The reduction in genetic load by purifying selection operating against recessive deleterious variants that have become exposed in a homozygous state.

Purifying selection – Selection that removes deleterious mutations.

Realised load – The proportion of deleterious mutations that are present as homozygotes, plus the fitness effects of the partially recessive, additive and dominant mutations that are in heterozygous loci.

Reinforcement – The translocation of an organism into an existing population of the same species within its natural range, with the aim to improve population viability (GOV.UK, 2024).

Reintroduction – The translocation of an organism inside its natural range, to areas from which it has been lost. Reintroduction aims to re-establish a viable population of the focal species within its natural range (GOV.UK, 2024)

Reintroduction programme – The release of individuals either raised or rehabilitated in captivity into their natural environment, to stabilise, reestablish, or increase *in-situ* populations that have suffered significant declines (AZA, 2024).

Reproductive isolation (c.f. Outbreeding depression) - a reduction in reproductive fitness of a population after gene flow due to either pre or postzygotic isolation, or a combination of both (Frankham et al., 2011)

ROH - Runs of Homozygosity

SAM – Sequence Alignment/Map

Selection coefficients – A measure of the relative reduction in contribution to gametes a genotype causes relative to other genotypes in a population (s).

Selfing – Where offspring are produced from one parent alone.

Singletons - Mutations occurring in one copy across all 37 individuals.

SNP – Single Nucleotide Polymorphism

UCE – Ultraconserved Elements

VCF – Variant Call Format

WAZA – World Association of Zoos and Aquariums

References

Adams, J.R., Vucetich, L.M., Hedrick, P.W., Peterson, R.O. and Vucetich, J.A. 2011. Genomic sweep and potential genetic rescue during limiting environmental conditions in an isolated wolf population. Proceedings of the Royal Society B: Biological Sciences. 278(1723), pp.3336–3344.

Ager, S.L. 2005. Familiarity breeds: incest and the Ptolemaic dynasty. The Journal of Hellenic Studies. 125, pp.1–34.

Albeshr, M. 2016. Conservation biology and genomics of a flagship endangered species: the Mauritian pink pigeon Nesoenas mayeri [Online]. Available from: https://ueaeprints.uea.ac.uk/id/eprint/59458/.

Al Hikmani, H., van Oosterhout, C., Birley, T., Labisko, J., Jackson, H.A., Spalton, A., Tollington, S. and Groombridge, J.J. 2024. Can genetic rescue help save Arabia's last big cat? Evolutionary Applications. 17(5).

Alhosain, A. 2018. Premarital screening programs in the Middle East, from a human right's perspective. Diversity Equality Health Care. 15.

Alkalbani, A., Alharrasi, M., Achura, S., Al Badi, A., Al Rumhi, A., Alqassabi, K., Almamari, R. and Alomari, O. 2022. Factors Affecting the Willingness to Undertake Premarital Screening Test Among Prospective Marital Individuals. SAGE Open Nursing. 8, p.237796082210781.

Allen, R.P. 1952. The Whooping Crane. Research Report No. 3. National Audubon Society., p.246.

Álvarez, G., Ceballos, F.C. and Berra, T.M. 2015. Darwin was right: inbreeding depression on male fertility in the Darwin family. Biological Journal of the Linnean Society. 114(2), pp.474–483.

Alvarez, G., Ceballos, F.C. and Quinteiro, C. 2009. The Role of Inbreeding in the Extinction of a European Royal Dynasty. PLOS ONE. 4(4), pp.1–7.

Anderson, R., Waayers, R. and Knight, A. 2016. Orca Behavior and Subsequent Aggression Associated with Oceanarium Confinement. Animals. 6(8), p.49.

Ando, H. 2019. Genetic and ecological conservation issues for oceanic island birds, revealed by a combination of the latest molecular techniques and conventional field work. Ecological Research. 34(2), pp.255–264.

Angert, A.L., Bontrager, M.G. and Ågren, J. 2020. What Do We Really Know About Adaptation at Range Edges? Annual Review of Ecology, Evolution, and Systematics. 51(1), pp.341–361.

Arenas, M., Araujo, N.M., Branco, C., Castelhano, N., Castro-Nallar, E. and Pérez-Losada, M. 2018. Mutation and recombination in pathogen evolution: Relevance, methods and controversies. Infection, Genetics and Evolution. 63, pp.295–306.

Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I.T., Novak, A.M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., Genereux, D., Johnson, J., Marinescu, V.D., Alföldi, J., Harris, R.S., Lindblad-Toh, K., Haussler, D., Karlsson, E., Jarvis, E.D., Zhang, G. and Paten, B. 2020. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature. 587(7833), pp.246–251.

Arya, D., Gautam, P., Upreti, D. and Sharma, M. 2024. Bioinformatics and its application in veterinary sciences. International Journal of Advanced Biochemistry Research. 8(6S), pp.260–264.

AZA 2024. Reintroduction Programs. [Accessed 10 December 2024]. Available from: https://www.aza.org/reintroduction-programs.

Badaruddoza, Afzal, Mohd. and Akhtaruzzaman 1994. Inbreeding and congenital heart diseases in a North Indian population. Clinical Genetics. 45(6), pp.288–291.

Baker, R.J. and Bickham, J.W. 1986. Speciation by Monobrachial Centric Fusions. Proceedings of the National Academy of Sciences of the United States of America. 83(21), pp.8245–8248.

Balick, D.J., Do, R., Cassa, C.A., Reich, D. and Sunyaev, S.R. 2015. Dominance of Deleterious Alleles Controls the Response to a Population Bottleneck. PLOS Genetics. 11(8), p.e1005436.

Bandoli, F. and Cavicchio, P. 2021. The COVID-19 Pandemic and the Fragile Balance of a Small Zoo: the Case of Pistoia Zoo in Italy. Journal of Applied Animal Ethics Research. 3(1), pp.57–73.

Barbosa, S., Mestre, F., White, T.A., Paupério, J., Alves, P.C. and Searle, J.B. 2018. Integrative approaches to guide conservation decisions: Using genomics to define conservation units and functional corridors. Molecular Ecology. 27(17), pp.3452–3465.

Barnett, R., Yamaguchi, N., Shapiro, B. and Sabin, R. 2008. Ancient DNA analysis indicates the first English lions originated from North Africa. Contributions to Zoology. 77(1), pp.7–16.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S. and Haussler, D. 2004. Ultraconserved Elements in the Human Genome. Science. 304(5675), pp.1321–1325.

Bell, D.A., Kovach, R.P., Robinson, Z.L., Whiteley, A.R. and Reed, T.E. 2021. The ecological causes and consequences of hard and soft selection. Ecology Letters. 24(7), pp.1505–1521.

Bell, D.A., Robinson, Z.L., Funk, W.C., Fitzpatrick, S.W., Allendorf, F.W., Tallmon, D.A. and Whiteley, A.R. 2019. The Exciting Potential and Remaining Uncertainties of Genetic Rescue. Trends in Ecology & Evolution. 34(12), pp.1070–1079.

Benson, J.F., Mahoney, P.J., Vickers, T.W., Sikich, J.A., Beier, P., Riley, S.P.D., Ernest, H.B. and Boyce, W.M. 2019. Extinction vortex dynamics of top predators isolated by urbanization. Ecological Applications. 29(3).

Benzenhöfer U and Ackermann H 2015. Die Zahl der Verfahren und der Sterilisationen nach der Gesetz zur Verhütung erbkranken Nachwuchses. Kontur, Münster.

Berger, D., Stångberg, J., Grieshop, K., Martinossi-Allibert, I. and Arnqvist, G. 2017. Temperature effects on life-history trade-offs, germline maintenance and mutation rate under simulated climate warming. Proceedings of the Royal Society B: Biological Sciences. 284(1866), p.20171721.

Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., Morales, H.E. and van Oosterhout, C. 2022. Genetic load: genomic estimates and applications in non-model animals. Nature Reviews Genetics. 23(8), pp.492–503.

BirdLife International 2020. Grus americana. The IUCN Red List of Threatened Species . [Online]. [Accessed 15 August 2024]. Available from: https://dx.doi.org/10.2305/IUCN.UK.2020-3.RLTS.T22692156A181242855.en.

BirdLife International 2021. Nesoenas mayeri, Pink Pigeon. The IUCN Red List of Threatened Species. [Online]. [Accessed 15 August 2024]. Available from: https://dx.doi.org/10.2305/IUCN.UK.2021- 3.RLTS.T22690392A179390191.en.

Boakes, E.H., Wang, J. and Amos, W. 2007. An investigation of inbreeding depression and purging in captive pedigreed populations. Heredity. 98(3), pp.172–182.

Bolton, R.L., Mooney, A., Pettit, M.T., Bolton, A.E., Morgan, L., Drake, G.J., Appeltant, R., Walker, S.L., Gillis, J.D. and Hvilsom, C. 2022. Resurrecting biodiversity: advanced assisted reproductive technologies and biobanking. Reproduction and Fertility. 3(3), pp.R121–R146.

Bortoluzzi, C., Bosse, M., Derks, M.F.L., Crooijmans, R.P.M.A., Groenen, M.A.M. and Megens, H.-J. 2020. The type of bottleneck matters: Insights into the deleterious variation landscape of small managed populations. Evolutionary Applications. 13(2), pp.330–341.

Bosse, M. and van Loon, S. 2022. Challenges in quantifying genome erosion for conservation. Frontiers in Genetics. 13.

Bossu, C.M., Rodriguez, M., Rayne, C., Chromczak, D.A., Higgins, P.G., Trulio, L.A. and Ruegg, K.C. 2023. Genomic approaches to mitigating genetic diversity loss in declining populations. Molecular Ecology. 32(19).

Breithoff, E. and Harrison, R. 2020. From ark to bank: extinction, proxies and biocapitals in ex-situ biodiversity conservation practices. International Journal of Heritage Studies. 26(1), pp.37–55.

Broad Institute 2019. Picard toolkit.

Burleigh M 1994. Death and Deliverance. 'Euthanasia' in Germany 1900–1945. Cambridge, UK.: Cambridge University Press.

Burt, A. 2000. Perspective: sex, recombination, and the efficacy of selection — Was Weismann right? Evolution. 54(2), pp.337–351.

Butler, Matthew J., Harris, G. and Strobel, B.N. 2013. Influence of whooping crane population dynamics on its recovery and management. Biological Conservation. 162, pp.89–99.

Butler, Matthew J, Harris, G. and Strobel, B.N. 2013. Influence of whooping crane population dynamics on its recovery and management. Biological Conservation. 162, pp.89–99.

Butlin, R.K. 2005. Recombination and speciation. Molecular Ecology. 14(9), pp.2621–2635.

Carbone, L., Nergadze, S.G., Magnani, E., Misceo, D., Francesca Cardone, M., Roberto, R., Bertoni, L., Attolini, C., Francesca Piras, M., de Jong, P., Raudsepp, T., Chowdhary, B.P., Guérin, G., Archidiacono, N., Rocchi, M. and Giulotto, E. 2006. Evolutionary movement of centromeres in horse, donkey, and zebra. Genomics. 87(6), pp.777–782.

Carlson E F 2001. The Unfit. A History of a Bad Idea. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Caven, A.J., Rabbe, M., Malzahn, J. and Lacy, A.E. 2020. Trends in the occurrence of large Whooping Crane groups during migration in the great plains, USA. Heliyon. 6(4), p.e03549.

CBD 2022. COP15: FINAL TEXT OF KUNMING-MONTREAL GLOBAL BIODIVERSITY FRAMEWORK [Online]. Montreal, Canada. [Accessed 16 August 2024]. Available from: https://www.cbd.int/article/cop15-final-text-kunming-montreal-gbf-221222.

Ceballos, F.C. and Álvarez, G. 2013. Royal dynasties as human inbreeding laboratories: the Habsburgs. Heredity. 111(2), pp.114–121.

Ceballos, G., Ehrlich, P.R. and Dirzo, R. 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proceedings of the National Academy of Sciences. 114(30).

Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics. 10(3), pp.195–205.

Chen, R.S., Soulsbury, C.D., Lebigre, C., Ludwig, G., van Oers, K. and Hoffman, J.I. 2023. Effects of hunting on genetic diversity, inbreeding and dispersal in Finnish black grouse (Lyrurus tetrix). Evolutionary Applications. 16(3), pp.625–637.

Chester Zoo 2020. An update on our ‘Save our Zoo’ campaign. [Accessed 18 June 2024]. Available from: https://www.chesterzoo.org/news/an-update-on-our-save-our-zoo-campaign/.

Christie, M.R., Marine, M.L., French, R.A. and Blouin, M.S. 2012. Genetic adaptation to captivity can occur in a single generation. Proceedings of the National Academy of Sciences. 109(1), pp.238–242.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D.M. 2012. 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. 6(2), pp.80–92.

Clegg, S.M. and Phillimore, A.B. 2010. The influence of gene flow and drift on genetic and phenotypic divergence in two species of Zosterops in Vanuatu. Philosophical Transactions of the Royal Society B: Biological Sciences. 365(1543), pp.1077–1092.

Cohen, E. and Fennell, D. 2016. The elimination of Marius, the giraffe: humanitarian act or callous management decision? Tourism Recreation Research. 41(2), pp.168–176.

Conde, D.A., Colchero, F., Gusset, M., Pearce-Kelly, P., Byers, O., Flesness, N., Browne, R.K. and Jones, O.R. 2013. Zoos through the Lens of the IUCN Red List: A Global Metapopulation Approach to Support Conservation Breeding Programs. PLoS ONE. 8(12), p.e80311.

Converse, S.J., Strobel, B.N. and Barzen, J.A. 2019. Reproductive Failure in the Eastern Migratory Population: The Interaction of Research and Management In: Whooping Cranes: Biology and Conservation. Elsevier, pp.161–178.

Cooke, S. 2011. Duties to Companion Animals. Res Publica. 17(3), pp.261–274.

Crow, J.F. 1970. Genetic loads and the cost of natural selection In: Ken-ichi K, ed. Mathematical Topics in Population Genetics. Springer, Berlin, pp.128–177.

Dalal, V., Pasupuleti, N., Chaubey, G., Rai, N. and Shinde, V. 2023. Advancements and Challenges in Ancient DNA Research: Bridging the Global North–South Divide. Genes. 14(2), p.479.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G. and Durbin, R. 2011. The variant call format and VCFtools. Bioinformatics. 27(15), pp.2156–2158.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M. and Li, H. 2021. Twelve years of SAMtools and BCFtools. GigaScience. 10(2).

Darwin, C. 1859. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life [Online]. London : John Murray, 1859. Available from: https://search.library.wisc.edu/catalog/9934839413602122.

Davydov, E. V, Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A. and Batzoglou, S. 2010. Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. PLOS Computational Biology. 6(12), pp.e1001025-.

Dennis, B., Assas, L., Elaydi, S., Kwessi, E. and Livadiotis, G. 2016. Allee effects and resilience in stochastic populations. Theoretical Ecology. 9(3), pp.323–335.

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

Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N.J.B. and Collen, B. 2014. Defaunation in the Anthropocene. Science. 345(6195), pp.401–406.

Doekes, H.P., Curik, I., Nagy, I., Farkas, J., Kövér, G. and Windig, J.J. 2020. Revised Calculation of Kalinowski's Ancestral and New Inbreeding Coefficients. Diversity. 12(4), p.155.

Drake, J.W., Charlesworth, B., Charlesworth, D. and Crow, J.F. 1998. Rates of Spontaneous Mutation. Genetics. 148(4), pp.1667–1686.

Durrell, G. 1977. The Stationary Ark [Online]. Fontana. Available from: https://books.google.co.uk/books?id=_75JpJ3xW4oC.

Dussex, N., Morales, H.E., Grossen, C., Dalén, L. and van Oosterhout, C. 2023. Purging and accumulation of genetic load in conservation. Trends in Ecology & Evolution.

EAZA 2024a. EAZA Biobank. [Accessed 16 August 2024]. Available from: https://www.eaza.net/conservation/research/eaza-biobank/.

EAZA 2024b. EAZA Ex situ Programmes. [Accessed 11 August 2024]. Available from: EAZA Ex situ Programmes.

ENCODE 2012. An integrated encyclopedia of DNA elements in the human genome. Nature. 489(7414), pp.57–74.

Fa JE, F.S.O.D. 2011. Captive breeding and zoos In: J. E. Fa, S. M. Funk and D. O'Connell, eds. Zoo Conservation Biology [Online]. Ecology, Biodiversity and Conservation. Cambridge: Cambridge University Press, pp.142–189. Available from: https://www.cambridge.org/core/product/990BC3A0042B135FBA1D21BB49217593.

Faircloth, B.C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics. 32(5), pp.786–788.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T. and Glenn, T.C. 2012. Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales. Systematic Biology. 61(5), pp.717–726.

Fareed, M. and Afzal, M. 2014. Evidence of inbreeding depression on height, weight, and body mass index: A population-based child cohort study. American Journal of Human Biology. 26(6), pp.784–795.

Fareed, M. and Afzal, M. 2016. Increased cardiovascular risks associated with familial inbreeding: a population-based study of adolescent cohort. Annals of Epidemiology. 26(4), pp.283–292.

Fareed, M., Kaisar Ahmad, M., Azeem Anwar, M. and Afzal, M. 2017. Impact of consanguineous marriages and degrees of inbreeding on fertility, child mortality, secondary sex ratio, selection intensity, and genetic load: a cross-sectional study from Northern India. Pediatric Research. 81(1), pp.18–26.

Femerling, G., van Oosterhout, C., Feng, S., Bristol, R.M., Zhang, G., Groombridge, J., P. Gilbert, M.T. and Morales, H.E. 2023. Genetic Load and Adaptive Potential of a Recovered Avian Species that Narrowly Avoided Extinction. Molecular Biology and Evolution. 40(12), p.msad256.

Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A.H., Xie, D., Chen, G., Guo, C., Faircloth, B.C., Petersen, B., Wang, Z., Zhou, Q., Diekhans, M., Chen, W., Andreu-Sánchez, S., Margaryan, A., Howard, J.T., Parent, C., Pacheco, G., Sinding, M.-H.S., Puetz, L., Cavill, E., Ribeiro, Â.M., Eckhart, L., Fjeldså, J., Hosner, P.A., Brumfield, R.T., Christidis, L., Bertelsen, M.F., Sicheritz-Ponten, T., Tietze, D.T., Robertson, B.C., Song, G., Borgia, G., Claramunt, S., Lovette, I.J., Cowen, S.J., Njoroge, P., Dumbacher, J.P., Ryder, O.A., Fuchs, J., Bunce, M., Burt, D.W., Cracraft, J., Meng, G., Hackett, S.J., Ryan, P.G., Jønsson, K.A., Jamieson, I.G., da Fonseca, R.R., Braun, E.L., Houde, P., Mirarab, S., Suh, A., Hansson, B., Ponnikas, S., Sigeman, H., Stervander, M., Frandsen, P.B., van der Zwan, H., van der Sluis, R., Visser, C.,

Balakrishnan, C.N., Clark, A.G., Fitzpatrick, J.W., Bowman, R., Chen, N., Cloutier, A., Sackton, T.B., Edwards, S. V, Foote, D.J., Shakya, S.B., Sheldon, F.H., Vignal, A., Soares, A.E.R., Shapiro, B., González-Solís, J., Ferrer-Obiol, J., Rozas, J., Riutort, M., Tigano, A., Friesen, V., Dalén, L., Urrutia, A.O., Székely, T., Liu, Y., Campana, M.G., Corvelo, A., Fleischer, R.C., Rutherford, K.M., Gemmell, N.J., Dussex, N., Mouritsen, H., Thiele, N., Delmore, K., Liedvogel, M., Franke, A., Hoeppner, M.P., Krone, O., Fudickar, A.M., Milá, B., Ketterson, E.D., Fidler, A.E., Friis, G., Parody-Merino, Á.M., Battley, P.F., Cox, M.P., Lima, N.C.B., Prosdocimi, F., Parchman, T.L., Schlinger, B.A., Loiselle, B.A., Blake, J.G., Lim, H.C., Day, L.B., Fuxjager, M.J., Baldwin, M.W., Braun, M.J., Wirthlin, M., Dikow, R.B., Ryder, T.B., Camenisch, G., Keller, L.F., DaCosta, J.M., Hauber, M.E., Louder, M.I.M., Witt, C.C., McGuire, J.A., Mudge, J., Megna, L.C., Carling, M.D., Wang, B., Taylor, S.A., Del-Rio, G., Aleixo, A., Vasconcelos, A.T.R., Mello, C. V, Weir, J.T., Haussler, D., Li, Q., Yang, H., Wang, J., Lei, F., Rahbek, C., Gilbert, M.T.P., Graves, G.R., Jarvis, E.D., Paten, B. and Zhang, G. 2020. Dense sampling of bird diversity increases power of comparative genomics. Nature. 587(7833), pp.252–257.

Fiorini, M.R., Dilliott, A.A. and Farhan, S.M.K. 2023. Evaluating the utility of REVEL and CADD for interpreting variants in amyotrophic lateral sclerosis genes. Human Mutation. 2023(1), p.8620557.

Fitzpatrick, S.W., Mittan-Moreau, C., Miller, M. and Judson, J.M. 2023. Genetic rescue remains underused for aiding recovery of federally listed vertebrates in the United States. Journal of Heredity. 114(4), pp.354–366.

Folk, M.J., J.A. Rodgers Jr., T.A. Dellinger, S.A. Nesbitt, J.M. Parker, M.G. Spalding, S.B. Baynes, M.K. Chappell and S.T. Schwikert. 2010. Status of non-migratory whooping cranes in Florida. In: B. K. Hartup, ed. Proceedings of the Eleventh North American Crane Workshop [Online]. Wisconsin Dells, Wisconsin: Baraboo, WI: North American Crane Working Group, pp.118-123. [Accessed 11 December 2024]. Available from: https://digitalcommons.unl.edu/nacwgproc/101/.

Forester, B.R., Murphy, M., Mellison, C., Petersen, J., Pilliod, D.S., Van Horne, R., Harvey, J. and Funk, W.C. 2022. Genomics-informed delineation of conservation units in a desert amphibian. Molecular Ecology. 31(20), pp.5249–5269.

Foster, K.P. 1999. The Earliest Zoos and Gardens. Scientific American. 281(1), pp.64–71.

Frankel, O.H. 1974. Genetic conservation: our evolutionary responsibility. Genetics. 78(1), pp.53–65.

Frankel, O.H. and Bennett, E. 1970. Genetic resources in plants - their exploration and conservation. Oxford, UK: Blackwell Scientific Publications.

Frankham, R. 1997. Do island populations have less genetic variation than mainland populations? Heredity. 78(3), pp.311–327.

Frankham, R. 2008. Genetic adaptation to captivity in species conservation programs. Molecular Ecology. 17(1), pp.325–333.

Frankham, R. 2016. Genetic rescue benefits persist to at least the F3 generation, based on a meta-analysis. Biological Conservation. 195, pp.33–36.

Frankham, R. 2015. Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. Molecular Ecology. 24(11), pp.2610–2618.

Frankham, R., Ballou, J.D., Eldrige, M.D.B., Lacy, R.C., Ralls, K., Dudash, M.R. and Fenster, C.B. 2011. Predicting the Probability of Outbreeding Depression. Conservation Biology. 25(3), pp.465–475.

Frankham, R., Ballou, J.D., Ralls, K., Eldridge, M.D.B., Dudash, M.R., Fenster, C.B., Lacy, R.C. and Sunnucks, P. 2017. Genetic rescue by augmenting gene flow In: Genetic Management of Fragmented Animal and Plant Populations. Oxford University PressOxford, pp.115–134.

Fritts, R. 2022. Cloning goes wild. Science. 375(6577), pp.134–137.

Funk, W.C., McKay, J.K., Hohenlohe, P.A. and Allendorf, F.W. 2012. Harnessing genomics for delineating conservation units. Trends in Ecology & Evolution. 27(9), pp.489–496.

Fusco, N.A., Carlen, E.J. and Munshi-South, J. 2021. Urban Landscape Genetics: Are Biologists Keeping Up with the Pace of Urbanization? Current Landscape Ecology Reports. 6(2), pp.35–45.

Galeota-Sprung, B., Sniegowski, P. and Ewens, W. 2020. Mutational Load and the Functional Fraction of the Human Genome. Genome Biology and Evolution. 12(4), pp.273–281.

García-Dorado, A. 2012. Understanding and Predicting the Fitness Decline of Shrunk Populations: Inbreeding, Purging, Mutation, and Standard Selection. Genetics. 190(4), pp.1461–1476.

García-Dorado, A. and Caballero, A. 2021. Neutral genetic diversity as a useful tool for conservation biology. Conservation Genetics. 22(4), pp.541–545.

García-Dorado, A. and Caballero, A. 2000. On the Average Coefficient of Dominance of Deleterious Spontaneous Mutations. Genetics. 155(4), pp.1991–2001.

Gelinas, S.M., Benson, C.E., Khan, M.A., Berger, R.M.F., Trembath, R.C., Machado, R.D. and Southgate, L. 2020. Whole exome sequence analysis provides novel insights into the genetic framework of childhood-onset pulmonary arterial hypertension. Genes. 11(11), p.1328.

General Assembly resolution 2200A (XXI) 1966. International Covenant on Civil and Political Rights [Online]. [Accessed 10 August 2024]. Available from: https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights.

Gilbert, T., Gardner, R., Kraaijeveld, A.R. and 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), pp.15–31.

Gillespie, J.H. 2000. Genetic Drift in an Infinite Population: The Pseudohitchhiking Model. Genetics. 155(2), pp.909–919.

Gilpin, M.E. and M. E. Soulé 1986. Minimum viable populations: processes of extinction In: M.E. Soulé, ed. Conservation Biology: The Science of Scarcity and Diversity. Sunderland, MA: Sinauer, pp.19–34.

Gilroy, D.L., Phillips, K.P., Richardson, D.S. and van Oosterhout, C. 2017. Toll-like receptor variation in the bottlenecked population of the Seychelles warbler: computer simulations see the 'ghost of selection past' and quantify the 'drift debt'. Journal of Evolutionary Biology. 30(7), pp.1276–1287.

Golden, K.E., Hemingway, B.L., Frazier, A.E., Scholtz, R., Harrell, W., Davis, C.A. and Fuhlendorf, S.D. 2022. Spatial and temporal predictions of whooping crane (Grus americana) habitat along the US Gulf Coast. Conservation Science and Practice. 4(6).

Goto, H., Ryder, O.A., Fisher, A.R., Schultz, B., Kosakovsky Pond, S.L., Nekrutenko, A. and Makova, K.D. 2011. A Massively Parallel Sequencing Approach Uncovers Ancient Origins and High Genetic Variability of Endangered Przewalski's Horses. Genome Biology and Evolution. 3, pp.1096–1106.

GOV.UK 2024. Species reintroductions and other conservation translocations. [Accessed 10 December 2024]. Available from: https://www.gov.uk/guidance/species-reintroductions-and-other-conservation-translocations.

Gregory, T.R. 2009. Understanding Natural Selection: Essential Concepts and Common Misconceptions. Evolution: Education and Outreach. 2(2), pp.156–175.

Groß, C., Bortoluzzi, C., de Ridder, D., Megens, H.-J., Groenen, M.A.M., Reinders, M. and Bosse, M. 2020. Prioritizing sequence variants in conserved non-coding elements in the chicken genome using chCADD. PLOS Genetics. 16(9), pp.e1009027-.

Groß, C., Derks, M., Megens, H.-J., Bosse, M., Groenen, M.A.M., Reinders, M. and de Ridder, D. 2020. pCADD: SNV prioritisation in Sus scrofa. Genetics Selection Evolution. 52(1), p.4.

Groß, C., de Ridder, D. and Reinders, M. 2018. Predicting variant deleteriousness in non-human species: applying the CADD approach in mouse. BMC Bioinformatics. 19(1), p.373.

Grossen, C., Guillaume, F., Keller, L.F. and Croll, D. 2020. Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nature Communications. 11(1), p.1001.

Guhlin, J., Le Lec, M.F., Wold, J., Koot, E., Winter, D., Biggs, P.J., Galla, S.J., Urban, L., Foster, Y., Cox, M.P., Digby, A., Uddstrom, L.R., Eason, D., Vercoe, D., Davis, T., Andrew, K., Argilla, L., Arnold, K., Bohan, J., Bolitho, L., Brown, N., Carpenter, J., Crane, J., Grant, M., Greaves, G., Halkett, B., Hannan, R., Haultain, S., Hitchcock, B., Joyce, L., Larcombe, S., Ledington, J., MacTavish, J., Marsh, P., Mingam, G., Moore, F., Murray, L., Nye, E., Osborne, J., Parker, L., Phillips, C., Phillips, R., Philp, B., Raemaekers, T., Rickett, J., Rouse, R., Sagar, R., Sherriff, A., Thompson, T., de Wetering, J. Van, van Zyl, N., Waite, J., Watts, J., Howard, J.T., Jarvis, E.D., Robertson, F.E., Robertson, B.C., Gemmell, N.J., Steeves, T.E., Santure, A.W., Dearden, P.K. and Team, K.R. 2023. Species-wide genomics of kākāpō provides tools to accelerate recovery. Nature Ecology & Evolution. 7(10), pp.1693–1705.

Gusset, M. and Dick, G. 2011. The global reach of zoos and aquariums in visitor numbers and conservation expenditures. Zoo biology. 30(5), pp.566–569.

Habic, A., Mattick, J.S., Calin, G.A., Krese, R., Konc, J. and Kunej, T. 2019. Genetic Variations of Ultraconserved Elements in the Human Genome. OMICS: A Journal of Integrative Biology. 23(11), pp.549–559.

Haldane, J.B. 1956. The relation between density regulation and natural selection. Proceedings of the Royal Society of London. Series B, Biological sciences. 145(920), pp.306–8.

Haller, B.C. and Messer, P.W. 2019. SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. Molecular Biology and Evolution. 36(3), pp.632–637.

Hamilton, J.A. and Miller, J.M. 2016. Adaptive introgression as a resource for management and genetic conservation in a changing climate. Conservation Biology. 30(1), pp.33–41.

Hampe, A. and Petit, R.J. 2005. Conserving biodiversity under climate change: the rear edge matters. Ecology Letters. 8(5), pp.461–467.

Hao, Y., Brockhurst, M.A., Petchey, O.L. and Zhang, Q. 2015. Evolutionary rescue can be impeded by temporary environmental amelioration. Ecology Letters. 18(9), pp.892–898.

Hare, B., Wobber, V. and Wrangham, R. 2012. The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. Animal Behaviour. 83(3), pp.573–585.

Harris, R.S. 2007. Improved pairwise alignment of genomic DNA. The Pennsylvania State University.

Harrisson, K.A., Pavlova, A., Telonis-Scott, M. and Sunnucks, P. 2014. Using genomics to characterize evolutionary potential for conservation of wild populations. Evolutionary Applications. 7(9), pp.1008–1025.

Hartup, B.K. 2018. Rearing and release methods for reintroduction of captive-reared Whooping Cranes (Chapter 20) In: J. French J B, Converse S J and Austin J E, eds. Whooping Cranes: Biology and Conservation. Biodiversity of the World: Conservation from Genes to Landscapes. San Diego, California, USA: Academic Press, pp.433–447.

Hawass, Z., Gad, Y.Z., Ismail, S., Khairat, R., Fathalla, D., Hasan, N., Ahmed, A., Elleithy, H., Ball, M., Gaballah, F., Wasef, S., Fateen, M., Amer, H., Gostner, P., Selim, A., Zink, A. and Pusch, C.M. 2010. Ancestry and Pathology in King Tutankhamun's Family. JAMA. 303(7), pp.638–647.

Hedrick, P.W. 1994. Purging inbreeding depression and the probability of extinction: full-sib mating. Heredity. 73(4), pp.363–372.

Henn, B.M., Botigué, L.R., Bustamante, C.D., Clark, A.G. and Gravel, S. 2015. Estimating the mutation load in human genomes. Nature Reviews. Genetics. 16(6), pp.333–343.

Herman William C 1948. The Last Passenger Pigeon. The Auk. 65(1), pp.77–80.

Hoffmann, A.A., Miller, A.D. and Weeks, A.R. 2021. Genetic mixing for population management: From genetic rescue to provenancing. Evolutionary Applications. 14(3), pp.634–652.

Honda, T., Nomura, T., Fukushima, M. and Mukai, F. 2002. Gene dropping analysis of founder contributions in a closed Japanese black cattle population. Animal Science Journal. 73(2), pp.105–111.

Howell, L.G., Frankham, R., Rodger, J.C., Witt, R.R., Clulow, S., Upton, R.M.O. and Clulow, J. 2021. Integrating biobanking minimises inbreeding and produces significant cost benefits for a threatened frog captive breeding programme. Conservation Letters. 14(2).

Howell, L.G., Mawson, P.R., Frankham, R., Rodger, J.C., Upton, R.M.O., Witt, R.R., Calatayud, N.E., Clulow, S. and Clulow, J. 2021. Integrating biobanking could produce significant cost benefits and minimise inbreeding for Australian amphibian captive breeding programs. Reproduction, Fertility and Development. 33(9), pp.573–587.

Huber, C.D., Kim, B.Y. and Lohmueller, K.E. 2020. Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. PLOS Genetics. 16(5), pp.e1008827-.

Hung, C.-M., Shaner, P.-J.L., Zink, R.M., Liu, W.-C., Chu, T.-C., Huang, W.-S. and Li, S.-H. 2014. Drastic population fluctuations explain the rapid extinction of the passenger pigeon. Proceedings of the National Academy of Sciences. 111(29), pp.10636–10641.

International Crane Foundation 2021. International Crane Foundation Annual Report 2021.

IUCN 2024. IUCN. 2024. The IUCN Red List of Threatened Species. Version 2024-1. [Accessed 29 January 2024]. Available from: 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., Whitford, H., Clavijo, B., Bunbury, N., Tyler, K.M., Ruhomaun, K., Grace, M.K., Bruford,

M.W., Jones, C.G., Tollington, S., Bell, D.J., Groombridge, J.J., Clark, M. and Van Oosterhout, C. 2022. Genomic erosion in a demographically recovered bird species during conservation rescue. Conservation Biology. 36(4), p.e13918.

Jensen, E.L., Gray, R. and Miller, J.M. 2024. Leveraging genomic load estimates to optimize captive breeding programmes. Molecular Ecology Resources.

Johnson, W.E., Onorato, D.P., Roelke, M.E., Land, E.D., Cunningham, M., Belden, R.C., McBride, R., Jansen, D., Lotz, M., Shindle, D., Howard, J., Wildt, D.E., Penfold, L.M., Hostetler, J.A., Oli, M.K. and O'Brien, S.J. 2010. Genetic Restoration of the Florida Panther. Science. 329(5999), pp.1641–1645.

Jones, C.G., Bunbury, N., Zuel, N., Tatayah, V. and Swinnerton, K.J. 2013. Pink Pigeon Nesoenas mayeri In: R. J. Safford and A. F. A. Hawkins, eds. The Birds of Africa. Vol. VIII: The Malagasy Region,. London, pp.484–489.

de Jong, M.J., van Oosterhout, C., Hoelzel, A.R. and Janke, A. 2024. Moderating the neutralist–selectionist debate: exactly which propositions are we debating, and which arguments are valid? Biological Reviews. 99(1), pp.23–55.

Kagan, R., Carter, S. and Allard, S. 2015. A Universal Animal Welfare Framework for Zoos. Journal of Applied Animal Welfare Science. 18(sup1), pp.S1–S10.

Kardos, M., Armstrong, E.E., Fitzpatrick, S.W., Hauser, S., Hedrick, P.W., Miller, J.M., Tallmon, D.A. and Funk, W.C. 2021. The crucial role of genome-wide genetic variation in conservation. Proceedings of the National Academy of Sciences. 118(48).

Kawakami, T., Smeds, L., Backström, N., Husby, A., Qvarnström, A., Mugal, C.F., Olason, P. and Ellegren, H. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Molecular Ecology. 23(16), pp.4035–4058.

Keulartz, J. 2015. Captivity for Conservation? Zoos at a Crossroads. Journal of Agricultural and Environmental Ethics. 28(2), pp.335–351.

Khan, A., Patel, K., Shukla, H., Viswanathan, A., van der Valk, T., Borthakur, U., Nigam, P., Zachariah, A., Jhala, Y. V, Kardos, M. and Ramakrishnan, U. 2021. Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. Proceedings of the National Academy of Sciences. 118(49), p.e2023018118.

Kimura, M. 1968. Evolutionary Rate at the Molecular Level. Nature. 217(5129), pp.624–626.

Kimura, M. 1991. The neutral theory of molecular evolution: A review of recent evidence. The Japanese Journal of Genetics. 66(4), pp.367–386.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M. and Shendure, J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics. 46(3), pp.310–315.

Kisling, V.N. (ed.). 2000. Zoo and Aquarium History. CRC Press.

Korneliussen, T.S., Albrechtsen, A. and Nielsen, R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 15(1), p.356.

Korneliussen, T.S. and Moltke, I. 2015. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. Bioinformatics. 31(24), pp.4009–4011.

Kottler, E.J., Dickman, E.E., Sexton, J.P., Emery, N.C. and Franks, S.J. 2021. Draining the Swamping Hypothesis: Little Evidence that Gene Flow Reduces Fitness at Range Edges. Trends in Ecology & Evolution. 36(6), pp.533–544.

Kreger, M. and Hutchins, M. 2010. Ethics of keeping mammals in zoos and aquariums. In:, pp.3–10.

Kubelka, V., Sandercock, B.K., Székely, T. and Freckleton, R.P. 2022. Animal migration to northern latitudes: environmental changes and increasing threats. Trends in Ecology & Evolution. 37(1), pp.30–41.

Kyriazis, C.C., Wayne, R.K. and Lohmueller, K.E. 2021. Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. Evolution Letters. 5(1), pp.33–47.

Lacy, R.C., Brookfield, Jpp.C.Z.S., IL and 2013 2019. Vortex: A stochastic simulation of the extinction process. Version 10.0.

Landrum, M.J., Chitipiralla, S., Brown, G.R., Chen, C., Gu, B., Hart, J., Hoffman, D., Jang, W., Kaur, K., Liu, C., Lyoshin, V., Maddipatla, Z., Maiti, R., Mitchell, J., O'Leary, N., Riley, G.R., Shi, W., Zhou, G., Schneider, V., Maglott, D., Holmes, J.B. and Kattman, B.L. 2020. ClinVar: improvements to accessing data. Nucleic Acids Research. 48(D1), pp.D835–D844.

Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M. and Maglott, D.R. 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Research. 42(D1), pp.D980–D985.

Langhammer, P.F., Bull, J.W., Bicknell, J.E., Oakley, J.L., Brown, M.H., Bruford, M.W., Butchart, S.H.M., Carr, J.A., Church, D., Cooney, R., Cutajar, S., Foden, W., Foster, M.N., Gascon, C., Geldmann, J., Genovesi, P., Hoffmann, M., Howard-McCombe, J., Lewis, T., Macfarlane, N.B.W., Melvin, Z.E., Merizalde, R.S., Morehouse, M.G., Pagad, S., Polidoro, B., Sechrest, W., Segelbacher, G., Smith, K.G., Steadman, J., Strongin, K., Williams, J., Woodley, S. and Brooks, T.M. 2024. The positive impact of conservation action. Science. 384(6694), pp.453–458.

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T. and Carey, V.J. 2013. Software for Computing and Annotating Genomic Ranges. PLoS Computational Biology. 9(8), p.e1003118.

Lee, A.P. and Venkatesh, B. 2013. Ultraconserved Elements (UCEs) in the Human Genome In: Encyclopedia of Life Sciences [Online]. Available from: https://doi.org/10.1002/9780470015902.a0020842.pub2.

Leung, Y.Y., Naj, A.C., Chou, Y.-F., Valladares, O., Schmidt, M., Hamilton-Nelson, K., Wheeler, N., Lin, H., Gangadharan, P., Qu, L., Clark, K., Kuzma, A.B., Lee, W.-P., Cantwell, L., Nicaretta, H., van der Lee, Sven, English, A., Kalra, D., Muzny, D., Skinner,

E., Doddapeneni, H., Dinh, H., Hu, J., Santibanez, J., Jayaseelan, J., Worley, K., Gibbs, R.A., Lee, Sandra, Dugan-Perez, S., Korchina, V., Nasser, W., Liu, X., Han, Y., Zhu, Y., Liu, Y., Khan, Z., Zhu, C., Sun, F.J., Jun, G.R., Chung, J., Farrell, J., Zhang, X., Banks, E., Gupta, N., Gabriel, S., Butkiewicz, M., Benchek, P., Smieszek, S., Song, Y., Vardarajan, B., Reitz, C., Reyes-Dumeyer, D., Tosto, G., De Jager, P.L., Barral, S., Ma, Y., Beiser, A., Liu, C.T., Dupuis, J., Lunetta, K., Cupples, L.A., Choi, S.H., Chen, Y., Mez, J., Vanderspek, A., Ikram, M.A., Ahmad, S., Faber, K., Foroud, T., Mlynarski, E., Schmidt, H., Schmidt, R., Kunkle, B., Rajabli, F., Beecham, G., Vance, J.M., Adams, L.D., Cuccaro, M., Mena, P., Booth, B.M., Renton, A., Goate, A., Marcora, E., Stine, A., Feolo, M., Launer, L.J., Koboldt, D.C., Wilson, R.K., van Duijn, C., Amin, N., Kapoor, M., Salerno, W., Bennett, D.A., Xia, L.C., Malamon, J., Mosley, T.H., Satizabal, C., Bressler, J., Jian, X., Nato, A.Q., Horimoto, A.R., Wang, B., Psaty, B., Witten, D., Tsuang, D., Blue, E., Wijsman, E., Sohi, H., Nguyen, H., Bis, J.C., Rice, K., Brown, L., Dorschner, M., Saad, M., Navas, P., Nafikov, R., Thornton, T., Day, T., Haut, J., Sha, J., Zhang, N., Iqbal, T., Zhao, Y., Below, J.E., Larson, D.E., Appelbaum, E., Waligorski, J., Antonacci-Fulton, L., Fulton, R.S., Haines, J., Farrer, L., Seshadri, S., Brkanac, Z., Cruchaga, C., Pericak-Vance, M., Mayeux, R.P., Bush, W.S., Destefano, A., Martin, E., Schellenberg, G.D., Wang, L.-S. and Project, A.D.S. 2024. Human whole-exome genotype data for Alzheimer's disease. Nature Communications. 15(1), p.684.

Levy, S.E. and Boone, B.E. 2019. Next-generation sequencing strategies. Cold Spring Harbor perspectives in medicine. 9(7), p.a025791.

Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27(21), pp.2987–2993.

Li, H. and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). 25(14), pp.1754–60.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Subgroup, 1000 Genome Project Data Processing 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16), pp.2078–2079.

Lindholm, J. 2013. 4. Zoo History In: M. D. Irwin, J. B. Stoner and A. M. Cobaugh, eds. An Introduction to the Science and Technology [Online]. University of Chicago Press, pp.31–42. Available from: https://doi.org/10.7208/9780226925325-005.

Lott, M.J., Wright, B.R., Kemp, L.F., Johnson, R.N. and Hogg, C.J. 2020. Genetic Management of Captive and Reintroduced Bilby Populations. The Journal of Wildlife Management. 84(1), pp.20–32.

Luu, P.-L., Ong, P.-T., Dinh, T.-P. and Clark, S.J. 2020. Benchmark study comparing liftover tools for genome conversion of epigenome sequencing data. NAR Genomics and Bioinformatics. 2(3), p.lqaa054.

Lynch, M., Ackerman, M.S., Gout, J.-F., Long, H., Sung, W., Thomas, W.K. and Foster, P.L. 2016. Genetic drift, selection and the evolution of the mutation rate. Nature Reviews Genetics. 17(11), pp.704–714.

Lynch, M., Blanchard, J., Houle, D., Kibota, T., Schultz, S., Vassilieva, L. and Willis, J. 1999. Perspective: Spontaneous Deleterious Mutation. Evolution. 53(3), pp.645–663.

Lynch, M., Conery, J. and Bürger, R. 1995. Mutational Meltdowns In Sexual Populations. Evolution. 49(6), pp.1067–1080.

Magro Moraes, A., de Souza Lima, J., Rocha Alexandre, B., Ayala-Burbano, P.A., de Freitas, P.D., Ruiz-Miranda, C.R. and Miyaki, C.Y. 2023. Genetic Management Applied to Conservation of Reduced and Fragmented Wild Populations In: P. M. Galetti Jr., ed. Conservation Genetics in the Neotropics [Online]. Cham: Springer International Publishing, pp.227–249. Available from: https://doi.org/10.1007/978-3-031-34854-9_10.

Marcinkowska, U.M., Moore, F.R. and Rantala, M.J. 2013. An experimental test of the Westermarck effect: sex differences in inbreeding avoidance. Behavioral Ecology. 24(4), pp.842–845.

Marino, L., Rose, N.A., Visser, I.N., Rally, H., Ferdowsian, H. and Slootsky, V. 2020. The harmful effects of captivity and chronic stress on the well-being of orcas (Orcinus orca). Journal of Veterinary Behavior. 35, pp.69–82.

Marks, P., Garcia, S., Barrio, A.M., Belhocine, K., Bernate, J., Bharadwaj, R., Bjornson, K., Catalanotti, C., Delaney, J., Fehr, A., Fiddes, I.T., Galvin, B., Heaton, H., Herschleb, J., Hindson, C., Holt, E., Jabara, C.B., Jett, S., Keivanfar, N., Kyriazopoulou-Panagiotopoulou, S., Lek, M., Lin, B., Lowe, A., Mahamdallie, S., Maheshwari, S., Makarewicz, T., Marshall, J., Meschi, F., O'Keefe, C.J., Ordonez, H., Patel, P., Price, A., Royall, A., Ruark, E., Seal, S., Schnall-Levin, M., Shah, P., Stafford, D., Williams, S., Wu, I., Xu, A.W., Rahman, N., MacArthur, D. and Church, D.M. 2019. Resolving the full spectrum of human genome variation using Linked-Reads. Genome Research. 29(4), pp.635–645.

Mathur, S. and DeWoody, J.A. 2021. Genetic load has potential in large populations but is realized in small inbred populations. Evolutionary Applications. 14(6), pp.1540–1557.

Mathur, S., Mason, A.J., Bradburd, G.S. and 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. 120(43), p.e2303043120.

Maxwell, S.L., Fuller, R.A., Brooks, T.M. and Watson, J.E.M. 2016. Biodiversity: The ravages of guns, nets and bulldozers. Nature. 536(7615), pp.143–145.

McGowan, P.J.K., Traylor-Holzer, K. and Leus, K. 2017. IUCN Guidelines for Determining When and How Ex Situ Management Should Be Used in Species Conservation. Conservation Letters. 10(3), pp.361–366.

Mellor, D.J., Hunt, S. and Gusset, M. 2015. Caring for wildlife: The world zoo and aquarium animal welfare strategy. WAZA Executive Office: Gland, Switzerland.

Miranda, R., Escribano, N., Casas, M., Pino-del-Carpio, A. and Villarroya, A. 2023. The Role of Zoos and Aquariums in a Changing World. Annual Review of Animal Biosciences. 11(1), pp.287–306.

Moerman, F., Fronhofer, E.A., Wagner, A. and Altermatt, F. 2020. Gene swamping alters evolution during range expansions in the protist Tetrahymena thermophila. Biology Letters. 16(6), p.20200244.

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. and Köster, J. 2021. Sustainable data analysis with Snakemake. F1000Research. 10, p.33.

Mooney, A., Conde, D.A., Healy, K. and Buckley, Y.M. 2020. A system wide approach to managing zoo collections for visitor attendance and in situ conservation. Nature Communications. 11(1), p.584.

Morales, H.E., van Oosterhout, C., Whitford, H., Tatayah, V., Ruhomaun, K., Groombridge, J.J. and Gilbert, M.T.P. 2024. The genome sequence of the Pink Pigeon, Nesoenas mayeri (Prévost, 1843). Wellcome Open Research. 9, p.336.

Morgan, K.N. and Tromborg, C.T. 2007. Sources of stress in captivity. Applied Animal Behaviour Science. 102(3–4), pp.262–302.

Morton, N.E., Crow, J.F. and Muller, H.J. 1956. An estimate of the mutational damage in man from data on consanguineous marriages. Proceedings of the National Academy of Sciences of the United States of America. 42(11), pp.855–863.

Mueller, J.C., Kuhl, H., Boerno, S., Tella, J.L., Carrete, M. and Kempenaers, B. 2018. Evolution of genomic variation in the burrowing owl in response to recent colonization of urban areas. Proceedings of the Royal Society B: Biological Sciences. 285(1878), p.20180206.

Muller, H.J. 1950. Our load of mutations. Am J Hum Genetics. 2(2), pp.111–176.

Muller, H.J. 1964. The relation of recombination to mutational advance. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 1(1), pp.2–9.

Murray, G.G.R., Soares, A.E.R., Novak, B.J., Schaefer, N.K., Cahill, J.A., Baker, A.J., Demboski, J.R., Doll, A., Da Fonseca, R.R., Fulton, T.L., Gilbert, M.T.P., Heintzman, P.D., Letts, B., McIntosh, G., O'Connell, B.L., Peck, M., Pipes, M.-L., Rice, E.S., Santos, K.M., Sohrweide, A.G., Vohr, S.H., Corbett-Detig, R.B., Green, R.E. and Shapiro, B. 2017. Natural selection shaped the rise and fall of passenger pigeon genomic diversity. Science. 358(6365), pp.951–954.

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C. and Durbin, R. 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics. 32(11), pp.1749–1751.

Nichols, S., Ewen, J.G., Gottelli, D., Grueber, C.E., Santure, A.W., Trask, A. and Brekke, P. 2024. Genetic rescue attempt in a small, inbred population of a wild endangered passerine. Biological Conservation. 290, p.110430.

Norscia, I., Hecker, M., Caselli, M., Collarini, E., Gallego Aldama, B., Borragán Santos, S. and Cordoni, G. 2024. Social play in African savannah elephants may inform selection against aggression. Current Zoology.

Novak, B.J., Gober, P., Bortner, R., Garelle, D., Wright, M., Chuven, J., Houck, M.L., Ryder, O.A., Milutinovich, D., Benavidez, J., Ryan, K., Walker, S., Arenivas, S.S., Aston, L., Russell, B., Marinari, P., Crosier, A., Helmick, K., Gibson, M.R., Carlsen, D.P., Swanson, B.J., Wisely, S.M., White, Z.S., Lynch, C. and Phelan, R. 2024. First endangered black-footed ferrets, Mustela nigripes, cloned for genetic rescue. bioRxiv., 2024.04.17.589896.

Ohta, T. 2000. Mechanisms of molecular evolution. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. 355(1403), pp.1623–1626.

van Oosterhout, C. 2023. AI-informed conservation genomics. Heredity.

van Oosterhout, C. 2020. Mutation load is the spectre of species conservation. Nature Ecology & Evolution. 4(8), pp.1004–1006.

van Oosterhout, C., Marcu, D. and Immler, S. 2022. Accounting for the genetic load in assisted reproductive technology. Clinical and Translational Medicine. 12(5).

van Oosterhout, C., Speak, S.A., Birley, T., Bortoluzzi, C., Percival-Alwyn, L., Urban, L.H., Groombridge, J.J., Segelbacher, G. and Morales, H.E. 2022. Genomic erosion in the assessment of species extinction risk and recovery potential. bioRxiv., 2022.09.13.507768.

O'Regan, H.J. and Kitchener, A.C. 2005. The effects of captivity on the morphology of captive, domesticated and feral mammals. Mammal Review. 35(3/4), pp.215–230.

O'Regan, H.J., Turner, A. and Sabin, R. 2006. Medieval big cat remains from the Royal Menagerie at the Tower of London. International Journal of Osteoarchaeology. 16(5), pp.385–394.

Owen, A., Wilkinson, R. and Sözer, R. 2014. In situ conservation breeding and the role of zoological institutions and private breeders in the recovery of highly endangered Indonesian passerine birds. International Zoo Yearbook. 48(1), pp.199–211.

Parks, M. and Lambert, D. 2015. Impacts of low coverage depths and post-mortem DNA damage on variant calling: a simulation study. BMC Genomics. 16(1), p.19.

Parnell, G. 1999. The royal menagerie at the Tower of London [Online]. Royal Armouries Museum. Available from: https://books.google.co.uk/books?id=1G5qAAAACAAJ.

Pérez-Pereira, N., Caballero, A. and García-Dorado, A. 2022. Reviewing the consequences of genetic purging on the success of rescue programs. Conservation Genetics. 23(1), pp.1–17.

Pierce, J. and Bekoff, M. 2018. A Postzoo Future: Why Welfare Fails Animals in Zoos. Journal of Applied Animal Welfare Science. 21(sup1), pp.43–48.

Pinto, A. V., Hansson, B., Patramanis, I., Morales, H.E. and van Oosterhout, C. 2024. The impact of habitat loss and population fragmentation on genomic erosion. Conservation Genetics. 25(1), pp.49–57.

Pinto, A. V, Hansson, B., Patramanis, I., Morales, H.E. and van Oosterhout, C. 2023. The impact of habitat loss and population fragmentation on genomic erosion. Conservation Genetics.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. and Siepel, A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Research. 20(1), pp.110–121.

Proctor R N 1988. Racial Hygiene: Medicine Under the Nazis. Harvard University Press, Cambridge, MA.

Proops, L., Burden, F. and Osthaus, B. 2009. Mule cognition: a case of hybrid vigour? Animal Cognition. 12(1), pp.75–84.

Prüfer, K. 2018. snpAD: an ancient DNA genotype caller. Bioinformatics. 34(24), pp.4165–4171.

Quinlan, A.R. and Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26(6), pp.841–842.

Rabier, R., Erlichman, A., Lesobre, L. and Robert, A. 2022. The necessity of considering founder kinships in conservation breeding programs. Animal Conservation. 25(6), pp.759–770.

Rabier, R., Robert, A., Lacroix, F. and 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), pp.422–435.

Rainer, J., Taliun, D., D'Elia, Y., Pattaro, C., Domingues, F.S. and Weichenberger, C.X. 2016. FamAgg: an R package to evaluate familial aggregation of traits in large pedigrees. Bioinformatics. 32(10), pp.1583–1585.

Ralls, K., Ballou, J.D., Dudash, M.R., Eldridge, M.D.B., Fenster, C.B., Lacy, R.C., Sunnucks, P. and Frankham, R. 2018. Call for a Paradigm Shift in the Genetic Management of Fragmented Populations. Conservation Letters. 11(2).

Ralls, K., Sunnucks, P., Lacy, R.C. and Frankham, R. 2020. Genetic rescue: A critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful genetic variation. Biological Conservation. 251, p.108784.

Rentzsch, P., Schubach, M., Shendure, J. and Kircher, M. 2021. CADD-Splice—improving genome-wide variant effect prediction using deep learning-derived splice scores. Genome Medicine. 13(1), p.31.

Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. and Kircher, M. 2019. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Research. 47(D1), pp.D886–D894.

Robinson, J.A., Räikkönen, J., Vucetich, L.M., Vucetich, J.A., Peterson, R.O., Lohmueller, K.E. and Wayne, R.K. 2019. Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. Science Advances. 5(5).

Robinson, Z.L., Bell, D.A., Dhendup, T., Luikart, G., Whiteley, A.R. and Kardos, M. 2021. Evaluating the outcomes of genetic rescue attempts. Conservation Biology. 35(2), pp.666–677.

Roe, K., McConney, A. and Mansfield, C.F. 2014. The Role of Zoos in Modern Society—A Comparison of Zoos' Reported Priorities and What Visitors Believe They Should Be. Anthrozoös. 27(4), pp.529–541.

Rollinson, N., Keith, D.M., Houde, A.L.S., Debes, P. V., McBride, M.C. and Hutchings, J.A. 2014. Risk Assessment of Inbreeding and Outbreeding Depression in a Captive-Breeding Program. Conservation Biology. 28(2), pp.529–540.

Rose, P., Reed, A., Hurn, S., Badman-King, A. and Rice, T. 2022. Does the sound environment influence the behaviour of zoo-housed birds? A preliminary investigation of ten species across two zoos. Behavioural Processes. 203, p.104763.

Rundle, H.D., Nagel, L., Boughman, J.W. and Schluter, D. 2000. Natural Selection and Parallel Speciation in Sympatric Sticklebacks. Science. 287(5451), pp.306–308.

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.

Saccheri, I. and Hanski, I. 2006. Natural selection and population dynamics. Trends in Ecology & Evolution. 21(6), pp.341–347.

Sachdeva, H., Olusanya, O. and 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), p.20210010.

Samuel, M.D., Liao, W., Atkinson, C.T. and LaPointe, D.A. 2020. Facilitated adaptation for conservation – Can gene editing save Hawaii's endangered birds from climate driven avian malaria? Biological Conservation. 241, p.108390.

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R.P., Banday, S., Mishra, A.K., Das, G. and Malonia, S.K. 2023. Next-Generation Sequencing Technology: Current Trends and Advancements. Biology. 12(7), p.997.

Schulte-Hostedde, A.I. and Mastromonaco, G.F. 2015. Integrating evolution in the management of captive zoo populations. Evolutionary Applications. 8(5), pp.413–422.

Seidl, A., Mulungu, K., Arlaud, M., van den Heuvel, O. and Riva, M. 2020. Finance for nature: A global estimate of public biodiversity investments. Ecosystem Services. 46, p.101216.

Seri, N.A. and Rahman, A.A. 2021. Impact of Climate Change on Migratory Birds in Asia. Pertanika Journal of Science & Technology. 29(4).

Shuker, D.M. and Kvarnemo, C. 2021. The definition of sexual selection. Behavioral Ecology. 32(5), pp.781–794.

Silla, T., Kepp, K., Tai, E.S., Goh, L., Davila, S., Ivkovic, T.C., Calin, G.A. and Voorhoeve, P.M. 2014. Allele Frequencies of Variants in Ultra Conserved Elements Identify Selective Pressure on Transcription Factor Binding. PLOS ONE. 9(11), pp.e110692-.

Singh, A. and Bahuguna, A. 2014. Digging out treasure from museum specimen: A comparison of extraction methods and PCR polymerase enzymes for DNA from old taxidermy samples. International Journal of Advanced Research. 2(9), pp.700–707.

Slatkin, M. 1987. Gene Flow and the Geographic Structure of Natural Populations. Science. 236(4803), pp.787–792.

Smeds, L. and Ellegren, H. 2023. From high masked to high realized genetic load in inbred Scandinavian wolves. Molecular Ecology. 32(7), pp.1567–1580.

Smith, E.H. 2019. Species review: Whooping crane (Grus americana). Crane conservation strategy., pp.223–243.

Smith, J.M. and Haigh, J. 1974. The hitch-hiking effect of a favourable gene. Genetical Research. 23(1), pp.23–35.

Smith, W.J. 2023. Rapid assessment of the extent of feral introgression in British and Irish Rock Dove (Columba livia) populations. Ibis. 165(4), pp.1432–1439.

Smith, W.J. and Clegg, S.M. 2023. Hybridisation and genomic diversity in British and Irish pigeons Columba livia. European Journal of Wildlife Research. 69(5), p.94.

Smith, W.J., Quilodrán, C.S., Jezierski, M.T., Sendell-Price, A.T. and Clegg, S.M. 2022. The wild ancestors of domestic animals as a neglected and threatened component of biodiversity. Conservation Biology. 36(3).

Speak, S.A., Birley, T., Bortoluzzi, C., Clark, M.D., Percival-Alwyn, L., Morales, H.E. and van Oosterhout, C. 2024. Genomics-informed captive breeding can reduce inbreeding depression and the genetic load in zoo populations. Molecular Ecology Resources. 24(7), p.e13967.

Spooner, S.L., Walker, S.L., Dowell, S. and Moss, A. 2023. The value of zoos for species and society: The need for a new model. Biological Conservation. 279, p.109925.

Sukumar, S., Krishnan, A. and Banerjee, S. 2021. An Overview of Bioinformatics Resources for SNP Analysis In: V. Singh and A. Kumar, eds. Advances in

Bioinformatics [Online]. Singapore: Springer Singapore, pp.113–135. Available from: https://doi.org/10.1007/978-981-33-6191-1_7.

Swinnerton, K.J., Groombridge, J.J., Jones, C.G., Burn, R.W. and Mungroo, Y. 2004. Inbreeding depression and founder diversity among captive and free-living populations of the endangered pink pigeon Columba mayeri. Animal Conservation. 7(4), pp.353–364.

Tadmouri, G.O., Nair, P., Obeid, T., Al Ali, M.T., Al Khaja, N. and Hamamy, H.A. 2009. Consanguinity and reproductive health among Arabs. Reproductive Health. 6(1), p.17.

Tallmon, D., Luikart, G. and Waples, R. 2004. The alluring simplicity and complex reality of genetic rescue. Trends in Ecology & Evolution. 19(9), pp.489–496.

Tataru, P., Mollion, M., Glémin, S. and Bataillon, T. 2017. Inference of Distribution of Fitness Effects and Proportion of Adaptive Substitutions from Polymorphism Data. Genetics. 207(3), pp.1103–1119.

Teicher, A. 2018. Caution, Overload: The Troubled Past of Genetic Load. Genetics. 210(3), pp.747–755.

Templeton, A.R., Hemmer, H., Mace, G., Seal, U.S., Shields, W.M. and Woodruff, D.S. 1986. Local adaptation, coadaptation, and population boundaries. Zoo Biology. 5(2), pp.115–125.

Thurman, T.J. and Barrett, R.D.H. 2016. The genetic consequences of selection in natural populations. Molecular Ecology. 25(7), pp.1429–1448.

Tizard, I. and Skow, L. 2021. The olfactory system: the remote-sensing arm of the immune system. Animal Health Research Reviews. 22(1), pp.14–25.

Tollington, S., Jones, C.G., Greenwood, A., Tatayah, V., Raisin, C., Burke, T., Dawson, D.A. and Groombridge, J.J. 2013. Long-term, fine-scale temporal patterns of genetic diversity in the restored Mauritius parakeet reveal genetic impacts of management and associated demographic effects on reintroduction programmes. Biological Conservation. 161, pp.28–38.

Trask, A.E., Fenn, S.R., Bignal, E.M., McCracken, D.I., Monaghan, P. and Reid, J.M. 2019. Evaluating the efficacy of independent versus simultaneous management strategies to address ecological and genetic threats to population viability. Journal of Applied Ecology. 56(10), pp.2264–2273.

Traylor-Holzer, K., Leus, K. and Bauman, K. 2019. Integrated Collection Assessment and Planning (ICAP) workshop: Helping zoos move toward the One Plan Approach. Zoo Biology. 38(1), pp.95–105.

Tsai, W.L.E., Schedl, M.E., Maley, J.M. and McCormack, J.E. 2020. More than skin and bones: Comparing extraction methods and alternative sources of DNA from avian museum specimens. Molecular Ecology Resources. 20(5), pp.1220–1227.

Turghan, M.A., Jiang, Z. and Niu, Z. 2022. An Update on Status and Conservation of the Przewalski's Horse (Equus ferus przewalskii): Captive Breeding and Reintroduction Projects. Animals. 12(22), p.3158.

van der Valk, T., de Manuel, M., Marques-Bonet, T. and Guschanski, K. 2021. Estimates of genetic load suggest frequent purging of deleterious alleles in small populations. bioRxiv., p.696831.

Vigeland, M.D. 2020. Relatedness coefficients in pedigrees with inbred founders. Journal of Mathematical Biology. 81(1), pp.185–207.

Visel, A., Prabhakar, S., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Afzal, V., Rubin, E.M. and Pennacchio, L.A. 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nature Genetics. 40(2), pp.158–160.

Wallace, B. 1975. Hard and Soft Selection Revisited. Evolution. 29(3), p.465.

Van de Walle, J., Larue, B., Pigeon, G. and Pelletier, F. 2022. Different proxies, different stories? Imperfect correlations and different determinants of fitness in bighorn sheep. Ecology and Evolution. 12(12), p.e9582.

Waller, D.M. 2015. Genetic rescue: a safe or risky bet? Molecular Ecology. 24(11), pp.2595–2597.

Wang, B., Ekblom, R., Bunikis, I., Siitari, H. and Höglund, J. 2014. Whole genome sequencing of the black grouse (Tetrao tetrix): reference guided assembly suggests faster-Z and MHC evolution. BMC Genomics. 15(1), p.180.

Warnefors, M., Hartmann, B., Thomsen, S. and Alonso, C. 2016. Combinatorial gene regulatory functions underlie ultraconserved elements (UCEs) in Drosophila. Molecular Biology and Evolution. 33, p.msw101.

Warnefors, M., Hartmann, B., Thomsen, S. and Alonso, C.R. 2016. Combinatorial Gene Regulatory Functions Underlie Ultraconserved Elements in Drosophila. Molecular Biology and Evolution. 33(9), pp.2294–2306.

Warr, A., Affara, N., Aken, B., Beiki, H., Bickhart, D.M., Billis, K., Chow, W., Eory, L., Finlayson, H.A., Flicek, P., Girón, C.G., Griffin, D.K., Hall, R., Hannum, G., Hourlier, T., Howe, K., Hume, D.A., Izuogu, O., Kim, K., Koren, S., Liu, H., Manchanda, N., Martin, F.J., Nonneman, D.J., O'Connor, R.E., Phillippy, A.M., Rohrer, G.A., Rosen, B.D., Rund, L.A., Sargent, C.A., Schook, L.B., Schroeder, S.G., Schwartz, A.S., Skinner, B.M., Talbot, R., Tseng, E., Tuggle, C.K., Watson, M., Smith, T.P.L. and Archibald, A.L. 2020. An improved pig reference genome sequence to enable pig genetics and genomics research. GigaScience. 9(6), p.giaa051.

Warren, W.C., Hillier, L.W., Tomlinson, C., Minx, P., Kremitzki, M., Graves, T., Markovic, C., Bouk, N., Pruitt, K.D., Thibaud-Nissen, F. and others 2017. A new chicken genome assembly provides insight into avian genome structure. G3: Genes, Genomes, Genetics. 7(1), pp.109–117.

Wedekind, C., Seebeck, T., Bettens, F. and Paepke, A.J. 1997. MHC-dependent mate preferences in humans. Proceedings of the Royal Society of London. Series B: Biological Sciences. 260(1359), pp.245–249.

Weinberg W 1912. Weitere Beiträge zur Theorie der Vererbung. 4. Ueber Methode und Fehlerquellen der Untersuchung auf Mendelsche Zahlen beim Menschen. Arch. Rass. Ges. Biol. 9, pp.165–174.

Wennerholm, U.-B. and Bergh, C. 2020. Perinatal outcome in children born after assisted reproductive technologies. Upsala Journal of Medical Sciences. 125(2), pp.158–166.

Westermarck, E. 1922. The history of human marriage.

White, S.L., Rash, J.M. and Kazyak, D.C. 2023. Is now the time? Review of genetic rescue as a conservation tool for brook trout. Ecology and Evolution. 13(5).

Whiteley, A.R., Fitzpatrick, S.W., Funk, W.C. and Tallmon, D.A. 2015. Genetic rescue to the rescue. Trends in Ecology & Evolution. 30(1), pp.42–49.

Wilder, A.P., Steiner, C.C., Hendricks, S., Haller, B.C., Kim, C., Korody, M.L. and Ryder, O.A. 2024. Genetic load and viability of a future restored northern white rhino population. Evolutionary Applications. 17(4).

Wilder, A.P., Supple, M.A., Subramanian, A., Mudide, A., Swofford, R., Serres-Armero, A., Steiner, C., Koepfli, K.-P., Genereux, D.P., Karlsson, E.K., Lindblad-Toh, K., Marques-Bonet, T., Munoz Fuentes, V., Foley, K., Meyer, W.K., Ryder, O.A., Shapiro, B., Andrews, G., Armstrong, J.C., Bianchi, M., Birren, B.W., Bredemeyer, K.R., Breit, A.M., Christmas, M.J., Clawson, H., Damas, J., Di Palma, F., Diekhans, M., Dong, M.X., Eizirik, E., Fan, K., Fanter, C., Foley, N.M., Forsberg-Nilsson, K., Garcia, C.J., Gatesy, J., Gazal, S., Genereux, D.P., Goodman, L., Grimshaw, J., Halsey, M.K., Harris, A.J., Hickey, G., Hiller, M., Hindle, A.G., Hubley, R.M., Hughes, G.M., Johnson, J., Juan, D., Kaplow, I.M., Karlsson, E.K., Keough, K.C., Kirilenko, B., Koepfli, K.-P., Korstian, J.M., Kowalczyk, A., Kozyrev, S. V., Lawler, A.J., Lawless, C., Lehmann, T., Levesque, D.L., Lewin, H.A., Li, X., Lind, A., Lindblad-Toh, K., Mackay-Smith, A., Marinescu, V.D., Marques-Bonet, T., Mason, V.C., Meadows, J.R.S., Meyer, W.K., Moore, J.E., Moreira, L.R., Moreno-Santillan, D.D., Morrill, K.M., Muntané, G., Murphy, W.J., Navarro, A., Nweeia, M., Ortmann, S., Osmanski, A., Paten, B., Paulat, N.S., Pfenning, A.R., Phan, B.N., Pollard, K.S., Pratt, H.E., Ray, D.A., Reilly, S.K., Rosen, J.R., Ruf, I., Ryan, L., Ryder, O.A., Sabeti, P.C., Schäffer, D.E., Serres, A., Shapiro, B., Smit, A.F.A., Springer, M., Srinivasan, C., Steiner, C., Storer, J.M., Sullivan, K.A.M., Sullivan, P.F., Sundström, E., Supple, M.A., Swofford, R., Talbot, J.-E., Teeling, E., Turner-Maier, J., Valenzuela, A., Wagner, F., Wallerman, O., Wang, C., Wang, J., Weng, Z., Wilder, A.P., Wirthlin,

M.E., Xue, J.R. and Zhang, X. 2023. The contribution of historical processes to contemporary extinction risk in placental mammals. Science. 380(6643).

Willi, Y., Kristensen, T.N., Sgrò, C.M., Weeks, A.R., Ørsted, M. and Hoffmann, A.A. 2022. Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. Proceedings of the National Academy of Sciences. 119(1), p.e2105076119.

Wood, J.L.A., Yates, M.C. and Fraser, D.J. 2016. Are heritability and selection related to population size in nature? Meta-analysis and conservation implications. Evolutionary Applications. 9(5), pp.640–657.

Wright, D.J., Spurgin, L.G., Collar, N.J., Komdeur, J., Burke, T. and Richardson, D.S. 2014. The impact of translocations on neutral and functional genetic diversity within and among populations of the Seychelles warbler. Molecular Ecology. 23(9), pp.2165–2177.

Wright, S. 1922. Coefficients of Inbreeding and Relationship. The American Naturalist. 56(645), pp.330–338.

Wright, S. 1949. Population Structure in Evolution. Proceedings of the American Philosophical Society. 93(6), pp.471–478.

Wu, M.Y., Chong, H.W., Tan, Y.Y., Lim, B.T.M., Wong, A.M.S., Oh, R.R.Y., Lee, J.G.H., Ng, E.Y.X. and Rheindt, F.E. 2024. Genomic data reveal shift in geographic source of an illegally traded songbird. Conservation Genetics. 25(1), pp.179–187.

Yang, F., Fu, B., O'Brien, P.C.M., Nie, W., Ryder, O.A. and Ferguson-Smith, M.A. 2004. Refined genome-wide comparative map of the domestic horse, donkey and human based on cross-species chromosome painting: insight into the occasional fertility of mules. Chromosome Research. 12(1), pp.65–76.

Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P. and Wang, L. 2014. CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics. 30(7), pp.1006–1007.

Zhou, C., Yu, H., Geng, Y., Liu, W., Zheng, S., Yang, N., Meng, Y., Dou, L., Price, M., Ran, J., Yue, B. and Wu, Y. 2019. A High-quality Draft Genome Assembly of the Black-necked Crane (Grus nigricollis) Based on Nanopore Sequencing. Genome Biology and Evolution.

Zimmermann, J., Hajibabaei, M., Blackburn, D.C., Hanken, J., Cantin, E., Posfai, J. and Evans, T.C. 2008. DNA damage in preserved specimens and tissue samples: a molecular assessment. Frontiers in Zoology. 5(1), p.18.