Better tools, better resources, better conservation: Integrating genome data into the conservation of the pink pigeon *Nesoenas mayeri*

Camilla E. Ryan

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

June 2021

Dedicated with love to Charlie Treadwell, making me a better person since 1991

A juvenile pink pigeon named Camilla, reared by Harri Whitford at Jersey zoo

Acknowledgements

If it takes a village to raise a child it takes a small country to get someone through their PhD.

I am very grateful to my supervisory team: Bernardo Clavijo (EI), Cock Van Oosterhout (UEA), Diana Bell (UEA), Carl Jones (DWCT), Ian Barnes (NHM), Matt Clark (NHM), Lawrence Percial-Alwyn (NIAB). In particular Cock Van Oosterhout for his continuous support and infectious enthusiasm and Bernardo who adopted me when I was a stray, has constantly supported me and inspired me with irrefutable logic. I will always be grateful to the Calvijo group (Ber, Ben, Gon, Jon and Luis) who adopted and potentially indoctrinated me (in a good way) to their way of research, I have become an infinitely better researcher because my position in the A-team. Not not only have they shared their knowledge but to a person have never failed to offer support or help whenever it may be needed.

I have been "visiting EI" for four years and I have enjoyed all of them a big thank you to most of the staff and students at EI. A special thanks has to go to Dr Angela Man who (miraculously) got me up and running and I was all the better for it! And Sasha Stanbridge who has been a great friend, co-conspirator and dog walking companion.

A big thank you to Dr Graham Etherington (EI) who was instrumental in securing the funding to sequence pink pigeon genomes using Chromium 10x technology and also set EI up as a CITES scientific institute. A thank you also to Frederica Di Palma (UEA) who supported this project and enabled the collaboration with Chromium 10x. Finally a thank you to Dr Deanna Church (10x Chromium) and Chromium 10x for supplying the kits, free of charge, that were used to sequence the pink pigeon.

My PhD was part of the EnvEast doctoral training partnership and I will be forever grateful for the opportunities they have given me. The rest of my cohort are an absolute mad bunch of brilliant researchers who have made this process that little bit easier.

I was incredibly fortunate to be able to spend some small part of my PhD working in the group of Klaus Koepfli. Klaus gave me a much needed boost of inspiration and introduced

me to people who have become firm friends. In particular Rebecca Gooley who is always willing to offer support and advice. Prior to my PhD my scientific career was launched from Cardiff University there have been so many people during my Masters, Undergraduate and in between that have taught me and made me feel welcome. Floor 5 - Bruford's group I am very grateful to you all in particular to Mike Bruford. Mike has encouraged me, looked out for me and supported me throughout my time at Cardiff. Mike's support and guidance made a huge difference to me and it is definitely (in large part) responsible for me ending up doing this PhD.

I have been incredibly fortunate to work with and meet some of the most dedicated individuals anyone is likely to come across. This is because my small contribution to the pink pigeon research is only possible because of the 40+ years of research that has come before me. I thank everyone who has contributed to any research I have used or has spoken with me in particular I would like to mention Vikash Tataya, Simon Tollington, Michelle Barrow and Sara Shopland. A special thanks to Harri Whitford (the pink pigeon studbook keeper) and absolutely brilliant person who has given me lots of support and information. Finally a huge thank you to Carl Jones without this brilliant, bird obsessed man there would be no pink pigeons and the world would be infinitely poorer because of it.

I wanted to also thank the original research group who started to shape the researcher I would become Mike Bruford's group from Cardiff University. To all those who got me through my Masters with good cheer and support. I must offer a special thank you to Mike Bruford who has always looked out for me and encouraged me as a researcher.

I don't think anyone could ask for a more supportive, loving (or mad) family. Almost ever Wilson has in their own way offered support and guidance to me at some point in my scientific career. I have been continually supported by Nonna who (I know) thinks I am slightly mad but has been a pillar of support for me my whole life.

My parents have always made sure I know how proud they are of me and done everything that they can to support my career. Mum thank you for the baking, the proofing, and the crazy glue. Dad thank you for the chocolate, coffee and all the notes of support. James you have been an incredible older brother and now you are an incredible father to Luke, the pictures and video calls have kept me going throughout the last days of my writing up. Cress you have been the best big sister always providing advice, support and thoughtful notes. A big thank you to my wonderful niece Fiammetta Lester who reminds me why I love doing what I do. A special mention for my "little" brother Hugh who has never given up on me and helped me more than he can know. I am especially lucky have the most wonderful aunt and uncle, Steph and Richard (aka the coot) they have helped shaped the person I am today I

don't think I can ever express how grateful I am to them for their continued and unwavering love and support.

When I moved to Norwich I found a home and my own family (and menagerie - living the dream). A big thank you to Zelda for all the cuddles and mini-adventures we have every day. I don't think a single day has gone by when you haven't made me smile or chuckle (I still do not understand the rules of the games we play). My greatest and most heartfelt thanks must go to my incredible partner Dr Ward. They have been a cheerleader, a defender, a supporter, a shoulder whenever any was called upon. Despite never being able to get away from me or my PhD they have never complained whenever I have needed help (whatever time of day or night) or failed to volunteer suggestions. They have also been the most incredible mentor and has helped shaped the scientist I have become. I honestly do not know how to thank you or how I would have manged to produce my *magnum opus* without you.

Abstract

Humans are driving the sixth mass extinction causing biodiversity to decline at an unprecedented rate. To halt these declines effective conservation strategies are vital, if possible, these should include genetic data to ensure that the extinction risk of a population is not being underestimated. This thesis presents resources and tools that have been used to integrate genetic data into the management of the endangered pink pigeon (*Nesoenas mayeri*). A pseudo-chromosome assembly enabled the first whole genome analyses of the pink pigeon (Chapter 2), these analyses provided insight into their past demography and revealed a surprisingly large amount of variation within its genome. This challenges previous results obtained using restriction site associated DNA sequencing (RAD-Seq) data from wild pink pigeons but current methods for processing RAD data produce biased data sets that under-estimate diversity. A novel tool, RADiKal is presented (Chapter 3), which extracts information directly from raw RAD reads and avoids biasing data sets with complex parameterisation. The painted chromosomes produced by RADiKal provide an overview about the levels of variation present in the wild population which are comparable to those observed from whole genome analyses. Despite increasing access to genetic data one of the greatest challenges is its integration into management, one solution is the use of population viability analysis (PVA). An updated PVA was produced for pink pigeons showing that without genetic rescue they could face extinction within 100 years (Chapter 4). Selecting which individuals are most valuable for a genetic rescue is not trivial, especially in the absence of empirical data. I Choose You (I.C.Y) is an easy-to-use tool that allows practitioners to select individuals for genetic rescue based on their genetic diversity measured using founder equivalents calculated from studbook data (Chapter 5). Overall this thesis aims to demonstrate how better tools can lead to better resources and better conservation.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Humans are driving the current mass extinction event to a point at which the levels of global biodiversity are declining at an unprecedented rate [14]. Historically the loss of global biodiversity has been correlated to the arrival of humans and their over exploitation of local resources, more recently these pressures have been aggravated by the human-mediated spread of pathogens and alien species [15]. Because of this many species currently exist as small fragmented populations suffering from degraded habitat, invasive species and reduced genetic diversity [16, 17]. Reduced genetic diversity together with ecological drivers act synergistically creating a positive feedback loop, known as an extinction vortex (see Figure 1.1) [18], this places further pressure on endangered species and significantly increases their risk of extinction [19]. In contrast to these dire statements we are also living through a period of rapid technological innovation providing us with many of the tools we need to understand, reduce and reverse the decline in global biodiversity [20, 21].

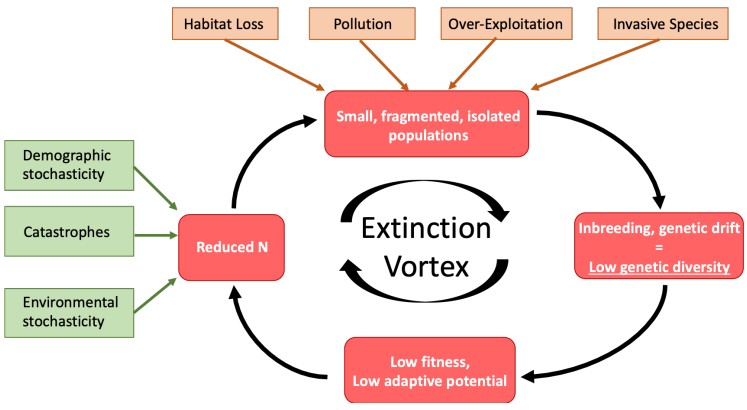


Figure 1.1 The extinction vortex describes a positive feedback loop whereby population size diminishes ultimately resulting in extinction. Natural impacts are seen in green boxes, anthropogenic influences in yellow boxes and the results of these processes are seen in red.

Modern efforts in conservation are now being helped by one of these rapidly developing technologies - genomics. Genomics is the application of large amounts of DNA sequence data to a variety of biological questions and has shown obvious promise in agriculture (breeding livestock adapted for climate change [22] and medicine (personalised and precision medicine see [23]) and we are now seeing examples of its functionality in conservation (for detailed list see supplementary material to [24]). However new technologies present several challenges, for example understanding the theory behind, and learning to use, the new technology and its associated tools properly and assessing whether they are relevant; new does not necessarily mean better [25]. These challenges can slow the integration of new technologies into areas with less funding and with high risk outcomes [26] such as conservation, where an incorrect management decision for a small population could lead to its extinction [27]. This integration is further impeded in conservation by the more immediate threats that often face small populations such as habitat destruction, poaching and disease [28].

1.1 The genetics of small populations

It is intuitive as to why habitat loss or poaching negatively impact a population but more complex to understand are the genetic dynamics of a small population which contribute to their decline as part of the extinction vortex. Some of the major genetic contributors to the extinction vortex are described as Allee effects. These are phenomena where the fitness of a population is associated with its size and below a critical density of individuals the population will lose fitness [29]. This loss in fitness does not necessarily mean the growth rate of the population becomes negative, however strong Allee effects will result in a negative growth rate and increased extinction risk [30]. The critical threshold below which Allee effects become important will depend on the cause or causes of the population decline which will lead to different Allee effects and are not limited to genetic impacts. For example populations can experience Allee effects that impact reproduction (example: mate-finding becomes harder in smaller populations), survival (example: predator dilution; fewer individuals in populations the higher the chance a particular individual is targeted) and a population may suffer from multiple Allee effects [29].

Genetic Allee effects can be attributed to two associated mechanisms, inbreeding and genetic drift [30]. Genetic drift is arguably the most important evolutionary force acting upon a population with a small effective population size [31, 32] and describes the stochastic process which defines the distribution of allele frequencies. With a low effective population size drift can cause deleterious alleles to accumulate within the genepool and drift towards fixation [32]. Drift also leads to an increase in homozygosity as it decreases the number of

alleles available within the genepool. Increased homozygosity within a population reduces the standing genetic variation for selection thereby making small populations less able to adapt and therefore less resilient to future changes [33]. Homozygosity is further increased by inbreeding whereby consanguineous mating increases the number of alleles that are identical by descent within a population. This leads to inbreeding depression where offspring are less fit than their parents and exhibit negative phenotypic traits. These can be shown as physical traits, such as the kinked tail of the Florida Panther [34] or can reduce longevity and/or fecundity [35]. Inbreeding depression is caused by two mechanisms, one is the loss of heterozygote advantage with increasing homozygosity in a population [36] and the the other is the expression of recessive deleterious alleles [37]. The accumulation of deleterious alleles in an individual is known as its genetic load and can be measured using lethal equivalents. The number of lethal equivalents in an organism represents a group of alleles which, if made fully homozygous, would be lethal [38, 37], for example an individual may have two lethal alleles or four alleles each of which causes a 50% reduction in survival. Lethal equivalents are calculated using a regression analysis (see Equation 1.1) of inbreeding versus a fitness trait, where **S** is the chosen trait **A** is the intercept, **F** is the inbreeding coefficient and β is the number of lethal equivalents in a haploid.

$$ln(S) = A^{-\beta F} \tag{1.1}$$

This calculation requires access to both pedigree and life history data which is challenging to gather for wild species and therefore most estimates are based on well studied systems (e.g. flies, mice, great tits) or captive animals [38, 39].

Whilst it is important to deal with the immediate threats faced by a declining population it is obvious that genetic factors will influence the fitness and relative resilience of a population and therefore must be considered as part of an informed management plan [40].

1.2 Conservation genetics & genomics

The concept of 'genetic conservation' - conserving genetic polymorphism - gained popularity in the 1970s because even before the advent of whole genome sequencing it was obvious to researchers that genetic variation was vital for the long term survival of a species [41, 42]. The study of genetic variation is still a major theme in conservation genetics but the field has expanded to include studies on inbreeding, adaptive potential, wildlife forensics, species identification and the delineation of conservation management units [40, 43, 44]. Table 1.1 shows examples where each of these have been successfully incorporated into management

decisions. Despite these successes there still remain challenges within conservation genetics and many researchers and practitioners believe there is a gap between the academic research generated in this field, its practical application and effective policy [45, 26, 46, 20, 47]. Although recent studies have shown that practitioners do think genetics is important they do not necessarily consider it important for every species. This is either because other threats are more immediate or because they have seen a species recover from extreme genetic bottlenecks with no genetic intervention [24, 48]. In reality the species cited such as Chatham Island Black Robin (*Petroica traversi*) or the Little Spotted Kiwi(*Apteryx owenii*) both suffer from inbreeding depression [48] but perhaps the main reason for the gap between academic research and practical application can be described as a lack of clear communication between academics and practitioners. The theory underpinning conservation genetics has its origins in population genetics which is complex, can be difficult to understand and is filled with academic jargon [26].

Another criticism of conservation genetic studies is that they use a small number of genetic markers (10s or 100s) compared to the number used in genomic studies (1000s) and that this limits the statistical power researchers have to accurately describe a study population [49, 50]. Furthermore genetic markers such as microsatellites are assumed to be evolving under neutral processes (genetic drift) and therefore are not suitable for studying adaptation and selection [25]. Although unsuitable for studying selection neutral genetic markers (such as microsatellites) can and have been used effectively in conservation (see Table 1.1) to study populations of endangered species where genetic drift is considered the predominant evolutionary force [11]. Specifically neutral markers (including microsatellite, amplified fragment length polymorphism (AFLP) and restriction fragment length polymorphisms (RFLP) markers) have been used to study heterozygosity, phylogeny, rates of gene flow between populations and population structure [51]. Neutrally evolving markers, like microsatellites, are also better suited to identifying certain population processes such as growth, contractions and divergence because neutral markers accumulate mutations at a faster rate than genetic markers under selection and can provide a more distinct measure of differentiation between populations [52].

The utility of neutral genetic markers has been increased by the use of Next Generation Sequencing (NGS) which enables researchers to generate thousands of markers throughout the genome. This not only led to an increase in statistical power which enabled researchers to make more robust inferences about population structure [50, 53] but also meant markers that were under selection could be generated and used to study functional diversity and adaptive potential [54]. To understand a population's ability to adapt and preserve their potential for adaptation in the future, functional diversity needs to be identified and conserved which requires characterising and studying loci under selection [55, 20]. This is an important

consideration for conservation projects because as populations are threatened they have three options dispersal, plasticity and adaptation [52], if a population is is unable to disperse and the changes needed to survive exceed the plastic abilities of the species then they will need to adapt [52, 20]. If the population is unable to adapt it will go extinct therefore it is vital to be able to characterise a populations functional diversity and adaptive potential to ensure that conservation projects creates healthy self-sustaining populations [20].

Despite the difference in the number of markers used in different studies, ranging from tens to millions, there is not a strict definition separating conservation genomics from genetics [40] but for the purposes of this thesis, unless otherwise stated, I consider any study using microsatellites, single loci or fewer than 1000 loci to be genetics and everything else genomics.

Table 1.1 Examples of how genetics has been successfully used as part of conservation managements

Example	**Molecular markers**	**Outcome**	**Ref**
Used microsatellites to accurately estimate relatedness, then assign breeding pairs based on this estimate in a captive population of Attwater's Prairie-Chicken *Tympanuchus cupido*	9/10 microsatellites	Decreased inbreeding and increased chick survival	[56]
Used eDNA to detect the presence of the American Bullfrog *Lithobates catesbeianus* at low densities in France where it is an invasive species so that effective management and eradication regimes could be implemented to protect the native wildlife	79bp of mitochondrial cytochrome *b*	American Bullfrogs were identified by eDNA in ponds deemed clear by traditional survey techniques	[57]
Identified the origin of sashimi being sold as legal whale products as belonging to species illegal to hunt. They also discovered a possible link with illegal trade in whale meat between Japan and Korea and therefore an infraction of the CITES agreement	Mitochondrial cytochrome *b* and control region sequences	Handed information over to the relevant authorities with some leading to criminal trials	[58]

Examples of how genetics has been successfully used as part of conservation managements (continued)

Example	**Molecular markers**	**Outcome**	**Ref**
Used genetic and ecological data together to show that the Massasauga Rattlesnake (*Sistrurus catenatus*) was vulnerable to losing adaptive diversity over time unless populations across three distinct regions were maintained	Microsatellites	Evidence for threatened status under the Endangered Species Act (ESA) and recommendation to conserve the rattlesnake over the three distinct regions	[59]
Molecular data indicated that the population of Amazon River Dolphin *Inia sp.* found in Bolivia was significantly diverged from other populations living in the Orinoco and Amazon rivers and represented a distinct evolutionary and management unit	400bp mitochondrial cytochrome *b* and 600bp mitochondrial control region	Call to review the taxonomy of the pink river dolphin in case one unit was in fact a cryptic species and increased conservation management	[60, 61].

1.2.1 Working with genomic data

Using genomic rather than genetic data, researchers are capable of answering questions about genetic diversity and neutral evolutionary forces with greater precision [21, 40] and can examine how these shaped different parts of the genome [49, 62]. In particular the inclusion of non-neutral loci has enabled some researchers to identify loci with adaptive potential and study the processes of natural selection and adaptation in endangered species [63, 44]. The study of adaptive potential and functional diversity describes the identification and characterisation of functional loci that, together with gene-environment interactions, phenotypic plasticity, epigenetic effects and natural selection allow a population to respond to changes in their environment [20]. This is important for endangered species given that the ultimate goal of conservation is to create a self sustaining population therefore a population needs to be capable of adapting to future challenges [64–66]. Genetic diversity at neutral loci could be considered a proxy for functional diversity and adaptive potential [46], a population with low heterozygosity will have limited scope for selection and therefore is likely to have limited adaptive potential. The assumption then is that the diversity at neutral loci is mirrored by that at functional loci however this relationship has not always been observed therefore to answer questions about adaptive potential, genome-wide data, which includes functional (or functionally linked) loci, are the most accurate [67, 25].

Despite the advantages of genomic compared to genetic data, including the increased precision and statistical power of analyses and breadth of questions researchers can ask [68, 40], there are disadvantages and challenges. The main disadvantage is that the generation of genomic data is still too expensive for many projects [28, 48]. This is a major consideration for conservation projects who typically struggle to find funding [26] and whose questions could be answered with the use of genetic markers [25] or the complete absence of genetic analyses - no amount of genetic data will stop poaching. However the price of genome sequencing is continually decreasing and is becoming comparable to genetic sequencing which will make it more accessible to projects with limited funding [26, 40, 63].

One of the reasons for the relatively high cost of genomic projects is the extensive bioinformatic post processing that is required to generate useable data from raw sequences [69]. NGS technology produces a huge number of raw sequences but interpreting this data is not trivial requiring access to high performance computing facilities and knowledge of the command line to run many different computational tools which are often designed using complex algorithms [63]. The development and quality control of these tools is often taken for granted [70] with papers describing the evolution of the sequencing machines in detail and simply stating that scientists now need to understand bioinformatics [71]. In reality selecting, understanding and correctly parameterising the appropriate software are crucial steps to

ensure that the final data set is accurate and not full of processing or sequencing artefacts. Sequencing artefacts tend to be well characterised for example it is widely known that fragment length is related to the quality of the Read-Twos in Illumina paired-end sequencing [72] but more complicated to characterise are processing artefacts - features internal to the software's algorithm or to the data processing strategy that will alter the data so that it becomes a reflection of the processing rather than biology. For example multiple studies examining the impact of parameterisation on the generation of Restriction Site Associated DNA (RAD) loci have shown that the number, polymorphism and even genetic differentiation of the loci recovered can be significantly altered by the parameter set used [73–75]. This problem is compounded by the fact that there are no tools that can assess definitively which set of loci are produced from RAD sites instead the results of the data-processing are often validated by prior assumptions. This is not a sufficient assessment and leads to biased data sets that may not adequately reflect the biology of the study species and this could lead to incorrect conclusions and management decisions.

The need for these complex tools to interpret genomic data also contributes to the "practitioner-academic" gap because of the resources and expertise required, this highlights the need for close collaboration between conservation practitioners and academics [48]. Practitioners should not have to understand the complex algorithms and parameterisation needed to produce the data but the academic who runs the analysis should understand it. Furthermore that academic should be able to effectively communicate all the necessary information to practitioners, including the uncertainties in the results, so that they can correctly interpret the data [48, 76]. For effective conservation this communication must be two-way, just as practitioners are not necessarily bioinformatics experts, bioinformaticians are unlikely to have an expert level of understanding about the study species. Practitioners, however, are usually experts in their study species and population biology therefore they will be able to guide bioinformatic analyses to ensure the data generated by academics is relevant and contributes to the management of the study species [26]. Furthermore practitioners are uniquely placed to highlight any population or species specific traits that could otherwise confound bioinformatic analyses. Ultimately management decisions rely on data therefore a thorough understanding of how that data is produced and processed is vital to ensure that the data used is high quality and accurate.

Another challenge is that the successful implementation of genomic methods into conservation management does not have the same proven track record as genetic methods [28]. The importance of considering genetic diversity in endangered populations has been acknowledged since the 1970's however genomics has really only been relevant for the past 20 years due to the development of NGS and the accompanying methodologies. However, we are now

beginning to see examples of how genomics has been successfully integrated into conservation management (see supplementary materials for [24]).

1.2.2 The importance of genome data in conservation

On numerous occasions when researchers have applied genomic data to projects that previously used genetic markers they find inconsistencies which have led to incorrect conclusions and management decisions [77, 40]. For example one study on the regionally endangered Eastern Tiger Salamander (*Ambystoma tigrinum*) used a dozen microsatellites, across 26 ponds [78]. These twelve loci showed that the populations had a low genetic diversity, low effective population size and that there was little genetic differentiation between the ponds. From these results the authors inferred high connectivity and migration between populations and advocated maintaining this connectivity by preserving habitat corridors [78]. However when these populations were re-examined using 5000 SNPs the authors found some major discrepancies between their results and those of the previous study [79]. Using genomic data their results showed significant differentiation and sub-structuring between the populations from different ponds caused by limited connectivity and due to the fine scale resolution of the data they were able to identify the presence of roads as a major barrier to the movement of these amphibians. Although the results of their research contradicted those from the previous study their main conservation recommendations remained the same and they advocated for increased connectivity [79]. Whilst the use of genomic data did not change the management advice for the Eastern Tiger Salamander, management recommendations did change when researchers used genomic markers to re-investigate the phylogeography of the Yellow-Legged Frog (*Rana boylii*). In 2011 researchers used 1525bp of the mitochondrial genome and 517bp of a nuclear intron and found little population sub-structure and therefore recommended that the management units for this species be determined by hydrologic boundaries [80]. When thousands of genomic markers were applied in a similar study years later researchers were able to detect significant population structuring into five distinct phylogenetic clades and therefore suggested management units based on these five clades and the genetic rescue of these clades by assisted migration of individuals between populations [77].

In this section I have discussed some of the merits of including genetic information and the transition between using a handful of genetic markers and genome-wide data. Ultimately the precise nature of the information required (genetic versus genomic) will depend on the nature of the research or management question and the resources available (funding and sampling) [25]. Importantly although there does seem to be obvious merits to producing and using genomic data, this thesis does not set out to argue genomics is better than genetics but

rather that the incorporation of genetic information is necessary for successful conservation management.

1.3 Genetic rescue

One conservation management framework, which has contributed to the recovery of several endangered populations [81–83], and for which genetic information is essential, is genetic rescue. Genetic rescue aims to increase the fitness of small, fragmented populations with low genetic diversity, high genetic load and inbreeding depression by introducing novel alleles [84, 11]. The introduction of alleles into a declining population can occur naturally via migration [85, 86] or it can be part of a human-mediated management strategy such as; translocating individuals from other wild populations [84], releasing captive bred individuals [87, 88] or even the individuals from a closely related sub-species [83, 89]. Genetic rescues have resulted in significant increases in composite fitness (combined fecundity and survival) that have been shown to persist beyond the first generation following an introduction, due to reductions in the frequency or the eradication of deleterious traits produced by inbreeding depression [90, 84, 91, 92].

One of the most well known examples of a successful genetic rescue is that of the Florida Panther (*Felis concolor coryi*). The Florida Panther declined to about 22 breeding animals in 1990 and displayed a number of physical traits associated with inbreeding depression such as kinked tails, heart defects and undescended testicles. Furthermore genetic analysis using microsatellites confirmed that the population had low genetic diversity [34]. As the population continued to decline the decision was made to translocate eight female Texan panthers (*Felis concolor stanleyana*) in an attempted genetic rescue. When the population was reassessed 10-15 years after the genetic rescue they found that heterozygosity and survivorship had increased, the census population size tripled and the effective population size doubled [83]. Furthermore the negative physical traits such as undescended testicles, had almost entirely disappeared [93, 83].

In the genomics era we are able to examine the impact of genetic rescue in finer detail and provide evidence for the continued investment and use of genomics as part of a well rounded conservation programme. The Swedish wolf project is a good example of how information acquired from genomic data can be used as evidence for both the impact of genetic rescue and the continued use of genomic data in conservation projects. The current Scandinavian grey wolf (*Canis lupus*) population was originally founded by the natural immigration of a breeding pair of wolves in the early 1980's [86], since then a further 12 wolves have immigrated with 5 of those becoming founders. A recent study using whole genome re-sequencing data from 97

wolves showed that resident wolves have entire chromosomes that are identical by descent and homozygous and that this is not the case for the immigrant wolves. Furthermore they showed that genomic data was a more accurate estimate of inbreeding than pedigree data which underestimated the levels of inbreeding in the population. This study demonstrated that the population of Scandinavian wolves are highly inbred, suggested candidate regions responsible for the inbreeding depression they experience and in terms of management, it showed the importance of genetic rescue to try and re-introduce variation into the population at a genome-wide level [62].

Despite both the theoretical and empirical benefits of genetic rescue there are relatively few examples of it being used to help conserve endangered species [91, 94, 95]. Some reasons for this are: socio-economical, a lack of clear guidelines, disruption to social systems and a concern for biosecurity [96, 91, 19]. The primary reason often cited is the fear of outbreeding depression which is the decrease in fitness that can occur in the generations following the crossing of individuals from separate populations [97, 96, 91, 94]. There are two mechanisms that can contribute to outbreeding depression; one is the loss of local adaptation in individuals in a population, known as extrinsic outbreeding depression and the other is the disruption of co-adapted gene complexes called intrinsic outbreeding depression [98, 99]. There is a growing body of evidence showing that the effects of outbreeding depression are often temporary, minor (or non-existent) and can be avoided entirely if populations meet the following criteria: (1) they are naturally outbreeding, (2) they share the same karyotype, (3) they have been isolated for less than 500 years, and (4) they come from similar environments [96, 91, 100].

Genetic rescue is not a panacea and there are many things for practitioners to consider before attempting it including which individuals to translocate, the financial and welfare burden on both the resident and introduced individuals, whether a single genetic rescue will be enough or what other management strategies may be needed alongside it [101]. However overall genetic rescues could make a significant contribution to lowering the extinction risk of many endangered populations with low genetic diversity and there is no clear evidence to support the scarcity of studies and reticence to utilise them as part of a balanced conservation strategy [91].

1.4 Genetics in management and policy

Inadequate conservation strategies can have disastrous consequences [27] but conservation is a ‘crisis discipline’ where practitioners may have to make decisions quickly based on incomplete knowledge [18, 102]. The time scales involved when working with endangered species often prevent scientists from being able to perform replicated experiments and long-term monitoring

[103]. Given these constraints and the potentially serious consequences of poor management plans, modelling provides a means to test different management strategies, such as genetic rescue, and observe the predicted outcome on the study species [104, 89, 105]. Population Viability Analysis (PVA), are programs designed to model the probability of extinction of populations given a set of demographic and environmental parameters and management strategies [106]. By modelling different management strategies PVAs provide quantitative output which can be used to compare the effectiveness of different conservation plans [101]. To this end PVAs have been used to examine the effectiveness of different contraceptive regimes on controlling populations [107], the impact of climate change on population viability [108], the impact of hunting or harvesting [109] and the outcomes of genetic rescue [101].

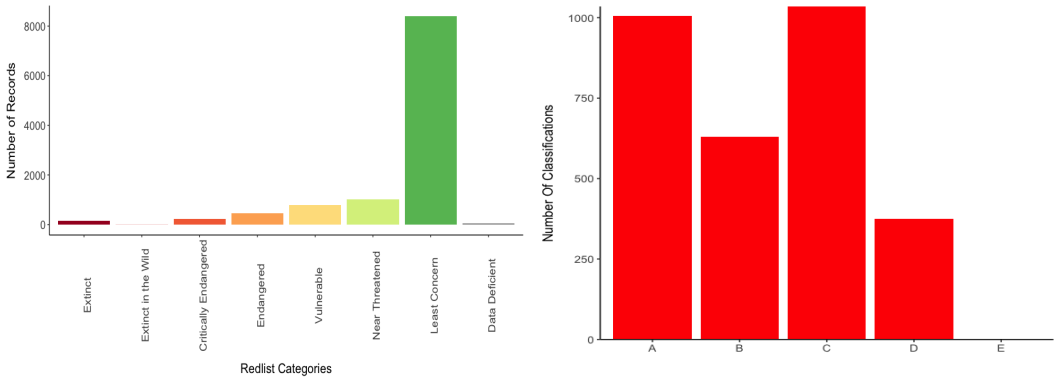
The use of PVA in assessing the likely outcomes of translocation events (including genetic rescues) has been recommended by both academics [110, 111, 101] and by the the International Union Conservation of Nature (IUCN) [112]. The IUCN is a membership union which brings together academics, practitioners and governments to collect data on global biodiversity, design management plans, and identify species in need of conservation action. One way the IUCN does this is through the creation and maintenance of the IUCN Red List of Threatened Species (hereafter Red List). The Red List was designed to *'provide information and analyses on the status, trends and threats to species in order to inform and catalyse action for biodiversity conservation'* [113]. Practically this means gathering a wealth of data about a species then using a standard, quantitative framework to gauge their risk of extinction [1]. The framework used to assign a species threat status is described in Figure 1.2a but in brief it includes the consideration of population size, rate of decline, range area, quantitative models (or a combination of these). The quantitative models described in category E refer to PVAs however in practice this category is rarely used to assess a species' level of risk ([113]), and this is certainly true based on data from threatened bird species (see Figure 1.2b, Figure 1.2c), even though this would be one way of assessing the need for genetic rescue and integrating genetic data in to policy and management [101]. Currently the genetic health of a species is not explicitly considered in Red Listings, which could lead to a misclassification of threat level and an underestimation of extinction risk [13]. The lack of genetic data in assessing the threat of a species is especially surprising given that the IUCN created the Conservation Genetic Specialist Group and advocates the importance of genetics in management plans [112].

The lack of inclusion of genetic data to describe the extinction risk of species by a global and prominent conservation body like the IUCN is perhaps symptomatic of the larger challenge of incorporating genetic data into policy. Despite numerous examples of the utility of genetic methods and the need to preserve genetic diversity to preserve viable populations

Criteria	Critically Endangered	Endangered	Vulnerable	Qualifiers
A1: reduction in population size	≥ 90%	≥ 70%	≥ 50%	Over ten years/three generations[a] in the past, where causes of the reduction have ceased and are clearly reversible and understood
A2–4: reduction in population size	≥ 80%	≥ 50%	≥ 30%	Over ten years/three generations[a] in past, future or combination
B1: small range	< 100km²	< 5000km²	< 20 000km²	Plus two of (a) severe fragmentation and/or few locations (1, ≤ 5, ≤10); (b) continuing decline; (c) extreme fluctuation
B2: small range	< 10km²	< 500km²	< 2000km²	Plus two of (a) severe fragmentation and/or few locations (1, ≤ 5, ≤10); (b) continuing decline; (c) extreme fluctuation
C: small and declining population	< 250	< 2500	< 10 000	Mature individuals. Continuing decline either: (1) over specified rates and time periods; or (2) with (a) specified population structure or (b) extreme fluctuation
D1: very small population	< 50	< 250	< 1000	Mature individuals
D2: very restricted population	N/A	N/A	< 20 km² area of occupancy or ≤ five locations	Capable of becoming Critically Endangered or even Extinct within a very short time frame
E: quantitative analysis	≥ 50% in ten years/three generations[a]	≥ 20% in 20 years/five generations[a]	≥ 10% in 100 years	Estimated extinction risk using quantitative models (e.g. population viability analyses)

a - whichever is longer

(a)



(b) **(c)**

Figure 1.2 Information about the IUCN criteria **(a)** shows the criteria that need to be met for different threatened categories adapted from [1]. **(b)** and **(c)** are based on available data from the IUCN [accessed 12 February 2019] for assessed bird species only. The number of assessed bird species falling in each category can be seen in **(b)** and of those in a threatened category (Vulnerable, Endangered, Critically Endangered) **(c)** shows which criteria were used to assign their threat level.

there is still a lack of policies that explicitly consider genetic diversity [76]. This gap between academic research and practical conservation highlights the need for further integration of genetics in management strategies by increasing academic and agency collaboration [26, 20]. One way in which this issue can be addressed is through the use of model species and case

studies to demonstrate the impact of various regimes and give clear examples to practitioners, academics and agencies alike [114, 26]. The pink pigeon (*Nesoenas mayeri*) is a threatened species with over 40 years of ecological data, genomic data and a successful captive population that would make a valuable model species for the integration of genetic data into balanced conservation plans.

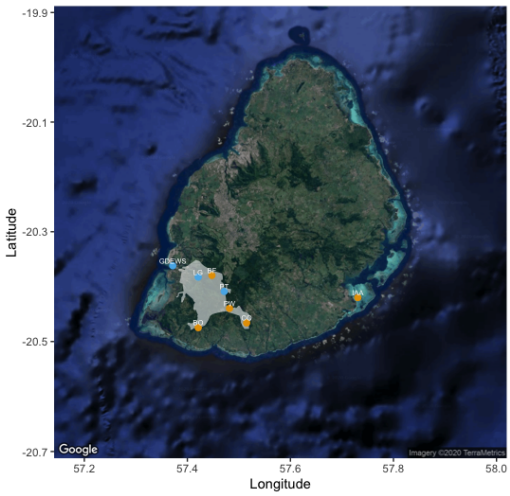
1.5 The pink pigeon *Nesoenas mayeri*

1.5.1 Pink pigeons a potted history

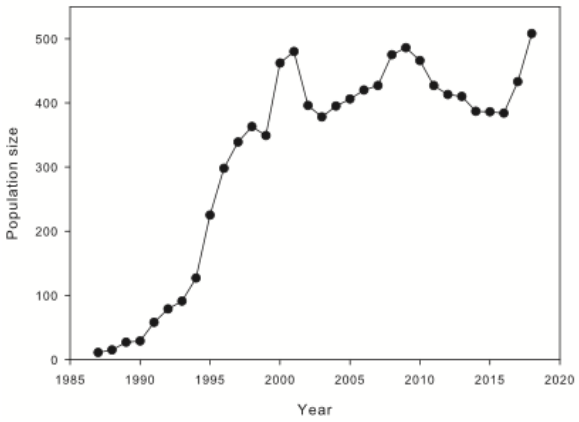
The pink pigeon (*Nesoenas mayeri*) is endemic to the island of Mauritius (1,865 km2 at 20.250S 57.50E) in the south-west Indian Ocean (see Figure 1.3b). Mauritius was uninhabited until 1598 when humans settled on the island and since then island has been markedly changed by a variety of anthropogenic activities. Mauritius lacked mammalian predators so the native wildlife made easy prey for, humans and subsequently a number of invasive predators. Many alien species have been introduced to Mauritius both accidentally (Black Rats *Rattus rattus* and Crab-Eating Macaques *Macaca fascicularis*) and intentionally (Feral Cats *Felis Catus*, Indian Mongooses *Herpestes auropunctatus*, Deer (*Cervus timorensis Blainville*))[115]. These alien species have predated and competed with native species and acted as zoonotic reservoirs for disease [116, 117]. As humans continued to colonise Mauritius, as well as introducing a variety of non-native species, they also cleared large swathes of forest for agriculture and infrastructure[118]. Mauritius has been farmed extensively since the 18th Century for sugar cane, tea and coffee with the result that an island that was originally covered in dense forest now has less than 2% of its native habitat remaining [119]. All of these factors have contributed to the decline and extinction of many of the islands native and endemic species and in the early 1970s the pink pigeon was on the brink of extinction with only about 20 individuals remaining. The pink pigeon had been suffering from habitat loss, predation from four introduced predators (Feral Cat (*Felis Catus*), (Black Rat *Rattus rattus*), Indian Mongoose (*Herpestes auropunctatus*) and Crab-Eating Macaque (*Macaca fascicularis*)), invasive pathogens and inbreeding depression [119–122]. After intensive conservation management, which included the creation of both *in situ* and *ex situ* captive breeding programmes, supplementary feeding, predator control and restoration of native habitat, the species reached its current size of approximately 400 (see Figure 1.3c) individuals in 1999 [122]. The road to recovery has not been smooth and in the early 1990s only circa. 10 wild individuals remained however since 1999 the wild population has remained relatively stable (circa. 400 individuals). Despite this recovery pink pigeons are still intensively managed, they are provided with supplementary feed, predator control is

ongoing, diseased individuals are treated [122, 2] and they are being successfully kept and bred in zoos in the UK, Europe and the US.

(a)



(b)



(c)

Figure 1.3 The location of the pink pigeon (**(a)** image taken from https://www.durrell.org) populations on Mauritius (1,865 km2 at 20.250S 57.50E) can be seen in **(b)** the white shaded area represents the Black River Gorges National Park which includes Pigeon Wood where the last remaining population of pink pigeons were found in the 1970s. The subpopulations are represented by circles where orange represents subpopulations for which there is data present in this thesis and blue subpopulations that have been formed since the collection of this data and therefore are not included in any analysis in this thesis. The formation of new subpopulations has been possible due to the pink pigeons recovery **(c)** since the beginning of the release program with the population numbering around 350-400 wild individuals today

Table 1.2 Birds translocated from GDEWS to Jersey and vice versa as part of the pink pigeon conservation programme

Year	To Jersey	To GDEWS
1978	3	-
1979	-	2
1981	6	-
1982	2	5
1986	6	-
1989	2	-
1990	1	10
1992	-	2
1993	8	-
1995	9	-
1996	-	5*
Total	37	24

1.5.2 Captive pink pigeons

The captive breeding programme was established on Mauritius in 1976 with three male birds taken from the remnant wild population at pigeon wood, another male and three females were taken from the remaining wild population from 1976-1981 and founded the *in situ* captive population at what is now known as Gerald Durrell endemic wildlife sanctuary (GDEWS, which was originally called the Black River Aviary). The *ex situ* captive population was founded in 1977 with three males and a single female. Early on in the programme researchers recognised the need to balance founder-representation in the captive and free-living populations as well as ensuring genetically valuable stock was shared between Jersey and GDEWS (see Table 1.2).

1.5.3 The ecology and biology of the pink pigeon

The pink pigeon is primarily an arboreal species that is well adapted to foraging for fruit, leaves, seeds and flowers of trees at the tips of branches [123]. Pink pigeons are monomorphic and the different sexes are distinguished by vocalisations and behaviours, they form long-term monogamous pairs usually with males defending their nesting territory [9, 119]. Breeding may occur all year around but there is a marked decrease in breeding between February and April [9]. Females lay 1-2 eggs which both parents will take turns to incubate for 14 days, once the eggs hatch (typically only one will survive) chicks are cared for by both parents which for the first three days includes being fed solely on the lipid-protein rich crop milk. Squabs fledge at 23 days post-hatching but remain dependant on their parents for food for a further

two weeks post fledging (a month if no supplementary food provided) and can accompany their parents up to six months post-hatching. Both multiple broods and overlapping clutches are known to occur in pink pigeons [9].

1.5.4 Challenges

Due to its impressive recovery the pink pigeon is considered an iconic conservation success story however they still require intensive management due to a number of challenges. For example pink pigeons experience high mortality during their first year of life. In free living birds in the IAA sub population it was calculated that out of all the eggs laid less than 12% would go on to fledge [120]. Whilst some of this mortality is due to predation the pink pigeon also suffers from inbreeding depression and susceptibility to disease both of which are likely to be compounded by the fact that they have low genetic diversity.

Inbreeding

Pink pigeons are more inbred than would be expected by chance in a randomly mating population and the level of inbreeding within certain sub-populations has increased over time [2]. Previous studies have identified some individuals with inbreeding coefficients that meant they shared the same proportion, or greater, of their genome as siblings ($F > 0.25$) [120]. It is not surprising therefore that pink pigeons suffer from inbreeding depression which manifests as reduced fertility (eggs laid, hatched), reduced longevity and the reduced survival of juvenile birds [120, 124, 2]. It has also been hypothesised that inbreeding increases the susceptibility of pink pigeons to diseases as as there is a negative relationship between genome-wide heterozygosity and susceptibility to infections by the protozoan parasite *Trichomonas gallinae* [2]

Diseases

Avian trichomonosis, caused by the protozoan parasite *Trichomonas gallinae*, is a major cause of mortality in young pink pigeons . *T. gallinae* is believed to have arrived on Mauritius with the introduction of several exotic dove species - the Madagascar Turtle-Dove (*Streptopelia picturata*), Zebra Dove (*Geopelia striata*), Spotted Dove (*Streptopelia chinensis*) and the Feral Pigeon (*Columba livia domestica*). Although these introductions occurred during the 18th or 19th centuries trichomonosis wasn't diagnosed in pink pigeons until 1992 [121, 125, 126]. It can be spread both directly from parents to offspring or indirectly at shared food and water

sources [125]. When infected, birds develop lesions and/or necrotic ulcerations in the upper digestive and respiratory tracts, in severe cases lesions can block the respiratory or tracheal openings and birds can die from either starvation or asphyxiation [127]. Trichomonosis related mortality in adults is low however over the past decade the parasite is responsible for killing over 50% of nestlings and juveniles and reducing the long-term survival of pink pigeons with sub-clinical infections [127]. Birds are screened using the InPouch TF method (BioMed Diagnostics, White City, Oregon) [128] and infected birds have been treated with carnidazole [116]. Survival of pink pigeons can be further impacted by the the blood parasite *Leucocytozoon marchouxi* but unlike *T. gallinae* there has been no direct mortality recorded that has been caused by *L. marchouxi* [117].

Genetic data

Over the last 30 years both genetic and genomic markers have been generated for the pink pigeon as part of the ongoing conservation programme. The first genetic data to be analysed for the pink pigeon was DNA fingerprinting in 1991, these markers indicated that the founders of the captive population were related which is in contrast to the assumption made by the pink pigeon studbook that all founders were unrelated [129]. Further evidence supporting this was gained by the sequencing of 730bp of the mitochondrial d-loop which showed a low haplotypic diversity, there were 3 haplotypes found in eleven birds, with most of the birds (78%) sharing a single haplotype [120]. More recently microsatellites were used to compare the free-living wild population and *ex situ* captive population. Using a panel of 22 microsatellites, researchers found 164 alleles across the wild population and both the *in situ* and *ex situ* captive populations; of those, 45 alleles were unique to wild individuals and 25 were found only in the *ex situ* captive population. Using these markers the effective population size for the pink pigeon in 2010 was calculated as fewer than 50 birds [13]. A reference genome was assembled from a captive male pink pigeon which had an N50 of 8Mbp [2] and 93% single copy complete orthologues from BUSCO analysis. This genome was used to help generate RAD markers from wild individuals, which led to the discovery of 45,841 SNP. Subsequent analysis using the SNP data showed that there was little population structure although the sub populations did appear to be impacted by drift both temporally and spatially [2]. Further analyses showed low genome-wide diversity in the wild pink pigeon population (Figure 1.4) which given its demographic history and current low population size is not surprising and is similar to other endangered bird species (see Table 1.3).

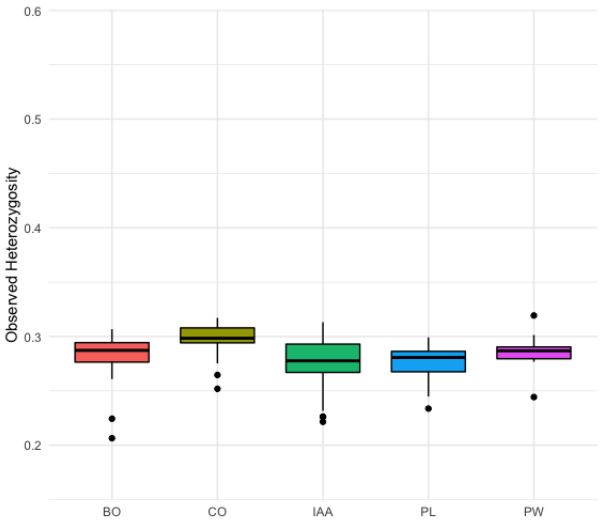


Figure 1.4 Observed heterozygosity from five pink pigeon sub populations using SNP data, recreated using data from [2]

Table 1.3 Estimates of observed heterozygosity in threatened bird species

IUCN Classification	Species	Ho	Ref.
VU	pink pigeon (*Nesoenas mayeri*)	0.530 / 0.280*	[2]
VU	North Island Brown Kiwi (*Apteryx mantelli*)	0.508	[130]
VU	Chatham Parakeet (*Cyanoramphus forbesi*)	0.349	[131]
EN	Floreana Mockingbird (*Mimus trifasciatus*)	0.250	[132]
EN	Galapagos Penguin (*Spheniscus mendiculus*)	0.063	[133]
EN	Crested Ibis (*Nippon nippon*)	0.121	[134]
CR	Northern Bald Ibis (*Geronticus eremita*)	0.553	[135]

* All values of observed heterozygosity are based on mean observed heterozygosity of a microsatellite panel except the asterisked value which is based on SNP data.

Genetic rescue

Since the conservation programme began, pink pigeons have been reintroduced from what is now called the Gerald Durrell Wildlife Sanctuary (GDEWS, the *in situ* breeding facility on Mauritius) and between 1987 and 2009 eight birds have been reintroduced into the wild from Jersey Zoo (Figure 1.5). Over the past few years there has been further discussions about reintroductions from the *ex situ* captive population when it was discovered that they harbour novel alleles and could provide a valuable source genetic diversity for a genetic rescue [13]. In the summer of 2019 the first three birds were sent to Mauritius from Jersey Zoo, where

they will be bred with wild birds in a government facility and then their F1 progeny will be used to supplement the wild population (Harriet Whitford and Carl Jones pers. comms.).

	Plaine Lievre	Bel Ombre	Ile aux Aigrettes	Combo	Lower Gorges
Release period	1987-1996	1994-2000 & 2009	1994-1996	1999-2000	2007-2009
Number of cohorts	28	21	9	9	13
Number of birds released	107	105	47	46	51
Number of captive-bred birds released	107	96	46	34	9
Number bred at GDEWS Mauritius	102	96	46	34	6
Number bred at DWCT Jersey	5	0	0	0	3
Number of wild-bred birds released	0	9	1	12	42
Number of Males:Females:Unknown sex	53:36:20	43:44:18	28:19:01	17:21:06	11:08:36
Number surviving to 30 days post-release	91	94	44	42	33

Figure 1.5 Introduction of birds into sub populations between 1987 and 2009. Birds were sourced from the wild, the *in situ* captive population (GDEWS) and *ex situ* captive population (Durrell Wildlife Conservation Trust (DWCT), now known as Jersey Zoo. Table taken from [3])

1.6 Thesis description

This thesis aims to demonstrate how genome-wide data can be effectively integrated into management and generate resources that will be useful for the continuing conservation of the pink pigeon. The successful conservation of the pink pigeon, or any endangered species, relies on accurate information to make informed management decisions and this thesis presents tools that aim to maximise the amount of information that can be gained from data which can then be used to understand the levels of genetic variation within a population. Throughout this thesis there is a strong emphasis on understanding the biases that may be introduced by data processing, parameterisation or untested prior assumptions and how this can generate incorrect or biased data sets that could lead to ill-informed management decisions.

Chapter 2 generates a high-quality pseudo-chromosome assembly reference genome for the pink pigeon using a reference assisted chromosome assembly method. This reference is used together with other whole genome data to examine genome-wide variation and inbreeding in the captive pink pigeon population as well as the ancient demography of the pink pigeon. This chapter demonstrates the power of whole genome data versus genetic data for ensuring accurate conclusions are drawn about the levels of variation. All the analyses for this chapter were conducted by Camilla Ryan under the supervision of Bernardo Clavijo

(Earlham Institute (EI)), Cock Van Oosterhout (University of East Anglia (UEA)) with help from Ben Ward (EI) and guidance from Gonzalo Garcia Accinelli (EI) and Jon Wright (EI).

Chapter 3 introduces RADiKal, a novel software for producing an overview of genome-wide population level variation using raw data from RAD-seq experiments. RADiKal produces an overview of the variation found in the wild pink pigeon population which is compared to the results from whole genome analysis in Chapter 2. The results from RADiKal emphasise the need for methods that are able to extract signals from raw data to reduce the biases that are caused by processing pipelines and provide a method for examining and testing prior assumptions. All the analyses and programming for this chapter were conducted by Camilla Ryan under the supervision of Bernardo Clavijo (EI), with help from Ben Ward (EI) and guidance from Gonzalo Garcia Accinelli (EI) and Jon Wright (EI).

Chapter 4 presents an updated Vortex model for the pink pigeon that incorporates microsatellite data and predicts the risk of extinction for the wild pink pigeon population given three scenarios. These scenarios are a naive "do nothing" scenario, the impact of demographic rescue and the impact of genetic rescue. The importance of accurate data and parameterising is also discussed with particular attention given to the modelling of genetic load by Vortex using lethal equivalents. All the analyses for this chapter were conducted by Camilla Ryan under the supervision of Cock Van Oosterhout (UEA).

Chapter 5 presents I Choose You (ICY) a tool designed for use by practitioners to help them select a suitable group of individuals for genetic rescues or reintroductions when no genetic data is available. ICY is a web-based tool that is easy-to-use and uses existing studbook data to calculate founder equivalents and use them in combination with mean kinship to select a group of individuals suitable for genetic rescue. All the analyses for this chapter were conducted by Camilla Ryan under the supervision of Bernardo Clavijo (EI), Cock Van Oosterhout (UEA) with help from Ben Ward (EI) and guidance from Gonzalo Garcia Accinelli (EI) and Jon Wright (EI).

Chapter 6 is a general discussion about the findings presented in this thesis including what they may mean for pink pigeon conservation. The implications of these findings for the field of conservation genetics as a whole are discussed.

All the laboratory work that produced the raw sequence data that were used in this study was undertaken by the Earlham Institute.

Chapter 2

Genome-wide variation and inbreeding in the pink pigeon *Nesoenas mayeri*

2.1 Introduction

The greatest unmet challenge in conservation is successfully integrating genetic data into management [51]. Although there are now several examples where integrating genetic data have improved conservation management [83, 11, 63] these are still in the minority [136]. A valuable genomic resource for the study of any endangered species is a reference genome [63]; an assembly that is considered the representative genome for a given species. Access to a reference genome has several advantages, in particular, it provides a stable coordinate system. This coordinate system enables the placement of single nucleotide polymorphisms (SNPs), insertions and deletions (indels), comparative studies with other organisms and the annotation of genes. This is especially important for studies that are trying to identify and study regions or genes that may confer an adaptive advantage and therefore need to consistently identify the same region in different individuals [137]. This same coordinate system can also be used to study variation across a genome, which is more informative than a single mean value of heterozygosity. A mean value condenses all the information from a genome into a single number that does not represent any single part of the genome. Access to a reference genome also allows for accurate calculations of mechanisms that otherwise may only be estimated. For example, it provides the location of areas with high linkage disequilibrium and recombination breakpoints which are important for identifying haplotypes

and understanding their inheritance [138]. Overall, although useful analyses can be conducted without a reference genome (see Introduction), access to a reference genome increases the precision of some and enables additional analyses [63]. It can even be more cost effective than more traditional methods [139, 63]. For example microsatellite markers have been traditionally used to assign parentage or examine population structure. Unless primers for microsatellite markers are already available, the development and production of these markers, is now more expensive than sequencing the whole genome which can then be mined for all microsatellite motifs [63]. Unfortunately, reference genomes are not widely used in conservation, this lack of integration into management (discussed further in the Introduction) may be due to logistics (obtaining samples from endangered species), financial costs, lack of expertise or a lack of knowledge about how genomic information can be practically useful to saving a species [140, 24]. However due to the decrease in sequencing costs and the efforts of consortia such as the Vertebrate Genome Project and Bird 10K an increasing number of high quality genomes are being sequenced [141, 63]. This has led to studies that demonstrate how access to a reference genome can advance our understanding of genomic mechanisms and how this understanding can be integrated in to the management of endangered species like the pink pigeon (*Nesoenas mayeri*) [136].

The ability to predict and understand a species response to the current anthropogenic or climate pressures is important to enable pro-active conservation measures [142, 143, 55]. Traditionally this has meant studying genetic diversity with the assumption being that the greater the genetic diversity within a species, the greater its ability to adapt to future challenges [41, 91]. However there are several species which appear to violate this assumption, for example many invasive species are very successful despite low initial levels of heterozygosity due to founder effects [144]. There are also several species that appear to have been living for millennia with low levels of variation because of historical demographic events [145–147]. One reason for this apparent paradox is that traditional methods, such as microsatellites, measured putatively neutral variation which would not be expected to reflect adaptive potential. Instead, whole genome data is required to infer signatures of selection and identify loci carrying adaptive genetic variation [148]. To study this requires access to a reference genome. Nevertheless, it is important to note that population level data cannot be assessed from a single genome [149], but the presence of a reference provides a template for resequencing experiments to generate population level data [139, 150, 55]. In other words, a reference genome enables population genetic analysis, but it is not sufficient by itself to conduct such analysis [149]. For example access to a reference genome allowed scientists studying the coral, *Acropora millepora*, to identify a signal of balancing selection in a region containing a single gene *sacsin* which is known to be up-regulated in response to elevated temperatures [55]. This same study also used whole genome data combined with environmental data to create a model capable of predicting the response of different corals to bleaching, which

poses the greatest threat to the survival of corals. This model should allow researchers to simultaneously identify populations that are susceptible to bleaching and those that may be resistant to bleaching and suitable for translocations to preserve and restore reefs [55]. This is just one example of how access to a reference genome and whole genome data is vital to the study of selection and adaptive potential and in the future there will be more studies harnessing the information within genomes to study these mechanisms and guide conservation [54].

Access to whole genome data also enables analyses of past demography which can provide valuable insight into the current status of species [147] and contribute to deciding on the best possible management strategy [146]. For example whilst it is known that inbreeding and low genetic diversity contribute to the extinction vortex the ability to quantify how much variation is considered low depends on the species, their life history traits, and historic events [54, 4]. A specific example of how historic events can shape heterozygosity can be seen in the vaquita (*Phocoena sinus*) [147], an endangered species of whale found only in the Gulf of Mexico. The vaquita is the most critically endangered marine mammal with fewer than 19 individuals left in the wild, and hence, it is perhaps unsurprising that they have low genetic diversity. However after whole genome data was used to perform a coalescence analysis and examine variation across the genome (see Figure 2.1) it was discovered that the low genetic diversity was principally due to historical bottlenecking as opposed to recent bottlenecking or inbreeding [147]. These findings imply that if the vaquita's main threat (gill-net bycatch) was removed, the species should be able to recover unhindered by their low diversity [147]. Other examples have emerged which demonstrate the power of historical demographic analysis, using whole genomes, to elucidate past extinction events [151], understand the future extinction risks of endangered species [147], define current patterns of heterozygosity and deleterious variants across the genome[4] and observe the impact of genetic rescue at the genome level [62]. There are several methods for inferring the historic demography of a species and most either require or are significantly improved by the use of a reference genome.

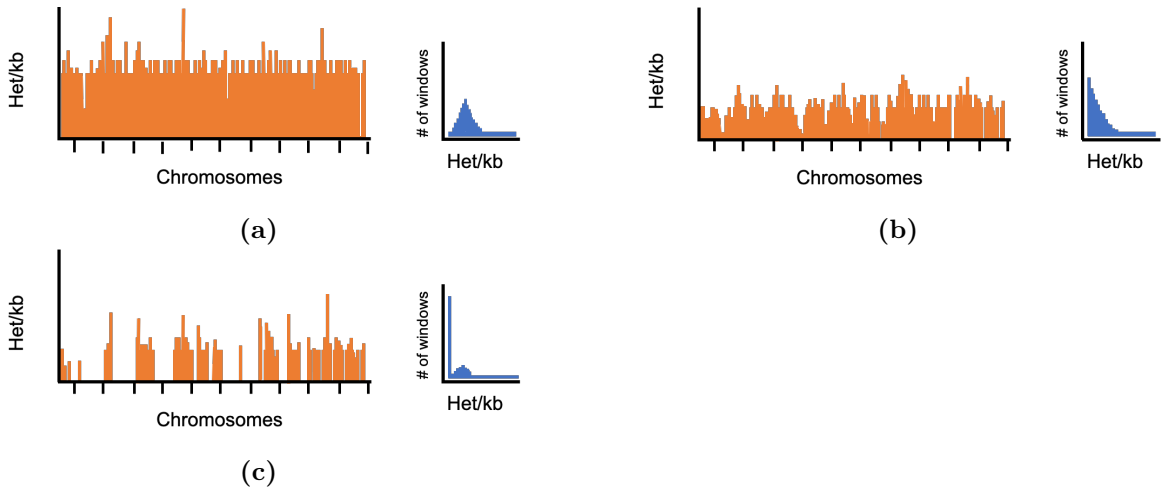


Figure 2.1 Illustration how different patterns of heterozygosity are reflect different demographic processes. In a healthy outbreeding population **(a)** with a large N_e expect high levels of heterozygosity across the genome **(b)** population/individuals with historic bottleneck **(c)** recent inbreeding based on [4]

Inbreeding can lead to inbreeding depression, which is a concern for managers of both wild and captive populations of endangered species because it contributes to a significant decline in population level fitness and exacerbates the extinction vortex [11, 5]. Although inbreeding can be monitored, to some extent, by pedigrees or using a small set of molecular markers neither provide the precision or accuracy that is possible when inbreeding is measured using whole genome data [152, 63]. One measure of inbreeding, that uses whole genome data, is the analysis of runs of homozygosity (RoH); the proportion of the genome found within chromosomes that are identical by descent [5]. Using RoH it is possible to not only accurately estimate the severity of inbreeding but also, indicate the regions of the genome affected by inbreeding and show the direct impact of genetic rescue [152, 4]. By studying the distribution of lengths of RoH it is also possible to distinguish between different historic demographic events [153, 5] (see Figure 2.7a); an outbred population with a large effective population size is expected to have a very few, short runs of homozygosity, whereas many, long runs of homozygosity are indicative of consanguineous mating and recent inbreeding in a population [154, 5, 4].

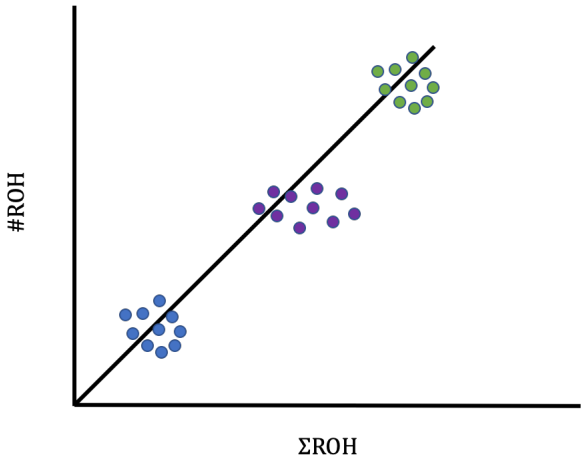


Figure 2.2 Relationship between the length of RoH (y-axis), sum of RoH (x-axis) and demography (represented by different colours). Blue shows a large outbred population, purple an inbred population and green shows a population that has been through a bottleneck. Figure adapted from [5].

Using whole genome data to study inbreeding can also further our understanding about the mechanisms of inbreeding depression, as is illustrated by a serious of studies conducted on the Isle Royale wolf population (*Canis lupus*) [155, 35, 4]. The isolated population of wolves on Isle Royale decreased to two highly inbred individuals in 2018 and, unsurprisingly, they suffered from severe inbreeding depression. The original founders are likely to have originated from a population of wolves in Minnesota which was outbred and had a large effective population size (N_e). Despite their contrasting demographies the two populations have the same number of putatively deleterious alleles. This suggests that the difference in fitness seen between the two populations does not come from an accumulation of recessive deleterious alleles in the Isle Royale population; but that the deleterious alleles are expressed more in the highly inbred population due to an increase in homozygosity [4]. As well as advancing our understanding about the mechanism of inbreeding depression, this finding raises an interesting question in terms of managing the Isle Royal wolf population which is; if wolves were to be translocated on to Isle Royale in a demographic and genetic rescue attempt, the obvious source population should be Minnesota wolves however given that they share the same deleterious alleles this may decrease the fitness of any future generations by increasing the frequency of these deleterious alleles in a small population. Therefore, it might better to consider alternative source populations with significantly different demographic histories [4].

The Isle Royale wolf studies have initiated a heated scientific debate about the best approach to genetic rescue. The majority of empirical and theoretical evidence supports the use of gene flow from genetically diverse populations as the best approach to reverse genetic erosion [91, 95, 51, 156], however, a few studies argue that minimising the introduction of harmful variation is a superior approach [157, 4]. For example by introducing individuals

from small, historically isolated populations where deleterious alleles may have been removed by prior purging [4]. The criticisms of these studies are that they fall prey to the survivorship fallacy; only the few highly inbred populations that survived can be studied not the many that went extinct [16]. But perhaps the most important criticism is that the evidence for these arguments is based on unrealistic simulations and poor parameterisation choices that contradict real-life observations [156]. The success of current genetic rescues (as well as theoretical studies) are evidence for the continued use of individuals from outbred populations with a large N_e and novel alleles [91, 95, 100, 156] wherever possible [158, 159]. Both inbreeding depression and selection are complex phenomena and the continued generation of reference genomes and whole genome data is therefore an essential part of research to continue to refine our knowledge about these phenomena and to allow managers to design highly effective translocations and management strategies.

Although there are many advantages to producing a reference genome there is also the need to ensure that the reference genome is as high a quality as possible and that any biases are considered so that their impact on downstream analyses can be identified and, if possible, limited [63]. The foremost bias common to every reference genome is the assumption that the individual chosen is representative of an entire species, however without any other information about the population (for example a pedigree) researchers have no guarantee that the individual they choose is representative i.e. has the correct number of chromosomes, has similar levels of heterozygosity and inbreeding compared to the other individuals in the population [140, 160]. When studying populations of endangered species that have experienced a recent bottleneck, where the population contains a small and random assortment of individuals, these assumptions are even more likely to be violated which researchers should consider when designing their experiments [161]. Biases and errors are also introduced during sample processing and sequencing (contamination, PCR errors, sequencing errors) and the genome assembly process [162, 163]. Decisions made during the genome assembly process will impact the completeness, correctness and overall quality of the final assembly which in turn will impact the reliability and accuracy of any downstream analyses [164]. These decisions include the choice of assembly software and parameterisation. Many assembly software packages can be compared to black boxes where users put in reads, and set parameters to produce an assembly. However, inside the black box each assembler will have a set of heuristics that can introduce biases into an assembly. For instance many assemblers were designed for work on the human genome, and therefore, a commonly used heuristic is the assumption of diploidy [165]. Clearly, if a species is a polyploid, the assembler would assemble it incorrectly assuming diploidy [166]. If the assembly was *de novo* and the ploidy unknown then the use of this heuristic and its impact may remain undetected and analyses performed using this genome would lead to spurious results. Although this may seem a fairly obvious source of bias each assembly process will add its own biases however because of the

"black box" effect these biases may not be obvious. Therefore it is important to scrutinise the assembly to discover any large mis-assemblies, biases or mistakes made by the assembly process and ensure the final assembly is high quality [162, 167].

Ideally a high quality reference genome will be complete, correct, contiguous, and sorted into chromosomes and haplotypes. However, even the human genome is not at this level [168] and most of the current software used in downstream analysis are designed to work with collapsed mosaic assemblies and not haplotype specific assemblies [139, 169]. Therefore, practically, a reference genome should be assembled to the highest possible standard given the available resources and should have the correct motifs, the correct number of times, in the correct order [140, 164]. Achieving a high quality, accurate reference genome is not trivial and especially not for projects with endangered species. For instance the first challenge in creating a reference genome is getting a high quality sample (usually blood or tissue) from which to extract DNA but this can be logistically challenging especially for endangered species therefore the sequencing strategy will often reflect quality of sample that has been obtained [137, 63]. For example long reads can significantly improve the structure of a genome assembly but they require large fragments of DNA which come from high quality, fresh samples [141, 139]. Instead to achieve the highest possible quality genome for an endangered species may require innovative techniques and hybrid frameworks that combine information from different data sets [170]. An example of such a method is reference assisted chromosome assemblies (RACA) which use one or more chromosome level assemblies to super-scaffold a target genome into pseudo-chromosomes by synteny [171, 172]. This increases the contiguity of the reference genome but contiguity does not necessarily improve the quality of the genome assembly because it does not check for completeness or correctness and therefore it is important to evaluate the completeness and correctness before using RACA [173]. This strategy has multiple advantages for endangered species and one significant disadvantage which is that the algorithms rely on synteny and so will produce the best results when using closely related species. However for vertebrates (where there are representative genome assemblies for almost every major clade [171]) such an approach provides an opportunity to generate genomes with improved contiguity, that allows the study of structural variants, and reduces the cost that would be needed to otherwise generate chromosome level information [174, 170]. Bird genomes are excellent candidates for reference assisted chromosome assemblies due to the relatively small size of bird genomes, the availability of a number of high quality assemblies (due to the efforts of B10K) [175, 176] and the high levels of synteny between avian genomes [177].

The primary aim of this chapter is to produce an improved reference genome for the pink pigeon that has been sorted into pseudo-chromosomes and gain a better understanding

of the genome-wide variation, inbreeding and the historic demography of the pink pigeon population.

2.2 Methods

All the scripts used to produce the data and analyses performed as well as the resultant genome and VCF files can be found in appendix A.1.

2.2.1 Genome assembly

Scaffolding into pseudo-chromosomes

During a previous study a reference genome was created for the pink pigeon from a deceased captive male bird. The protocols used are described in detail in [2]] but briefly a paired end library (2 x 251bp) was sequenced on an Illumina HiSeq 2500 and assembled *de novo* using DISCOVAR [178]. This assembly was then scaffolded by SOAPdenovo [179] using data from sequenced long mate pair libraries (sizes 4, 8, 12 Kb). This produced a high quality draft reference assembly with an N50 of 8mb and (93%) BUSCO groups identified as single copy and complete. For this study it was necessary to improve the contiguity of the assembly further by organising the genome into pseudo-chromosomes to enable certain genome-wide analyses and provide a high quality resource for future studies.

Birds are highly suited to guided assemblies using the genome of a closely related species due to their high levels of synteny [175]. The turtle dove (*Streptopelia turtur*) is the closest relative to the pink pigeon with a high quality, chromosome level assembly, which contains 30 autosomes (1-33 but missing chromosomes 16,29 and 31), one sex chromosome (Z) chromosome and 316 unplaced scaffolds. Therefore, the turtle dove genome (GCA_901699155.1, downloaded from https://www.ncbi.nlm.nih.gov/assembly/GCA_901699155.1) was used to assemble pink pigeon into pseudo-chromosomes using RagTag v1.01 [180]. RagTag was run with default parameters apart from two parameters: (1) to allow RagTag to infer gap length with a minimum allowed gap of 1bp and (2) output all unplaced contigs into the sequence record "chr0". Once the pink pigeon genome had been assembled into 31 pseudo-chromosomes all sequences which couldn't be aligned (those in chr0) or that were aligned to unlocalised scaffolds were put into a single chromosome "ChrU" with each sequence separated by 100 "N". ChrU was not included in any downstream analyses but it was important to include these sequences so that reads could be accurately mapped to the assembly during whole genome

resequencing analysis. This means that the number of misalignments are reduced when reads were aligned to the correct part of the genome.

Genome assemblies using Chromium 10x reads

Whole blood samples were collected from six captive pink pigeons from Bristol and Jersey Zoo (see Table 2.1). The pink pigeon is a protected species under Appendix III of the Convention on International Trade of Endangered Species of Wild Flora and Fauna (CITES). All due diligence was taken to follow the required procedures, blood samples were taken by veterinarians and sent to the Earlham Institute, which is a registered CITES scientific research institute (GB035)).Because of the protection afforded the pink pigeon sampling was opportunistic and because of this there is a pair of siblings included in this study (S6W1688 and S7W1687) and individuals S3B7462 and S4B7703 appear to have similar kinship coefficeint. The impact of including samples on analysis discussed

High molecular weight DNA was extracted and five libraries were created using 10x Chromium technology and sequenced on an Illumina Novaseq using paired end (2 x 150bp) reads. Sample were pooled and run together with a predicted coverage of circa. 20x. The reads were assembled using 10x Chromium propriety software Supernova v2.1.1 [181] using –maxreads=all, –accept-extreme-coverage and output as a .fasta file with a single record per scaffold (–pseudo-hap).

Table 2.1 Information about individuals from which samples were taken and used in whole genome resequencing and genome assembly. This table includes pairwise kinship estimates between each of the seven individuals calculated in PMx v1.5.6 using studbook data.

Location	**Sample ID**	**Unique ID***	1618	1593	1606	1626	1625	1545
Jersey Zoo	S2B7805	1618	0.5381	0.0819	0.1403	0.103	0.103	0.1043
Jersey Zoo	S3B7462	1593	0.0819	0.6743	0.2736	0.0493	0.0493	0.0876
Jersey Zoo	S4B7703	1606	0.1403	0.2736	0.5466	0.0902	0.0902	0.1095
Bristol Zoo	S6W1688	1626	0.103	0.0493	0.0902	0.5441	0.3179	0.0721
Bristol Zoo	S7W1687	1625	0.103	0.0493	0.0902	0.3179	0.5441	0.0721
Jersey Zoo	S1272**	1545	0.1043	0.0876	0.1095	0.0721	0.0721	0.5372

* This refers to a unique ID given to each pink pigeon when they are entered into the international studbook

** This is the individual used to create the reference genome and unlike the other assemblies was not created using Chromium 10x technology

Quality assessment of genome assembly

The traditional measures of genome quality are N50 and BUSCO analysis which assess the contiguity and completeness of the genome however these alone are inadequate to assess whether the genome is complete, correct and contiguous therefore a k-mer spectra was also produced using the k-mer analysis toolkit (KAT) v2.4.2 [182] which assessed how much of the information in the reads was included in the assembly, and whether there might be any large assembly errors.

In order to assess that all of the correct motifs were present in the genome the correct number of times (i.e. the genome was complete and correct) the k-mers from the reads used to build the assembly were compared to the final assembly. If a k-mer appears once in the reads it should only appear once in the genome, if a k-mer does not appear in the reads it should not be in the genome as this implies the assembler has created content. To assess this the –comp function from KAT was used and k-mer spectra generated to allow a qualitative assessment of the genome assembly quality. An example of k-mer spectra and how they are interpreted can be seen in Figure 2.3

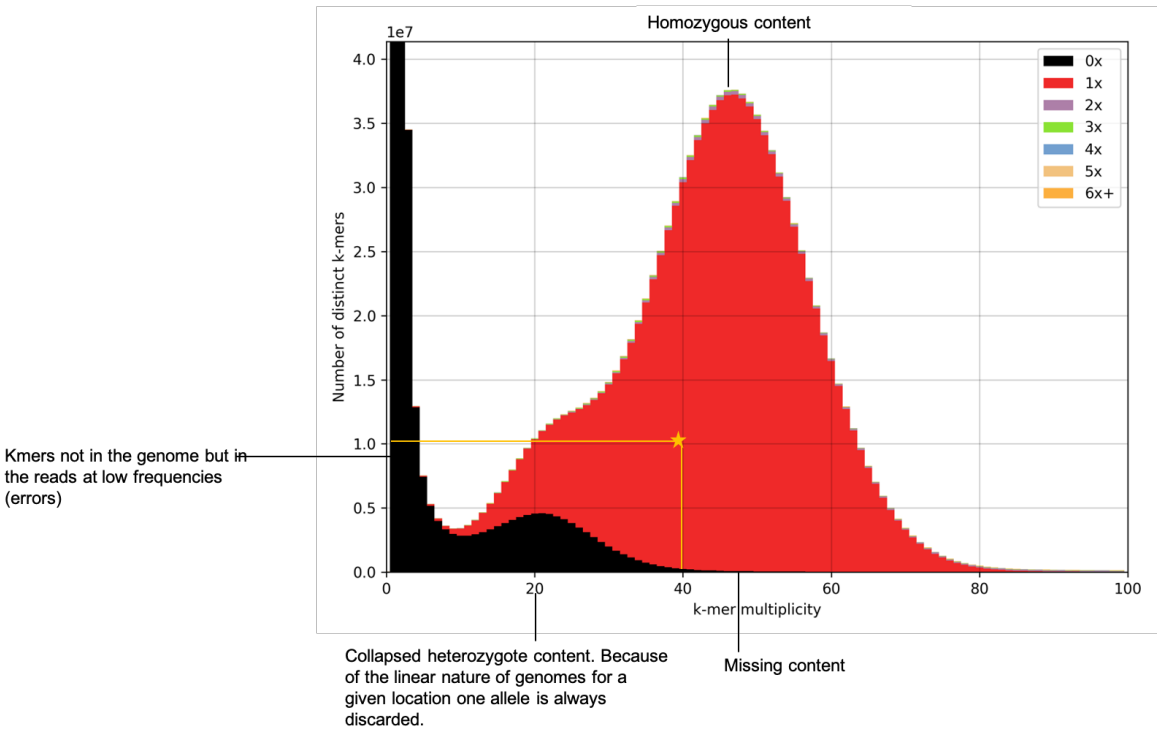


Figure 2.3 An example of a k-mer spectra. A k-mer-spectra is a stacked histogram where the colour represents the number of times the k-mers appear in the assembly for example red means they appear once in the assembly. For a well assembled but unphased diploid genome you would expect the majority or all of the plot to be red, if it was phased the homozygous peak would be purple (k-mers appearing twice, once in each allele) but the heterozygote peak would still be red as each k-mer only appears once in the assembly. The y axis is the number of distinct k-mers and the x axis is the number of times those k-mers appear in the reads - k-mer coverage. This is slightly different to bp coverage of the genome although at longer read lengths k-mer coverage approximates bp coverage. In the example below the yellow star and lines show where 100,000,000 distinct k-mers appear once in the genome and 10 times in the reads. K-mers missing from the assembly are in black, low frequency k-mers that are missing from the assembly are most likely errors and can be seen in the first black peak whereas any data missing from the assembly will be at higher frequencies in this example there is only a very small amount of missing data that can just be seen as a thin black line at the bottom. The missing data under the peak that represents the heterozygote content is expected as the assembly is a collapsed one and therefore discards one of a pair of alleles so you expect a black peak about half the height of the red peak. Any k-mers found to the left of the error distribution would represent k-mers present in the assembly not in the reads and represent mis-assemblies.

BUSCO analysis searches databases of 4915 housekeeping genes [183] that are considered necessary for an organism to survive therefore if a genome is complete it should have all or almost all of them. However BUSCO results are not in themselves a guarantee of a high quality genome, the genes included in the BUSCO databases have been chosen because they are common and easy to assemble therefore if an assembly is high quality then you should be able to detect almost all the BUSCO genes. An assessment of the completeness of the

assembled genome was calculated using BUSCO v3.1.0 [184, 183] and the aves database (aves_odb9).

N50 is the contig size at which more than 50% of the genome is contained in contigs more than or equal to that size and it is used to describe the contiguity of a genome. Whilst N50 is widely used and is useful for giving an overview of contiguity in a genome the results must be interpreted with caution because it does not provide information about the correctness or completeness of an assembly [173]. Abyss v1.9.0 [185] was used to calculated assembly statistics (abyss-fac) for each genome which included N50 and genome size estimates.

2.2.2 Data processing

Whole genome re-sequencing

Chromium 10x reads were processed to remove their barcodes using process_10xReads.py v0.0.2 (https://github.com/ucdavis-bioinformatics/proc10xG.git), the reads were then treated as 150bp paired end. The reads from each individual were mapped to the reference genome using Bowtie2 [186] using a very-sensitive local alignment allowing a single mismatch. Once mapped, duplicates were marked using GATK v4.6.1.0 [187] MarkedDuplicates, then variants were called for each sample using GATK HaplotypeCaller. The calls from each were combined using GATK CombineGVCFs and genotypes called for the whole population (n = 6) using GenotypeGVCFs. Each individual's genotype calls were put in separate vcf files and filtered by depth using Vcftools v0.1.13 [188] where the minimum depth was set to five and the maximum depth was calculated for each individual as two times the mean depth (see Table 2.2). After each individual had been filtered separately for depth the separate files were merged using Vcftools *merge* and a final round of filtering performed that ensured only biallelic SNPs with quality scores of at least Q20 and that did not deviate from the Hardy-Weinberg equilibrium ($p < 0.01$) remained.

Table 2.2 The mean depth of coverage calculated across the genome for the pink pigeon samples and the values used for filtering out putative errors (2x depth)

Sample	Mean depth	2x depth
S1272*	59.3736	118.747
S2B7805	17.583	35.166
S3B7462	19.4108	38.821
S4B7703	18.5392	37.078
S6W1688	19.2853	37.571
S7W1687	16.5053	33.011

* This is the reference individual which is why it has significantly greater depth than the other samples. Unlike the other samples S1272 was not sequenced using Chromium 10x technology.

2.2.3 Data analysis

Heterozygosity

Observed heterozygosity (H_o) was calculated by Vcftools v0.1.13 as the number of heterozygous genotypes divided by the total number of called genotypes. Variation across the genome was calculated separately for each individual in non-overlapping 1MB bins.

Runs of homozygosity

Vcftools v0.1.13 was used to extract runs of homozygosity for each individual. These runs were then grouped by length into short (0.1 MB $\geq$ RoH $<$ 1 MB), medium (1 MB $\geq$ RoH $<$ 10 MB), and long (10 MB $\geq$ RoH $<$ 100 MB). Runs of more than 100Kb were used to calculate inbreeding using F_{RoH} (see Equation 2.1) and runs less than 100 KB were ignored. Vcftools uses a hidden markov model and considers a region a run of homozygosity if the homozygous state is the most likely state in a region of at least 1 cM and containing at least 50 SNPs with a minimum minor allele frequency of 5% [189].

$$F_{RoH} = \frac{\sum RoH > 100KB}{\sum(BP)} \tag{2.1}$$

Where F_{RoH} is the proportion of the genome that is identical by descent and is used as a molecular estimate of inbreeding, RoH is the sum of runs of homozygosity over 100 KB and BP is the sum of all base pairs in the genome (genome size).

R ZooRoH

By examining the size of regions of the genome that are homozygous by descent (HBD) realised inbreeding coefficients can be estimated and by examining the lengths of these segments are informative about inbreeding events that happened at different points in the past. For example recent inbreeding will result in long runs of homozygosity that have not yet had time to be broken up by recombination. RZooRoH v0.2.3[190] uses a hidden Markov model to sort the different lengths of HBD present in a population into generation classes so that the contribution of different inbreeding and past demographic events can be estimated and visualised. Two RZooRoH models were created for the pink pigeon whole genome resequencing data to provide an estimate of realised inbreeding coefficients and to understand the history of inbreeding in the pink pigeon. The models achieve this by sorting the HBD segments into classes (k) where k is approximately double the the number of generations from the time of inbreeding. Given the generation time of the pink pigeon is 5.6 a k of two is equal a single generation or 5.6 years.

R ZooRoH was used to examine possible key points in the pink pigeons history the VCF file produced for the whole genome sequencing data was converted into the required Oxford Gen format using Plink v1.9 [191]. Two models were run, both used the default parameters provided by R ZooRoH but included predefined values of k. The first model was run to examine how different historic events may have impacted the pink pigeon, the values chosen for k and the reasons behind their choice can be seen in Table 2.3. The second model was created to examine events further back in time and because R ZooRoH recommends using a series of rates with a constant ratio between rates the following 14 values of k were modelled (the last value represents the non-HBD segment): 2,4,8,16,32,64,128,256,512,1024,2048,4096,8192, 16384. The output from RZooRoH provides both the overall realised inbreeding coefficient for each individual and estimates of the contribution of each predefined HBD class to the overall inbreeding coefficient.

PSMC

To study any large-scale historic changes in the effective population size of the pink pigeon a Pairwise Sequential Markovian Coalescent (PSMC) model was used. PSMC studies recombination patterns across the genome to infer historical changes in N_e using the inverse relationship between N_e and number of coalescent events [192]. PSMC requires a consensus genome sequence for each individual, to create this consensus, reads were mapped, duplicates marked in the same way as whole genome resequencing above including filtering for depth but excluding any other filters and for these file Samtools v1.7 [193], –mpileup was used to

Table 2.3 Values of k chosen for the R ZooRoH model and the reason behind their inclusion.

k	Generations	Years	Reason*
2	1	5.6	Detect recent inbreeding
4	2	11.2	Detect recent inbreeding
8	4	22.4	Detect recent inbreeding
16	8	44.8	Near the start captive breeding programmes (1970s)
18	9	50.4	Start captive breeding programmes (1970s)
42	21	117.6	Start of 1900s
78	36	201.6	1800s after the establishment of sugar cane
142	71	397.6	Humans arrive on Mauritius
143	71.5	400.4	Non-HBD segment

* Chronology for reasons provided were taken from [115].

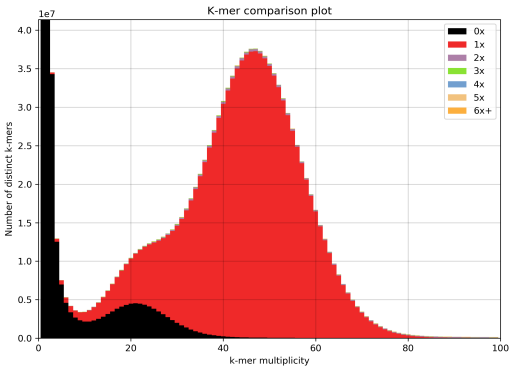
call genotypes. The script vcfutils.pl *vcf2fq* was then used to convert the vcf to a diploid consensus sequence. Each consensus sequence was formatted into suitable input for PSMC using *fq2psmcfa* and *splitfa* from the PSMC software package. The final model for each individual was parameterised following guidelines from [194] which applied following time intervals to multiple species of bird "4+30*2+4+6+10" and the model was bootstrapped 100 times. To plot the results of the PSMC requires a generation time and mutation rate to be able to scale the plot correctly, the generation time used was 5.6 years which was taken from pink pigeon studbook data using PMx v1.5.6 [8]. Presently, no mutation rate has been calculated for the pink pigeon, and therefore a value of $4.598^{e\text{-}10}$ was used for the domestic pigeon (*Columba livia*) taken from supplementary material of [194]. Although this may differ from the pink pigeon's mutation rate, the results will from the model will still be useful because mutation rate impacts the plot in a predictable manner; it does not change the shape of the plot, but rather could shift the plot along the x-axis resulting in incorrect estimates of the timing of events and could impact the estimates of N_e. For example a halved mutation rate would shift the curve further back in time and double the N_e estimates [195]. All plots were created using psmc_plot.pl.

2.3 Results

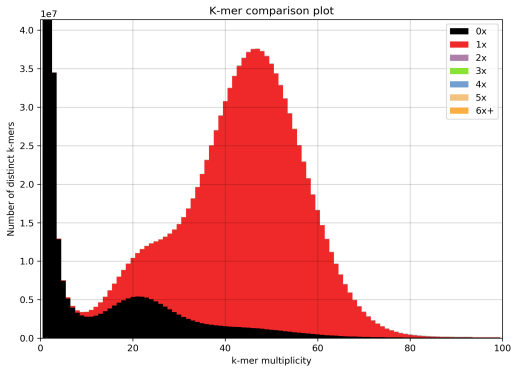
2.3.1 Genome assembly

The turtle dove genome was used to successfully assemble the pink pigeon genome into 31 pseudo-chromosomes. The resulting super-scaffolded assembly was high quality with good completeness, correctness and contiguity. A quality assessment of the genomes assembled using Chromium 10x data can be found in appendix A.1.

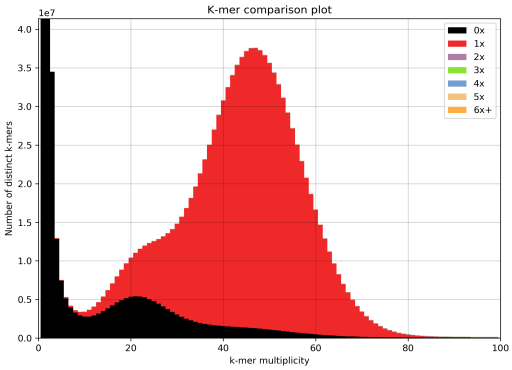
The k-mer spectra (see Figure 2.4) show that most of the content in the reads is captured by the original assembly and there are no mis-assemblies (Figure 2.4a). Some small amount of content is lost when the genome is scaffolded by synteny into pseudo-chromosomes (Figure 2.4b), and a little more content is lost if ChrU is discarded (Figure 2.4c), this result is confirmed by the results from Abyss in Table 2.4. Even so, the majority of the content is still present and there are no large mis-assemblies. Surprisingly the k-mer spectra shows that there are comparable levels of heterozygosity in the pink pigeon reference genome and the human genome, which can be judged by comparing the size of the heterozygous peak and homozygous peak; in all k-mer spectra produced the heterozygous peak is about 1/4 of the height of the homozygous peak (see Figure 2.4d). This is unexpected given the different current and recent past demographics of humans and pink pigeons and indicates that the pink pigeon may have more diversity than was initially thought.



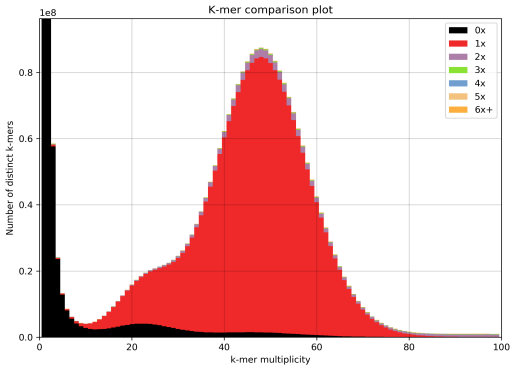
(a) Original pink pigeon assembly



(b) pink pigeon pseudo-chromosomes with ChrU



(c) pink pigeon pseudo-chromosomes without ChrU



(d) Human genome assembly

Figure 2.4 K-mer spectra comparing **(a)** the original pink pigeon assembly, **(b)** the 31 super-scaffolded pseudo-chromosome assembly with ChrU, **(c)** the super-scaffolded pseudo-chromosomes without ChrU (unplaced, unlocalised scaffold) and **(d)** a human genome assembly . The k-mer spectra for the human assembly was generated in the same way but using the GRCh38 genome and reads created from a sample taken from individual HG002 (Male, White, Ashkenazim Jewish). These figures all look very similar and show three things, firstly that almost all the information from the reads was captured by the original assembly (a), secondly making the assembly more contiguous by scaffolding into pseudo chromosome did not results in a large loss of information (a,b,c) and thirdly that the pink pigeon has similar levels of heterozygosity to the human genome (a,d) which can be seen by comparing the ratio between the height of the first red peak (representing heterozygote content) and the second red peak (representing the homozygous content)

A BUSCO analysis identified 93.1% of the 4915 BUSCO genes, which were present as single copy and complete when considering all 31 pseudo-chromosomes and the ChrU. The latter contains unplaced or unaligned scaffolds. When only considering the pseudo-chromosome, 91.3% of the BUSCO genes are identified, which still shows a high level of completeness.

The results from Abyss show that the super-scaffolding resulted in a highly contiguous assembly with an N50 of 78.4 MB.

Table 2.4 Comparison of assembly statistics produced by abyss v1.9.0

Name	n	Min	N50	E-size	Max	Sum	% 'Ns'*
Original assembly	443409	500	8076185	1.06E+07	3.87E+07	1.17E+09	2.361
Turtle Dove	357	5486	8.01E+07	1.02E+08	2.21E+08	1.18E+09	0.329
PP synteny**	32	23396	1.18E+08	1.22E+08	2.20E+08	1.27E+09	7.544
PP synteny***	31	23396	7.84E+07	1.02E+08	2.07E+08	1.05E+09	2.547

* Percentage of the assembly comprised of the ambiguous base 'N'. Calculated as the number of N's/Sum.

**This is the genome produced by synteny including all unlocalised scaffolds and unaligned contigs put into a single sequence record (ChrU), this means that the N50 for this assembly is inaccurate as it includes ChrU.

***This is the genome produced by synteny but only containing the 31 pseudo-chromosomes which are based on the 31 chromosomes of the turtle dove.

2.3.2 Variation

After filtering, 7,387,518 SNPs (and all individuals, n = 6) remained for analysis. The pink pigeon shows fairly even levels of variation across the genome, with only a few regions of low diversity. This pattern of variation is more closely resembled that of a large outbred population or one that may have experienced an historic decrease (see Figure 2.1), rather than a recently bottlenecked population. Each individual shows a similar pattern with notable regions of low diversity in chromosome 1,2,3,4,5. Individual S1272 which is the reference bird has lower levels of variation and shows some regions that are entirely homozygous, for instance much of chromosome 12.

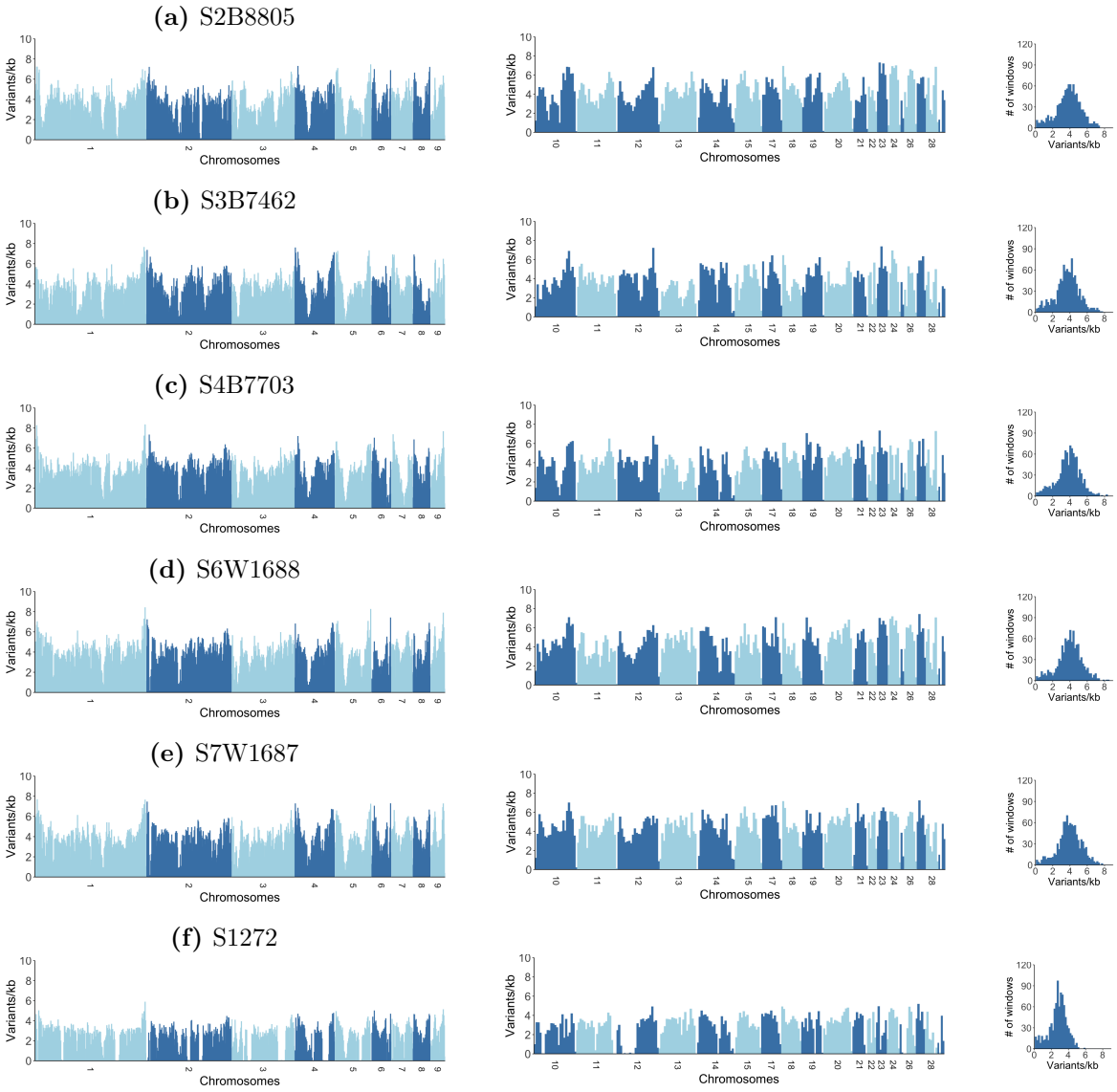


Figure 2.5 Distribution of heterozygosity across the genome, heterozygosity is calculated as number of variants per 1000bp measured in non-overlapping 1MB bins. For clarity the genome of each individual has been divided into macro-chromosome (on the left hand side n = 9) and micro-chromosomes (on the right hand side n = 21) alternating shades of blue have been used to distinguish adjacent chromosomes. Micro-chromosome 16, 29 and 31 are missing from the assembly and micro-chromosomes 25,27,30,32 and 33 are not labelled because there was no space to clearly label them. The histograms on the far right show histograms of per-window heterozygosity

The pink pigeon shows reasonably high levels of observed heterozygosity (H_o,mean proportion of heterozygous sites per individual genome)) compared to 40 other avian species (see Figure 2.6). The reference individuals (S1272) does have lower observed H_o than the population mean (n = 6) however it still has higher levels of H_o than many species with large

current population sizes such as the rock dove (*Columba livia* , > 20,000,000) and the Great crested grebe (*Podiceps cristatus*, >500,000 individuals).

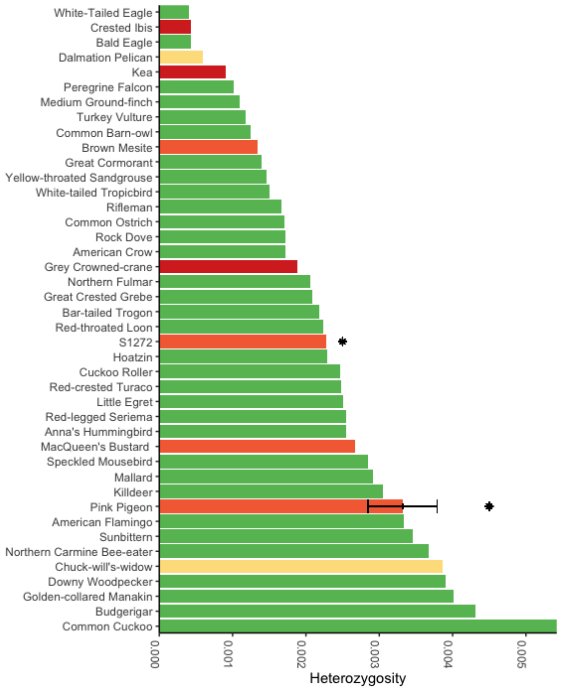


Figure 2.6 Observed heterozygosity (H_o) in different bird species the colour represents the current IUCN classification (as of September 2020), green = Least Concern, yellow = Near Threatened, orange = Vulnerable, red = Endangered. The data have been taken from [6] where the authors calculated heterozygosity as the proportion of sites that are heterozygous across a whole genome, further information about these data can be seen in appendix A.1. There are two columns representing the pink pigeon, the "Pink Pigeon" column is the mean heterozygosity of the six whole genome sequences and the individual that was used to create the reference genome (S1272) is also included separately, both these columns are emphasised by asterisks. add SD add link to appendix and how het calculated citation

2.3.3 Runs of homozygosity

The captive pink pigeon population has a mean inbreeding coefficient of 0.1360 ±0.1004, 0.0983 ±0.0537 or 0.2020 ±0.0544 as measured by F_{ped}, F_{RoH} and ZooRoH respectively (see Table 2.5). When considered by individual, the pedigree and molecular data show similar values (see Table 2.5) with some notable exceptions. Whilst all methods agree that individual S3B7462 is the most inbred, the pedigree value is nearly double that measured using F_{RoH} but agrees closely with the value calculated by ZOORoH. The individual S2B7805 has the lowest inbreeding coefficient when measured by F_{ped} but has the second highest when calculated empirically using F_{RoH} and ZOORoH.

Table 2.5 Inbreeding coefficients calculated from pedigree and molecular data.

ID	F_{ped}	F_{RoH}	ZooRoH
S1272	0.0744	0.1829	0.2191
S2B7805	0.0752	0.1048	0.1980
S3B7462	0.3487	0.1821	0.3114
S4B7703	0.0932	0.0616	0.1513
S6W1688	0.0815	0.0833	0.1780
S7W1687	0.0815	0.0595	0.1540
Mean	**0.1360**	**0.0983**	**0.2020**
Standard deviation	**0.1004**	**0.0537**	**0.0544**

The relationship between the size of runs of homozygosity and the time at which the demographic event occurred can be used to understand the processes that may have shaped the pink pigeons genome. To study this in the pink pigeon runs of homozygosity were classified as short (0.1 MB $\geq$ RoH $<$ 1 MB), medium (1 MB $\geq$ RoH $<$ 10 MB), and long (10 MB $\geq$ RoH $<$ 100 MB). No long runs of homozygosity were found in the pink pigeon (see Figure 2.7a), three individuals (S1272, S2B7805, S3B7462) showed some medium sized runs of homozygosity but for only S1272 did this constitute a substantial amount of the runs found Figure 2.7b. Most of the runs of homozygosity were short (0.1 MB $\geq$ RoH $<$ 1 MB).

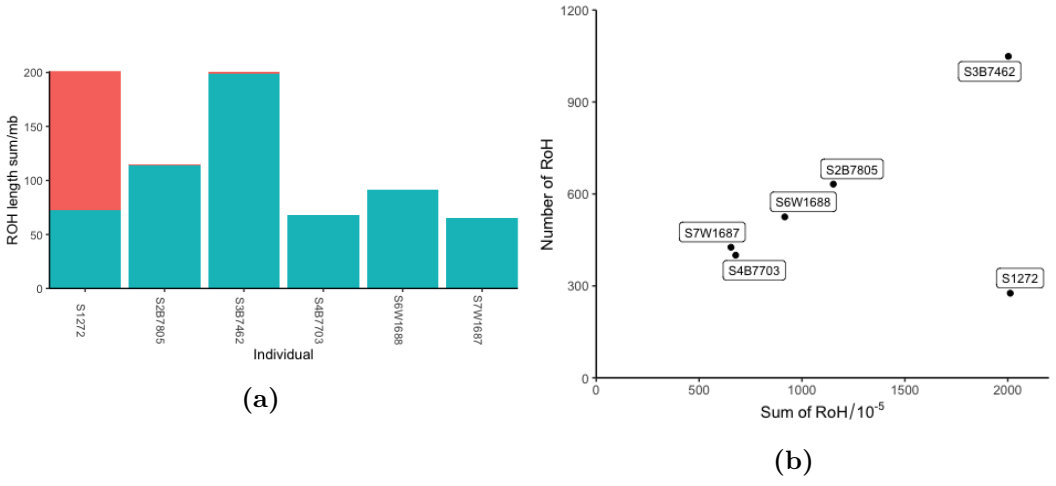
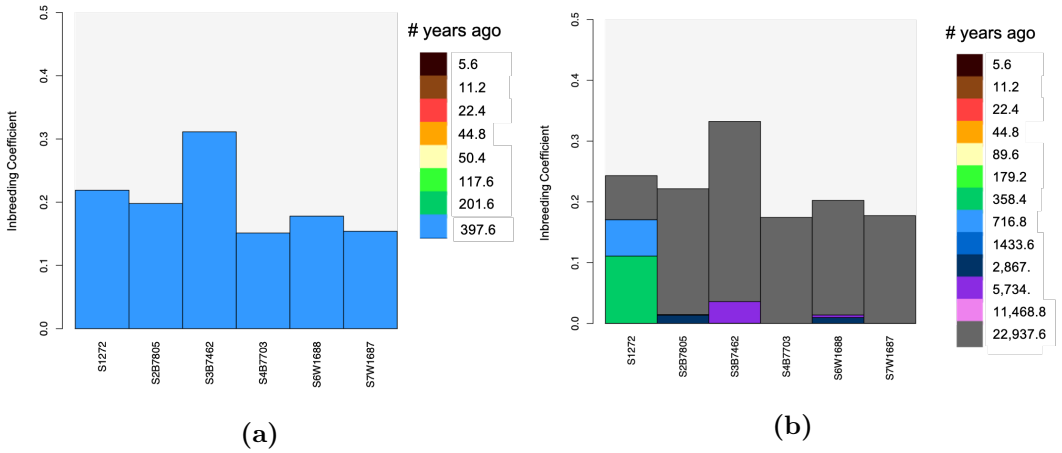


Figure 2.7 Runs of homozygosity in the pink pigeon. **a** the proportion of the pink pigeons genome that is contained within runs of homozygosity classified as short (0.1 MB $\geq$ RoH $<$ 1 MB), medium (1 MB $\geq$ RoH $<$ 10 MB), and long (10 MB $\geq$ RoH $<$ 100 MB). Short runs are in blue, medium runs are in red, there were no long runs present. **b** The relationship between the number of runs in each individual and the amount of the genome contained within those runs measured in base pairs (bp)

The result of no recent inbreeding events was confirmed using R ZooRoH which detected no recent homozygous by descent (HBD) segments. In the first model (see Figure 2.8a), aimed at identifying historic events that may have impacted the pink pigeon, all HBD segments were classed at k = 142 (71 generations). The results of the second model (see Figure 2.8b), aimed at identifying more ancient events that may have impacted the pink pigeon, shows that one individual (S1272) has HBD segments in class k = 128 (64 generations) about 358 years ago. Most HBD segments were classified as more ancient with HBD segments in k = 512 (256 generations, 1,433 years), k = 1024 (512 generations, 2,867 years) or k = 8192 (4096 generations, 22,937 years).



(a) **(b)**

Figure 2.8 Realised inbreeding coefficients from 7,387,518 SNPs using RZooRoH models. Each bar represents an individual (x-axis), displaying overall individual inbreeding coefficients (y-axis) and the proportion of genome assigned to specific homozygosity by descent (HBD) classes. HDB class values are assigned in concordance with the length of the run of homozygosity, where longer runs of homozygosity indicate more recent inbreeding events. The colour indicates the time to inbreeding event. These times were calculated from the modelled HBD classes, where a class is approximately double the generation time to an inbreeding event.

2.3.4 PSMC

The results from the PSMC indicate the pink pigeon went through two cycles of increasing then decreasing effective population size (N_e) see Figure 2.9. The first noticeable increase starts before the beginning of the Pleistocene (circa. 2.5 million years ago (MYA)) and reaches its peak with an N_e of 2 - 3 million in the Pleistocene. This is followed by a decline to an N_e of 500,000 around 1 MYA. The second cycle started less than 1 MYA with a sharp increase in N_e 5.5 - 12 million, followed by a sharp decline around 200 KYA to an N_e of approximately 250,000. This last cycle finishes just before the start of the last glacial period

(110 - 12 thousand years ago (KYA)). The shape of the PSMC graphs are relatively similar between individuals, which gives confidence in the demographic inference.

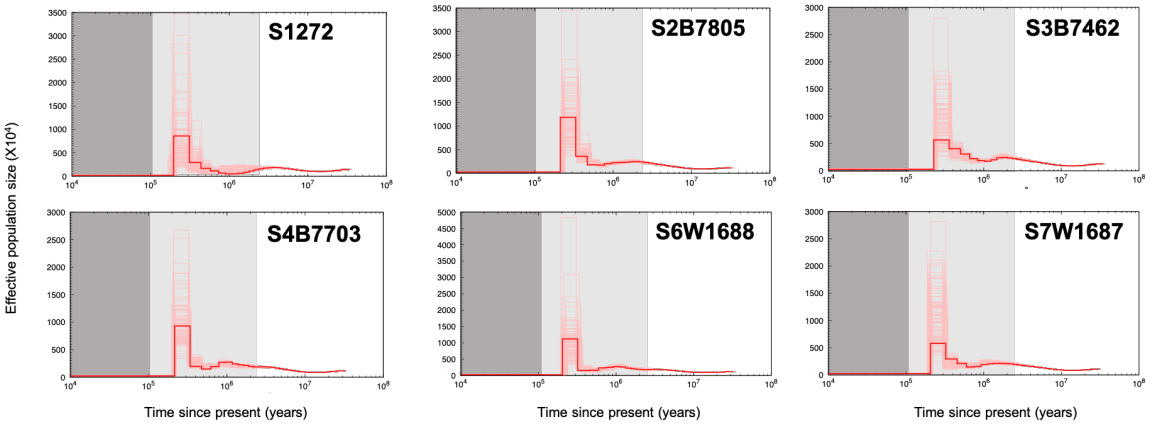


Figure 2.9 PSMC results for the pink pigeon these plots show that the effective populations size (N_e) of the pink pigeon went through two obvious cycles of increasing then decreasing N_e. The red curve is the PSMC estimate for the original data, the pink curves show the estimates for 100 bootstrapped sequenced. The shaded areas represent the Pleistocene period (2.58 MYA to 11.7 KYA 2,580,000) with the last glacial period (LGP) in darker grey (110 - 12 KYA). The N_e is estimated to reach a maximum of between 5 - 10 million individuals circa 200 KYA.

2.4 Discussion

Access to a high quality reference genome enables a greater breadth of analyses including the ability to provide an coordinate system for comparative research, measure processes across the genome that would otherwise be estimated and test pre-conceived ideas about a population [163, 63]. In this chapter the pink pigeon's reference genome was successfully super-scaffolded to create 31 pseudo-chromosomes that enabled analyses across the genome. Using this new reference genome model, genome-wide analyses were performed that revealed past demographic events that may have shaped the genome as well as a comparably large amount of variation and no evidence of recent inbreeding. These results are unexpected but encouraging because they show that the pink pigeon is less inbred and has more genetic variation than previously observed [9, 120, 2].

Using hybrid approaches for genome assembly should become more common with the increased number of chromosome-level assemblies produced by projects such as Vertebrate 10K (https://vgp.github.io/genomeark/), the DNA Zoo (https://www.dnazoo.org/). As well as the many research groups that are taking advantage of new technologies and the falling prices of sequencing [196]. Whilst it will likely be more accurate to produce a chromosome level assembly using long reads and physical mapping, this is not always an option for many

endangered species where funding may be limited or acquiring samples may be logistically challenging [137]. In this chapter the reference genome of the pink pigeon was successfully super-scaffolded into pseudo-chromosomes resulting in a genome model that showed a high level of contiguity and completeness. There are a couple of caveats, firstly the quality of the final assembly will depend upon the quality of the genome you are aligning to. The turtle dove genome assembly possesses 316 unplaced scaffolds and although pink pigeon reads mapped to these locations, these scaffolds could not be used in further analyses. Assessment of the k-mer spectra showed that some real content is present in these scaffolds, and these data have been lost from further analyses. Some of these scaffolds will belong to missing chromosomes, in the turtle dove genome the micro-chromosomes 16, 29 and 31 are missing. For this study, it did not matter that there were missing chromosomes, but it will restrict some more specific analyses. For example chromosome 16 in chickens contains the Major Histocompatibility Complex (MHC) [197] therefore it would be needed for studies interested in disease resistance, for example in the case of the pink pigeon looking for resistance to avian trichomonosis. Secondly the quality of the final assembly will necessarily depend on the levels of synteny between the two genomes, due to the high levels of synteny present in birds this is perhaps less of a concern than it may be with other taxa [198]. However preliminary alignments of pink pigeon reads to the turtle dove genome and several other high quality bird assemblies were performed that confirmed that this assumption was valid for the pink pigeon (see appendix A.1). There are several tools such as mummer [199] and online tools like genome ribbon [200] that make it is quick and easy to visualise and assess alignments. This highlights the utility of scaffolding by synteny to produce chromosome-level assemblies to enable whole genome analyses. Overall a high quality, pseudo-chromosome level assembly was produced where high quality is defined as an assembly that captures the majority of the information in the reads (Figure 2.4c), is highly contiguous (N50 = 7.84E+07), and in the right order (organised into chromosomes) [164]. Access to this high quality assembly enabled the first whole genome analysis of the pink pigeon and provides a useful resource for future research.

The reference genome enabled several analyses which used whole genome data from six captive pink pigeons these analyses showed relatively high levels of variation within the pink pigeon genomes. Based on previous research ([129, 120, 2]) there were two prior assumptions which were that the pink pigeon has low genetic diversity and is highly inbred. The pink pigeon declined to 20 individuals in the 1970s and the *ex situ* captive population was founded from a subset of those individuals over the course of several years (founders = 17). Since the beginning of both the *in situ* and *ex situ* captive breeding programmes there have been reports of inbreeding depression manifesting as skeletal deformities and poor fecundity [120]. An analysis of over a thousand zoo birds also detected a high genetic load (see Chapter 4). Furthermore, DNA fingerprinting indicated that the founders were related and a study of

mtDNA showed that there were only a few haplotypes [120]. Given this history and the starting priors it was initially surprising that the pink pigeon's k-mer spectra showed that they had similar levels of heterozygosity compared to a human genome. When variation was studied across the genomes (n = 6), levels also appeared high with few regions of extreme variation, and a near normal distribution of heterozygosity that was similar to the expected variation found in a population that had experienced a historic decline or a large outbred population [4]]. One individual (S1272) showed regions with pronounced low variation which could have been indicative of recent inbreeding [152] however recent inbreeding would have resulted in long runs of homozygosity that had yet to be broken up by recombination which were not found. Individual S1272 is same individual used to create the reference genome therefore it is possible that the increased homozygosity found in this individual is as a result of reference bias. This is where reads which are identical to the reference will be mapped preferentially to those with any differences, even if those differences are real polymorphisms [169].

Reference bias may also provide an explanation for the large discrepancy seen between the inbreeding coefficients calculated for individual S1272 using different methods (F_{ped}, F_{RoH}, ZooRoH). Inbreeding coefficients are a measure of the amount of an individuals genome that is identical by descent (IBD) and they can be calculated using molecular data (F_{RoH}) or pedigrees (F_{ped}). Whilst it is possible to accurately track inbreeding through a pedigree there are practical difficulties involved in maintaining accurate pedigrees. For example, unknown parentage and the assumption that all founding individuals were unrelated even though this is usually untrue [56, 201, 202]. These difficulties can lead to inaccurate estimations about the amount of inbreeding present in an individual or a population [203]. These difficulties can be surmounted by using empirical measures such as runs of homozygosity (F_{RoH}) which provide a quantitative method to measure inbreeding in individuals using molecular data [204]. The results from calculating the inbreeding coefficient from molecular data shows that the levels of inbreeding calculated are different depending on the method used to calculate them (see Table 2.5). ZooRoH and the vcftools (which was used to identify runs of homozygosity used to calculate F_{RoH}) use similar models to identify regions that are IBD but different algorithms to calculate inbreeding coefficients. ZooRoH uses the forward-backward algorithm to calculate the inbreeding coefficient [205] whereas F_{RoH} was calculated using equation Equation 2.1 using only segments $>$ 100 KBP, these differences may explain some of the discrepancies seen between the two measurements. The largest discrepancy in results is for individual S1272 which is the least inbred according to pedigree but one of the most inbred when using molecular methods. This may indicate a mistake in the individual's pedigree or the greater homozygosity (also detected in the variation plots) may be due to the reference bias as discussed above. The three measurements all agree that the pink pigeon captive

population is inbred and increasing the sample size may help to definitively quantify the amount of inbreeding present in the population.

Although the pink pigeon appears inbred there are no indications of recent inbreeding events in their genomes (i.e. long runs of homozygosity). However the estimates of inbreeding coefficients obtained from each method did vary, sometimes significantly. Both Vcftools and ZooROH use a hidden Markov model to identify runs of homozygosity and both use the forward-backward algorithm to calculate the probability of a segment belonging to a particular class. Unfortunately the documentation for vcftools –LROH function is minimal however it seems that the major difference between the two methods is that vcftools discards any runs shorter than 100kb [189]]. This may explain why zooROH calculated higher inbreeding coefficients than F_{ROH} if was able to include more segments in the analysis. When considering individual inbreeding coefficients there were discrepancies between the two methods when considering the absolute values but both methods agree on which individuals have higher or lower inbreeding coefficients (this is useful information for practitioners more than exact values) and both methods indicate that the runs of homozygosity present are not from recent inbreeding. Instead the majority of ROH were small ($<$ 1mb) consistent with a historic population crash or historic inbreeding events [5] and the results from ZooRoH are consistent with these observations. ZooRoH classifies HBD into different classes that can be used to calculate when in the past there were inbreeding events [205]. According to the models produced in this chapter the earliest inbreeding events happened 350 - 400 years ago which is around the time when Mauritius was first settled. But the majority of events that were detected were even more ancient (1,433,2867,22,937 years ago). In both models the majority of HBD segments belonged to the oldest classes and this may suggest even more ancient inbreeding events. Both models used pre-defined classes therefore ZooRoH can only classify segments in those classes when in fact the segments may belong to older events, this can be seen by comparing the results from the two models in Figure 2.8. Another consideration is that when calculating runs of homozygosity data is taken from across all chromosomes (micro and macro chromosomes) however it is known that the two chromosome types evolve at different rates (typically micro chromosomes have a higher recombination rate [177]) this may mask signs of inbreeding by breaking up runs of homozygosity [206, 177]. This may also explain why many of the low diversity regions seen in the pink pigeon are found on macro-chromosomes (see Figure 2.5). However as a preliminary test ZooRoH was re-run (using the same parameters as above) using only macro then micro chromosomes and in this instant both gave similar results (for all 6 individuals, mean F for macro-chromosomes = 0.220 $\pm$0.054, mean F for micro-chromosomes = 0.241 $\pm$0.103). Although these were only preliminary results and will require further analysis to fully understand the impact of chromosome type on calculating runs of homozygosity, they indicate that, in the present study, the results of inbreeding analyses were not simply a function of which chromosomes were

sampled. This increases the confidence of the current interpretations of the data presented that suggest the pink pigeon genome has been shaped by historic demographic processes.

The idea of a historic crash is supported by a PSMC analysis which shows that, like several other avian species, the pink pigeon experienced a sharp reduction in effective population size (N_e) during the last glacial period (110 KYA - 12 KYA) [194]. PSMC analysis are capable of identifying events that happened between 10 KYA and 10 MYA [192], this makes it an excellent model for studying the species of Mauritius, a volcanic island which was formed between 8 - 10 MYA [207]. The pink pigeon is believed to have originated from Madagascar and to have colonised Mauritius during the last glacial maxima when the sea levels were significantly lower [115]. Because Mauritius is a volcanic island, it had no native flora and fauna, which means that the initial colonising species faced relatively little competition. This allowed their populations to rapidly increase in size [208], which could explain the large increase in N_e seen around 220 KYA. After the initial increase the effective population size remained high for thousands of years before a sudden crash circa. 100 KYA. It is possible that the crash was due to sea levels rising stopping movement from Madagascar, the sea levels rose and fell throughout the ice age, however given the current timings the crash happened before the final sea level rise [115]. Another explanation is that the crash was caused by volcanic activity. Mauritius was an active volcanic island until about 25 KYA [207] therefore it is possible that there was a spate of volcanic activity that resulted in a population crash [208]. It is also possible that the increase in N_e reflects the ancestral population on Madagascar, and the crash represents the founder event of birds colonising Mauritius. This would imply that the pink pigeon are only relatively recent colonisers of Mauritius. Without further information, for example sub fossil evidence or an accurate mutation rate (which would provide more accurate timing for events), it is impossible to conclusively say what caused the historic demography. However it can confidently concluded that after a population expansion, there was a relatively recent historic population crash.

A study of genome wide data from the pink pigeon revealed higher levels of variation than expected and an apparent lack of recent inbreeding, instead it seems possible that the genome reflects ancient demographic processes. These are interesting observations and show the importance of assessing prior assumptions, however sane these assumptions may seem. The limitations of this study are discussed below before considering future work. There are two limitations to this study design which are the small sample size and no samples from wild individuals. Such constraints are common when working with endangered species however it does mean further work is needed to confirm these findings in more individuals and especially in individuals from the wild population. The sample size used for this study was small (n =6) and half of the birds were related to some degree according to the studbook estimated pairwise kinship Table 2.1. This means caution should be exercised before extrapolating these

results to the rest of the population. Relatedness may have influenced how certain SNPs were called during the GATK pipeline where SNPS are re-called depending on likelihood and levels within the population. Therefore what may appear a common SNP because it was seen in three birds may be a result of their relatedness. Despite this I think it is unlikely to impact many SNPS because they were first called on a per genome basis before population level data was used to adjust genotypes. Importantly the results from RADiKal (Chapter 3), which uses kmer based analysis on raw data RAD sequencing data from 180 wild pink pigeons, and indicates similar levels of variation across the pink pigeons genome. Although it is encouraging that the results from each individual were very similar suggesting that these patterns of variation, inbreeding and demography should be present in the rest of the captive population. It is possible that the wild birds show different patterns of variation and inbreeding but if the hypothesis about an ancient population crash, as shown by the PSMC analysis, is correct you would expect to see the same results from a wild pink pigeon genome. Secondly it is also possible that the current genomes reflect the intensive management present in the *ex situ* captive population, this intensive management led to a rapid increase in the number of captive individuals after the arrival of pink pigeons in captive collections. Using studbook information keepers were sure to pair individuals with as low mean kinship as possible to reduce inbreeding, although there have been instances of unintentional sib-sib matings (Harriet Whitford, pink pigeon studbook keeper, pers. comms.). Given the lack of long runs of homozygosity found in the pink pigeons, it appears that the management of the *ex situ* birds has been successful in trying to maintain a genetically healthy captive population. The wild population has also been intensively managed and descends from the remaining individuals found at Pigeon Wood, the majority of whom were used for *in situ* captive breeding. Similar to the *ex situ* programme, practitioners were mindful of managing the birds to reduce inbreeding. However, initially, it was more important to increase the census size as quickly as possible and therefore genetic health was, understandably, not the priority [120]. This can be seen by the calculated inbreeding coefficients and cases of inbreeding depression seen in the previous studies [129, 120]. Given these differences it is possible that the wild population (mainly derived from those *in situ* captive birds) may show different patterns of inbreeding and variation in their genomes and this would be interesting to test especially as management practices have been mindful of genetic variation between the wild and captive populations and therefore they have exchanged individuals between the two. Because founders and wild individuals were added over time, it may be that the influx of novel genotypes acted as a genetic rescue in the captive population. The influx of a single individual with novel genotypes can be detected as a reduction in homozygosity across the genome [152]. There are many possible explanations for the patterns seen in the pink pigeons genome, however, to test these hypotheses requires new sampling of both wild and captive individuals. Although the results from this chapter do not support the current

assumptions about heterozygosity and inbreeding in the pink pigeon, the genome provides a valuable resource that can be used to investigate inbreeding and heterozygosity further.

Heterozygosity is commonly used as a proxy for genetic health and adaptive potential however as more studies use whole genome data this proxy is being brought in to question. Several studies have shown that species with comparatively little heterozygosity have survived for millennia [146, 147]. A comparison of 41 avian species shows that there is a wide range of heterozygosity that appears to have little correlation with census size or Red List classification [6]. The pink pigeon which was reported to have low heterozygosity, based on genetic markers [129, 120] and RAD-seq data [2], shows surprisingly high levels of variation across the genome and no evidence of recent inbreeding. One possible explanation is that studies often use a single, condensed value of heterozygosity which is an average across the genome but does not represent any single part of the genome. This is why it is important to examine variation across the genome and may suggest that novel tools are needed that allow a population level view of variation across the genome. One such tool is presented in Chapter 3. As well as novel methods, novel approaches are needed for example due to the increase in sequencing and the ability to generate chromosome level assemblies it is now possible to study haplotypes and structural variation. In a recent study it was discovered that the adaptation of three sunflower species (*Helianthus annuus, Helianthus petiolaris and Helianthus argophyllus*), to different environments, was due to non-recombining haplotype blocks found throughout the genome [138]. Nevertheless using genetic diversity as a measure of a population's genetic health has enabled many successful conservation management plans for example picking source populations for genetic rescues [83]. It is also an intuitive measure and practitioners understand the need to preserve genetic diversity and manage breeding programmes accordingly [129]. In the absence of the appropriate data it must still be recommended to treat populations with low genetic diversity as compared to other similar populations as having poorer genetic health because the ability for a species to adapt is likely to be greater in a population with more diverse gene pool [54, 46]. More broadly, however, the mechanisms of adaptation are varied and complex therefore it may not be surprising that a single condensed value of heterozygosity alone is inadequate to study the phenomena.

A similarly complex phenomena is that of inbreeding depression even though the pink pigeon showed no evidence of recent inbreeding at the molecular level they have shown signs of inbreeding depression such as skeletal deformities and low fecundity [120]. If the pink pigeon has a very high genetic load then inbreeding depression may be possible without a large increase in homozygosity [36], therefore it would be useful to quantify the genetic load in the pink pigeon genome (see Chapter 4). This should be possible given the current genomic resources which would also allow the identification of putatively deleterious variants by comparison with an annotated bird genome such as the chicken. If causal variants behind

some of these negative traits and their distribution in the population are identified it may then provide information on the best way to manage the birds *vis a vis* translocations and genetic rescues [4, 156].

2.4.1 Conclusion

A high quality reference genome was created which will provide a valuable resource for future studies to try and understand the mechanisms of adaptation and inbreeding depression in the pink pigeon. The results from the current whole genome analysis are encouraging as they suggests that the pink pigeon is not as genetically depauperate as originally thought but has reasonably high levels of variation across its genome. This study also highlights the importance of whole genome data to ensure that past genetic studies are representative of the whole genome.

Chapter 3

RADiKal: a novel tool for generating an overview of genome-wide variation in a population at-a-glance

3.1 Introduction

Incorporating genetic data into conservation management is vital for the long-term survival of endangered populations [209, 210] however, the cost of generating whole genome sequencing is still prohibitive for many projects [12]. Instead, multiple methods now exist that sub-sample the genome creating 1000s of markers suitable for population genomic analysis but for a fraction of the cost of sequencing the equivalent number of whole genomes [211]. Restriction site associated DNA sequencing (RAD-seq) methods are an example of this, RAD-seq experiments use restriction enzymes to sub-sample the genome resulting in 100s - 1000s of markers genome-wide. This ability to sub-sample the genome at restriction sites combined with current high throughput technologies allows 10s - 100s of individuals to be sequenced simultaneously generating genome-wide population data. Although the development of RAD-seq methods has helped to provide access to genomic markers for projects in underfunded fields, including conservation genetics, new methods also bring new challenges. For example, whilst genomic data has been shown to provide more accurate information about endangered populations than genetic markers [79], it also requires more bioinformatic processing to prepare the raw data for analysis [136]. Bioinformatic pipelines are designed to remove errors

introduced during sequencing or during library preparation but the processing itself can remove information from the raw data and create significant biases [74]. These biases can impact downstream analysis so that the results of an analysis reflect the data processing steps and not the biological content of the markers [212]. Therefore new bioinformatic solutions are needed that extract as much information as possible from the raw data, increase the repeatability of experiments and give confidence that the resultant data, and any inferences drawn from it, are accurate. This is particularly important if the markers will be used in analyses that define management plans for populations of endangered animals. In these situations biased data could lead to poor management decisions resulting in the extinction of a population. The following introduction provides a brief description of RAD-seq methods and their importance in the field of conservation followed by an examination into the challenges of processing RAD-seq data and the consequences of using prior assumptions to test whether or not the results of data processing are correct. The software package RADiKal is introduced as a solution to these challenges because it provides a method to extract genome-wide data from raw unprocessed RAD-seq reads, thereby reducing the amount of information lost during processing. Using this information RADiKal produces painted chromosomes that provide an overview of the variation present in the genome of each individual in the population. This overview provides researchers with the ability to assess the levels of variation within the population at-a-glance and enables any prior assumptions, about the population, to be tested.

3.1.1 Brief history of RAD-seq

First published in 2007 Restriction Site Associated DNA sequencing (RAD-seq) is a form of reduced representation sequencing where a small proportion of a genome can be sequenced and 1000's of markers generated [213, 214]. RAD methods use one (or more) restriction enzymes that recognise palindromic motifs throughout the genome and digest DNA at those locations. The digested DNA is then sheared further, amplified, size selected and sequenced on a Next Generation Sequencing (NGS) platform (see Figure 3.1 for detailed description) [213–216].

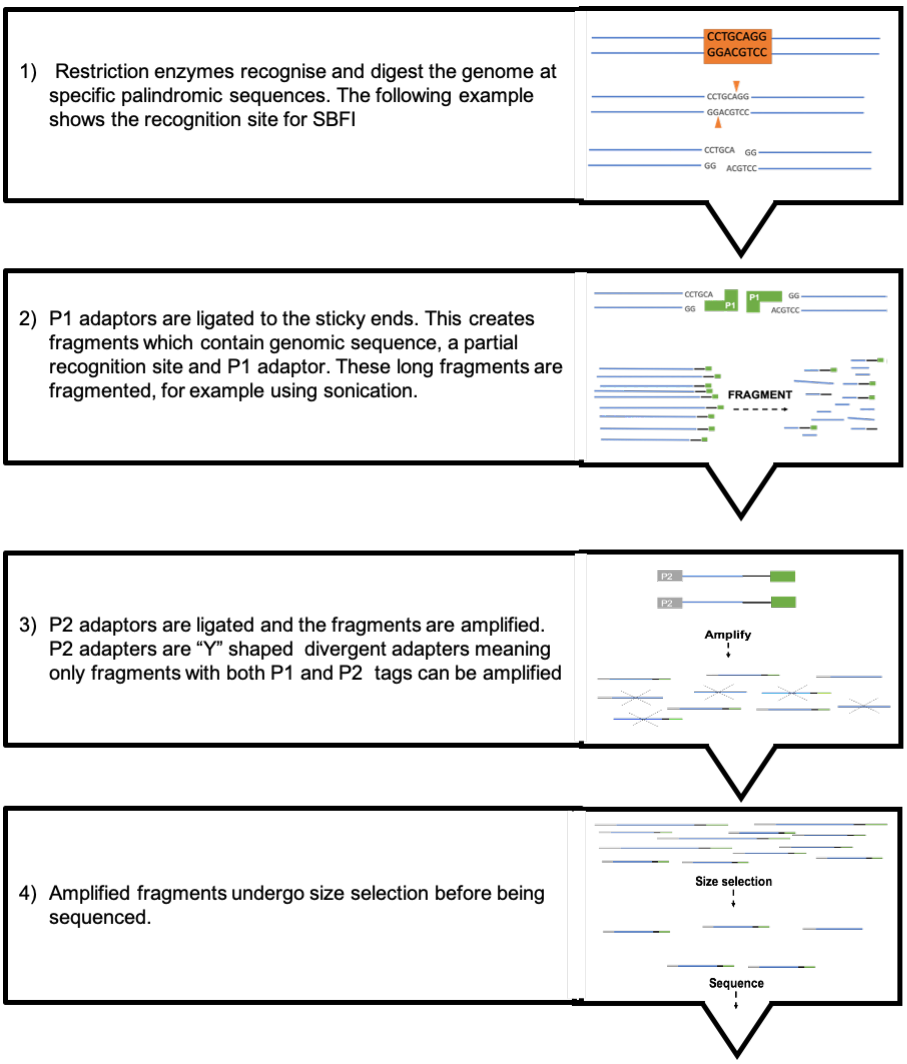


Figure 3.1 A description of a typical workflow for the traditional single digest RAD experiment.

RAD markers were originally developed as a method to improve genotyping capabilities, particularly in organisms that lacked a reference genome [217], by increasing the number of markers produced and reducing the costs [214]. The ability to generate so much data cost-effectively, without the need for a reference genome helped to expand the influence of genomic research to a variety of fields studying non-model organisms [218, 196]. These fields included population genetics [218], evolutionary biology [219], agriculture [220, 221, 196] and conservation genomics [222, 209].

3.1.2 Use in conservation and ecology

RAD sequencing methods have been a popular choice in the fields of conservation and ecology because (1) RAD markers can be analysed without the need for a reference genome or any knowledge of the genome sequence [223], (2) are relatively inexpensive, (3) give more information than traditional genetic markers and (4) some protocols can handle degraded DNA samples [224]. When studying endangered animals samples can be difficult to obtain therefore researchers may need to use poor quality, degraded samples which are unsuitable for whole genome sequencing but incredibly valuable in terms of the information they contain [137]. RAD-seq methods have been applied to DNA extracted from a variety of sources including: fins clips [225], snake scales [226], insects [223, 227], museum specimens [228, 224, 229], leaves [230, 231], blood meal [232], as well as more conventional sources like blood and tissue [233]. The popularity of RAD-seq comes from its ability to generate genomic data for non-model species (without a reference genome) and from its versatility. For a given budget researchers can select the sequencing strategy that will best answer their research question by altering the number of individuals sequenced, or the enzyme cutter [218, 211]. This versatility is reflected by the variety of studies in the field of conservation genomics that have used RAD-seq methods. For example the single nucleotide polymorphisms (SNPs) called from RAD-seq methods have been used in order to resolve phylogeny [233, 234], population structure [235] and conservation management units [236], examine levels of genome-wide heterozygosity [223, 233], assess the effectiveness of captive breeding [209], create panels for parentage analysis [237] , identify adaptive loci [100], study demographic history [223] and assess the consequences of genetic rescue [100].

As the list above indicates, RAD-seq has made valuable contributions to the fields of conservation and ecology in a relatively short time, however there are also a number of biases and challenges that must be considered when using RAD-seq methods [12, 238]. These biases include the lack of repeatability [215, 239]; allelic dropout [240, 241]; non-random sampling of haplotypes [240], sequencing error and failing to consider the impact of linkage [242, 238].

3.1.3 The Biases and challenges of working with RAD data

In the decade since the first study using RAD markers was released, over 600 papers have been published using RAD methods Figure 3.2.

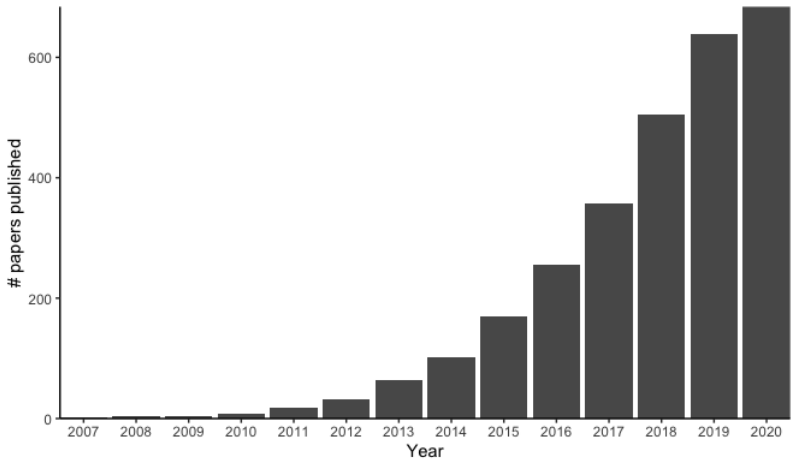


Figure 3.2 The increasing number of studies using RAD markers from its discovery in 2007 to current time. The following terms were searched for in Scopus (RAD sequencing OR RAD tag OR RAD markers or RAD-seq) (the term genotype by sequencing or GBS was not included in the search due to its ambiguity) and the metadata downloaded [accessed on the 10th April 2020].

Such widespread usage has led to the continued development and refinement of the laboratory protocols (see Table 3.1) and software used for generating RAD data [215, 243, 12]. The development of laboratory protocols has followed a trend of decreasing cost, increasing accuracy and versatility [12] while software developments have been focusing on the daunting task of repeatably identifying homologous sites across individuals and separating them from errors [244, 245].

Table 3.1 A selection of the most commonly used RAD methods. The choice of method will depend on the focus of the study, the study organism and the budget allocated to the project [12]. "Advantages" refers to advantages to a particular RAD method compared to the original RAD sequencing method.

Name	Brief description	Advantages	Ref.
RAD	Uses a single enzyme to digest the genome and broad size selection (see Figure 3.1).	NA	[213, 214]
ddRAD	ddRAD (double digest RAD) uses two restriction enzyme to generate fragments and has a specific size selection step	ddRAD was designed to increase repeatability, by generating precise fragment size, and reduce the cost of library preparation.	[215]
2b-RAD	Uses type IIB enzymes which cleave up and downstream of the recognition site creating short uniform tags between 33-36bp long	simple protocol, easy to adjust marker density and due to shorter markers require less sequencing to get equivalent depth.	[216]
RAPTURE	RAPTURE (RAD Capture) combines the use of restriction enzymes with a capture approach by separating library preparation and sequencing. Baits are designed on specific RAD locations, then the RAD library is created and the baits used to capture only complementary sequences.	Increases repeatability as baits target the exact same RAD sites in each samples and reduces clonal reads.	[246]
ezRAD	Uses frequent cutting enzymes to create fragment sizes compatible with Illumina TruSeq technology.	Can avoid amplification of fragments due to TruSeq technology, because TruSeq is commercially available digests can be sent to commercial facilities making it accessible to research groups without the technical expertise for library prep.	[247]
hyRAD	Uses RAD fragments (produced using ddRAD) to create biotinylated probes and to capture homologous fragments from shotgun sequencing.	Allows sequencing of degraded historic samples previously unsuitable for RAD, reduces allelic dropout and unlike Rapture doesn't require knowledge of target sequence.	[224]

As RAD-seq methods have gained in popularity the scrutiny of these methods has also increased [242, 73]. Researchers have discovered biases and errors attributable to the laboratory protocols followed, the bioinformatic tools used or inherent within the use of restriction enzymes [240, 248, 238, 242, 73]. An example of an inherent biases in RAD-seq studies is the assumption that the cut sites will be randomly distributed throughout the genome [214, 215]. This has been proven to be incorrect and instead studies have shown that the distribution and retrieval of RAD sites across the genome is non-random [248]. For instance a greater number of RAD sites in the Hawaiian amakihi (*Hemignathus virens*) genome are located on the micro-chromosomes [206] possibly due to a higher GC content [206, 249]. This uneven sampling of the genome needs to be considered when researchers want to use RAD loci to study adaptation because it is possible that entire portions of the genome will not be covered by markers or linked markers especially if linkage blocks are smaller than the gap between RAD sites [242].

Other forms of bias and error include: the lack of repeatability [215, 239]; allelic dropout [240, 241]; PCR clones [250]; or bioinformatic processing [239, 12, 74].

3.1.4 Bioinformatic processing

Bioinformatic processing is used to prepare raw data for data analysis and it is often used to reduce the impact of biases and errors, for example by detecting PCR duplicates [251]. However the decisions made when processing RAD data can influence the results of an analysis more than the biological information contained within the RAD markers [7, 74, 245, 252]. This is an uncontroversial but worrying statement which has obvious implications about the accuracy of results from RAD experiments. The foremost cause of this is the high levels of parameterisation required to identify RAD loci [253, 254] coupled with a lack of easy validation to confirm whether or not something is a RAD site. User-defined parameters provide access to the "black box" of software allowing researchers to make decisions about the processing of their data. The suitability of the chosen parameters are not judged on how well the output reflects the true biological input but rather on the output meeting biologically arbitrary thresholds [137] (see Figure 3.3). Overall the large number of possible parameters means that testing and deciding on a parameter set can be time consuming and also leads to problems of repeatability [7, 255, 256]. To try and mitigate these problems researchers have developed road maps [7] or software packages that automatically test multiple parameters sets [256]. But none of these solutions address the underlying problem that data is processed and judged based on abstraction, and that real information is inevitably being discarded. It should be noted here that not all thresholds are meaningless but that does not mean that they accurately reflect the biology of an organism, here this biology is their genomic sequence.

For example it is common to trim reads that come off an Illumina sequencer as it has been shown that sequence quality declines towards the end of the reads and therefore there are more likely to be errors in the sequence [72]. Such parameterisation does not reflect the biology of the sequence but is attempting to counteract a bias imposed by the sequencing method. Furthermore it is common to apply the same trimming to each read based on the mean quality score of the sequencing run for each read type which will likely result in the loss of real information from many reads whose quality is above the average. Finally even if the quality score is bad that does not necessarily mean that the base is wrong because the Illumina quality score represents the probability that a base was called incorrectly [257] . A recent genome assembly software (SDG) [164] is designed to use untrimmed raw reads to ensure the maximum amount of information is retained and instead use properties of the data (often kmers) that can more reliably distinguish errors from information to remove errors [164].

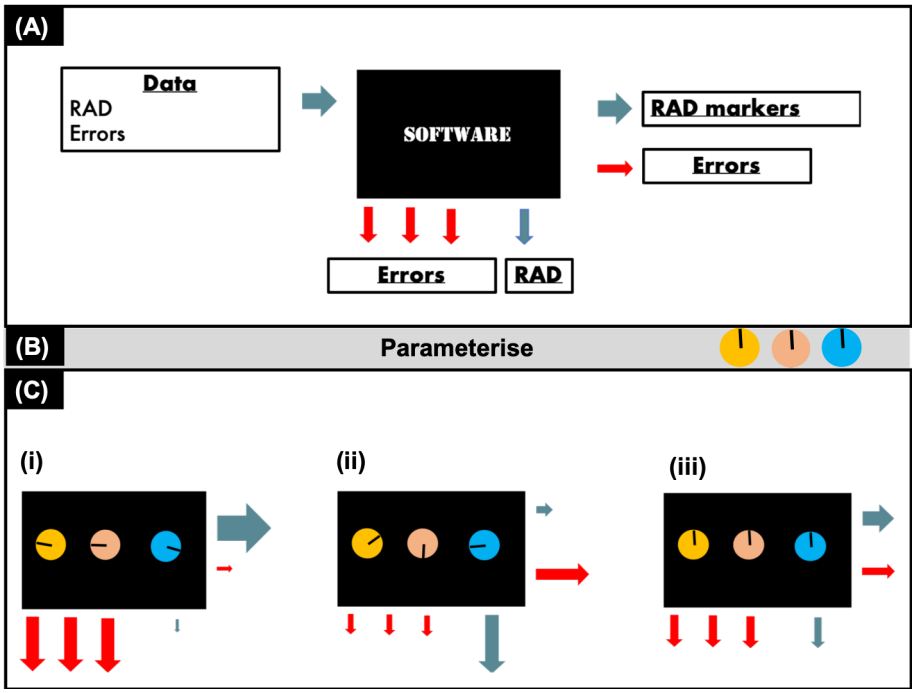


Figure 3.3 (A)Software is often described as a "black box" in to which data is put, some algorithms applied and then the results come out. With RAD software the algorithms are designed to identify RAD loci (here green arrows) and discard errors (here red arrows). In the diagram here everything underneath the box has been discarded due to the parameters set and everything to the right of the black box is the data that will be used in further analysis. (B) Parameters (here seen as three coloured dials) are designed to give the user some control over the functions within the "black box" in order to make sure that the algorithms are tailored to their particular data set. Optimising parameter settings is time consuming and there is no easy validation to help a user select the parameters that produce output data that best reflects the input data and captures the maximum amount of information. This is because the suitableness of different parameters is not judged on the information of the input data but rather on how the output data looks, often using thresholds with little biological meaning. For example (C) if the proportion of RAD data is judged by the relative size of the green arrows and the errors the red arrows, assuming the same input data, it can be seen that parameter set (i) gives the best data for downstream analysis because it retains the most information from the input data and removes the most errors. However in reality we do not know how much real information from the input data has been retained. Instead it may be that parameter set (ii or iii) passes the threshold commonly used to decide if the parameters produce "good" data, which is loci have to be present in 80% of the samples (R80 method, [7]]) but this does not tell you how much information has been lost from your input data or how many errors have been included.

3.1.5 Parameterisation and priors

A common method to assess whether the output of data processing has been successful is to compare the output to your prior assumptions (hereafter priors). Every experiment will start with some priors about the data or system that is being examined, and these priors can be useful for planning experiments. For example before starting a *de novo* whole genome

sequencing experiment, it is unlikely that the size of the target species' genome will be known, which makes it difficult to ensure that there is sufficient sequencing coverage. A reasonable prior would be to use the genome size of a closely related species for the necessary calculations then once the genome had been sequenced and assembled this prior would be tested. The results of the analysis or processing should be used to test your prior and see if it is accurate but the prior should not be used to confirm that your analysis is correct because this could lead to biased and incorrect results. Despite this, priors are often used as justification for the accuracy of bioinformatic processing, and the results this produces, without any evidence that the priors are correct [258]. When studying endangered species a common prior is that the species will have low levels of heterozygosity because, many threatened species have been shown to have lower levels of heterozygosity than their non-threatened counterparts [259]. However if a species is assumed to have low levels of heterozygosity and the RAD sites are processed, and a set of parameters chosen, because of this assumption you are biasing the data to be less diverse than it is, especially as RAD-seq analysis are known to underestimate diversity [258]. This could give an inaccurate measure of the genetic diversity of an endangered species and has the potential to interfere with the study of their adaptive potential if signals are lost and important regions of the genome overlooked.

The potential problems caused by using priors to assess parameterisation and the impact that parameterisation can have on the identification of RAD loci can be exemplified by examining how the most common pipeline, Stacks, processes RAD markers. Stacks is the most popular software for identifying RAD loci [260], it is a versatile software that enables users to process RAD sites *de novo* as well as with a reference genome and provides a number of tools for population genomics, phylogenetics and genetic mapping [261, 262]. Stacks gets its name from the algorithm it uses to recover RAD sites, which "stacks" short reads then, after applying a number of parameters generates RAD loci and alleles [261, 262]. Assuming a reference genome is not present then Stacks does the following (see **??** for default values used in data processing):

1. For each individual, stacking identical reads into putative alleles – a stack is only considered a "Stack" if it contains **m** or more reads.
2. Stacking alleles into putative loci, two stacks are considered to represent different alleles at the same loci if there are fewer than **M** mismatches between the two.
3. Stacks from all individuals in population are catalogued to provide a list of all loci and alleles in a population. If there were different alleles fixed in different individuals these will be considered different monomorphic loci as long as there are fewer than **n** mismatches. Stacks will merge the alleles recognising that they are two alleles originating from the same loci.

If a sequence or stack exceeds any of the thresholds set by the listed parameters they are discarded to, in theory, ensure that the final data set contains only RAD loci. Such an approach is designed to decrease errors but it also discards real RAD loci/alleles resulting in a reduced and biased data set [212]. In particular, highly polymorphic regions will be discarded as they cannot easily be stacked if there are too many differences between sequences [254]. This will bias the final data set resulting in lower diversity estimates and neglecting scientifically interesting and informative regions of the genome [258, 254]. This bias will not be obvious because it agrees with the prior assumption that endangered species have low genetic diversity.

If a reference genome is available another common method to process RAD-Seq is to map to a reference genome (this can be done as part of the STACKS pipeline as well) this has a similar problem to STACKS in that polymorphic regions that are more diverged than allowed by the mapping parameters will be discarded as being erroneous. Studies are beginning to assess the biological impact of such processing biases and although have found when studying pink pigeon found had made a significant difference to the conclusions being drawn (see case study below).

The pink pigeon a case study

RAD-Seq data was generated for 180 wild caught pink pigeons as part of a previous PhD project [2]. This data was processed by mapping the RAD reads to the reference genome using BWA mem [263] and then SNPs were called using SAMtools mpileup, BCFtools [263] and VCFtools [188] (details available in [2]). The results of this study confirmed the prior assumption that the pink pigeon has low genetic diversity probably due to the population bottleneck the species went through in the 1970s and their subsequent low population size (pre-recovery) [2]. At the beginning of this current PhD project the raw unprocessed RAD data was broken down into k-mers and the coverage of these k-mers calculated using KAT [182]. This coverage was then plotted against the reference genome (assembled as part of [2]) to have a better understanding of the diversity present in the RAD data before any processing had taken place (bar demultiplexing). An example of the results found can be seen in Figure 3.4 which showed that, even using data from only 11/180 individuals at a single site, multiple haplotypes (6) were present. Although only a single site has been presented in this case study this pattern (multiple distinct haplotypes) was seen throughout the genome during preliminary investigations. To have been able to detect so many distinct haplotypes, from a small sample of individuals, seemed incongruous the conclusion that the pink pigeon had low genome wide diversity therefore it seems likely that a significant number of real RAD sites (or haplotypes) had been lost that during bioinformatic processing.

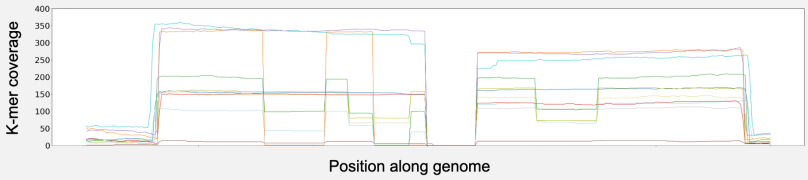


Figure 3.4 K-mer coverage at a single RAD site plotted for 11 wild pink pigeons. Where coverage is high it indicates that the individual is homozygous for the same allele as the reference genome it was compared to, where coverage drops by half it indicates that the individuals is heterozygote and has only a single allele matching the reference individual. Where k-mer coverage is 0 (apart from the region which represents the cut site) it indicates that the individual is homozygous for an alternate allele (compared to the reference genome).

In the above case study the method of processing the RAD data led to a misleading picture of the genetic diversity in the pink pigeon. However this is not an isolated incident and many studies have shown that how a RAD-Seq dataset is processed can alter the number of SNPs and the number of polymorphic loci obtained from that dataset [7**?**]. A few studies have sought to quantify the impact this may have on downstream analyses however results

have varied. Some studies have shown that the choice of parameters used in processing can alter the phylogenetic relationships recorded [264] and produce different patterns of population differentiation [260, 74]. Whilst other studies have reported that whilst the number of SNPs and polymorphic loci may differ depending on how the RAD data is processed these differences did not alter the results of downstream analysis [265]. Overall it appears the choice of parameters and the impact of processing is dependant on both the dataset and what analyses the data will be used for [260, 74, 258]. The question therefore remains - how can researchers consistently process their RAD data so that it resembles the biology of their study organism and not artefacts of the processing. One method would be to have a method that enabled researchers to verify whether the sequence data they were looking at derived from a RAD site or represented non-target sequencing error. If this was possible it would enable researchers to confidently separate errors from RAD sites without the need for strict parameterisation. The case study above showed a very simple qualitative method of visually assessing RAD sites - namely plotting the number of kmers present at each point across the genome. Such a method is not commonly used or integrated in to RAD processing so it could be argued that currently there is no method to visually assess raw RAD data to a) provide an easy overview to assess priors, inform experiments and examine biases and b) confirm that the information being captured is RAD data and not errors. Therefore a novel approach is needed [266] to provide tools that allow researchers to easily and visually assess their RAD data to ensure that valid RAD sites are being reported and provide a method to examine any prior assumptions.

3.1.6 A RADiKal new idea

RADiKal (RAD sequencing investigative Kmer analysis) is an alignment free software package that has been designed to identify RAD sites. RADiKal uses raw (demultiplexed) reads produced by RAD-seq and only a few basic prior assumptions to produce painted chromosomes that provide an overview of the variation within a sequenced population in a single figure. The output produced by RADiKal can then be used to visually assess RAD sites, examine the diversity in a population and allow more accurate downstream processing of the data. Ultimately RADiKal provides an interface that allows users access to the large amount of information, present in the raw RAD reads, that is often discarded by RAD-seq processing software.

3.2 Methodology

The scripts needed to run RADiKal and produce the analyses performed in this chapter can be found, along with an example of RADiKal's output in appendix A.2.

3.2.1 Data

RAD-seq data from 87 wild pink pigeons that had been generated as part of a previous study [2] was used. For this data DNA was extracted from blood and restriction digested using high fidelity SbfI (New England Biolabs). The resulting paired end (2 x 150bp) libraries were sequenced on an Illumina Hiseq 2500.

RADiKal requires a reference genome and for this study the pink pigeon (*Nesoenas mayeri*) pseudo-chromosome (n = 31) genome assembly produced in Chapter 2 was used. This genome was produced by the software RagTag v1.01 [180], which identified the syntenic regions between the pink pigeon reference genome (for details on the original reference genome see [2]) and the turtle dove (*Streptopelia turtur*). The turtle dove was chosen because it is the closest relative of the pink pigeon with a chromosome level assembly and the two species showed high levels of synteny.

The levels of variation from the RAD-seq data displayed by RADiKal were compared to those seen in whole genome data from a population of captive pink pigeons (n = 6). These are the same data that are described in Chapter 2, libraries for each bird were prepared using 10x Chromium technology and sequenced on an Illumina Novaseq using paired end (2 x 150bp) reads.

3.2.2 A RADiKal algorithm

The aim of any software program that processes RAD-seq data is to separate the signal present in the data from the noise caused by errors so that homologous RAD sites can be confidently identified in different individuals. Most RAD software map whole reads to identify RAD sites instead, RADiKal decomposes all reads into k-mers (sequence sub-strings of length k, default k = 27) and projects those k-mers onto the genome. The correct k-mer value is one that provides sufficient resolution with the larger the k value the more distinct the kmers are likely to be but this will reduce k-mer coverage at each site [182]. Throughout k = 27 is used for analyses in the k-mer analysis toolkit (KAT)although this is the default k-mer size

used by KAT it has also been shown to provide good resolution for genomes of a similar size to the pink pigeon or slightly larger such as the human genome [182].

K-mers have a number of useful properties that make them well suited to classifying RAD sites and these are listed below:

1. k-mers are computationally efficient
2. For ever single read of length L, L - k (+1) k-mers are created. The decomposition of reads to independent k-mers reduces the size of the sampling unit. This is a useful property because if there is an error in a read then that whole read will be discarded however if a single k-mer is in error then all the other k-mers (that do not contain the error) from that read are still retained. This maximises the information that is retained, amplifies signals in the data and eliminates the need for trimming reads, or parameterising to ensure errors are not included.
3. Errors are easier to detect using k-mers because sequencing errors are expected to be random and occur at low frequency. This means that k-mers resulting from errors have a distinct distribution from the main distribution of k-mers see Figure 3.5.

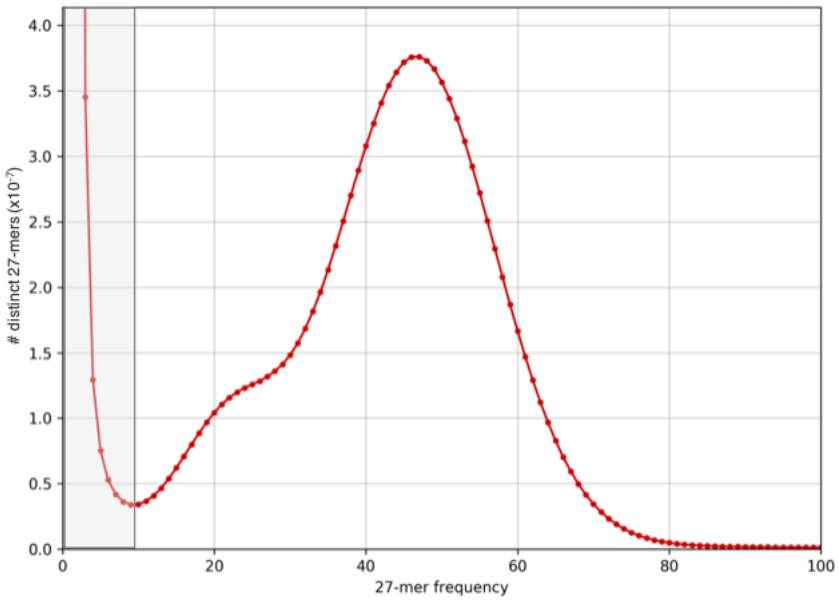


Figure 3.5 A k-mer frequency spectra generated by counting all 27-mers (k-mers of length 27) present in the pink pigeon reads. The distribution to the left which is emphasised by a grey box represents k-mers that are likely errors - a lot of distinct k-mers that only appear a few times.

4. K-mers are not mapped but rather projected on to the genome as counts. All the k-mers in the reads are added to a hash table and the number of times they occur are recorded.

Then for a given position in the genome these counts are projected so that if the k-mer "ATG" appeared 30 times in the hash table whenever the sequence "ATG" occurs in the genome the number 30 will be recorded instead. This projection eliminates the need to use a mapper and all the associated parameterisation and data losses that can occur [267]. For example mappers are most sensitive to highly polymorphic data which will not map well to a genome region and therefore may be discarded, artificially reducing heterozygosity.

5. K-mers can be represented canonically, which means they are strand neutral. If k-mers are canonical then both the forward and reverse k-mer sequences are counted and the counts added to whichever k-mer is the lexicographical smaller. K-mers can also be represented non-canonically (strand specific). RADiKal uses non-canonical k-mers which mean k-mers from the forward and reverse strand are different and therefore retain information about strand orientation. Although the genome is not phased this information can be used on a site-by-site basis and allows us to use this information to ensure the sequence comes from a valid RAD site which would have sequence present on both strands.

An overview of RADiKal's algorithm is presented in Table 3.2. Briefly RADiKal uses population level data to create a classifier based on information from every putative RAD site in the genome. This classifier is then used to score every RAD site in each individual of the population. Whereas many software packages offer a static model, which must then be used for any species regardless of its genome diversity, RADiKal creates a new classifier for every data set so that the classification reflects the biological diversity of the organism being studied as well as the sequence quality and characterisation of the data set.

Table 3.2 Overview of the RADiKal algorithm

Input:	All raw (demultiplexed) read ones catted into single file (.fastq) Raw sequencing reads from all individuals (.fastq) Reference genome (.fasta) Enzymes cutter motif
Output:	Scores for each individual in the population, at each putative RAD location across all chromosomes

Steps:

1. Load reference and genome and create reverse complement of the reference genome.

2. Load read files and create non-canonical k-mer counts of reads on both the forward and reverse complemented genomes.

3. Find all the putative RAD sites in the forward reference genome and return coordinates for each site to create a window around the putative RAD site

4. For each putative RAD site get the k-mer projection for the window on both the forward strands.

5. For each putative RAD site get the k-mer projection for the window on both the forward strands.

6. Run checks on the putative RAD sites, normalise the k-mer counts to avoid the impact of unequal sequencing coverage across the genome and individuals. Do this for sites on both strands.

7. For each site, on each strand, create a scoring matrix by filling in an empty scoring matrix using the normalised k-mer coverage.

8. Generate the same number of putative non-RAD sites as putative RAD sites by finding non-RAD windows and repeating all of the above steps except the checks. This is so we have some data to show our classifier what a RAD site is and some to show it what a RAD site is not.

9. Create a relative frequency matrix for both strand for the RAD by summing all the putative RAD scoring matrices from the relevant strands then calculating relative frequency of scores at each position

11. Final scoring matrix (classifier) created by subtracting the non-RAD from putative RAD relative frequency matrices for both the forward and reverse strands.

For each individual:

12. Get k-mer projections for each putative RAD site on the forward and reverse strands, these are the same sites that have been used to make the classifier.

13. Use the k-mer projections and the scoring matrix to score each site for each strand the total score being the sum of the two.

Dependencies:

SDG [164]
Python3
R v3.6.1

3.2.3 Constructing the classifier

RADiKal classifies RAD sites using information about strand orientation and the pattern of k-mer coverage at putative RAD sites. To identify this pattern the classifier is built using both RAD and non-RAD sites, however because we do not know which sites are RAD to start with, RADiKal uses three priors to select putative RAD sites.

1. RADiKal searches the genome for all instances of the cut motif, because RAD sites in the genome must have this motif.
2. The k-mer projection at that location for both strands, must not be composed of all zeros - i.e. there are no k-mers covering that region. If either the forward or the reverse strand is composed of all zeros then the site is excluded
3. There must be more than a single k-mer in that region and this must be true for both strands.

Whilst any site with only a single k-mer (or no k-mers) is obviously not a RAD site it could be argued that a site with only two k-mers present is also unlikely to represent a real RAD site. Deciding a threshold for how many k-mers need to be present to constitute a RAD site runs the risk of excluding highly polymorphic sites and RADiKal aims to keep as much information as possible with as few thresholds as possible. It was therefore decided that sites that were obviously not RAD sites (no coverage, or single k-mer coverage) would be

excluded from the construction of the matrix and accept that there would be some putative RAD sites that were not real RAD sites. Preliminary testing showed that this did happen but the non-RAD sites initially described as true RAD sites were; (1) in the minority, (2) did not impact the effectiveness of the classifier and (3) could be removed and the classifier regenerated without using those sites.

All the sites that have the three priors listed above are added to a list of putative RAD sites and are used to construct the classifier. Non-RAD sites are random locations across the genome but represent the same length of sequence that a RAD site covers. RAD sequencing tends to cover about 1% of the genome therefore whilst it is possible that a non-RAD site position may randomly land in a true RAD site it is unlikely. Preliminary results show that this is rare (a partial site was only seen once or twice across all tests which included 100s - 1000s of sites) and does not effect the functionality of the classifier however, as before these sites can be identified by the classifier itself and removed if necessary and the classifier re-generated without them.

RADiKal creates two classifiers, one for the forward strand of the genome and one for the reverse strand, this is possible because of the use of non-canonical k-mers which retain information about strand specificity. For each site a score for both the forward and reverse site's sequence is generated and a total score calculated by summing the two scores together. It is this total score that RADiKal uses when assessing the efficacy of the classifier and generating plots.

3.2.4 Scoring individuals

RADiKal generates scores (sum of scores from both strands) for each individual at each of the putative RAD sites used to construct the classifier. This method does not score RAD sites that may be present in the reads but are absent from the genome. Instead it ensures that homologous sites are scored in each individual in the population and this provides a robust data set with which to examine population level genome-wide diversity.

3.2.5 Painting Chromosomes

RADiKal extracts the signal from the raw RAD data using a classifier to score individuals at homologous sites. There is great potential for how these scores can be used but initially they are projected onto pseudo-chromosomes which gives the user an instant overview of the global diversity within the population and any differences between individuals or chromosomes. To amplify any signals and observe changes in variation RADiKal divides each chromosome into

50 bins of equal lengths, takes the mean scores from each bin and then bins those scores into ten distinct levels and plots the results (0-10). The number of bins was chosen after preliminary trials showed that this produced the clearest global signal however, this signal may vary for different species therefore, it may be necessary for the user to alter these values for other experiments. Altering the number of bins is akin to focusing a camera to try and find the clearest signal, too many bins will dilute the signal at any given region as a mean will be taken across many sites conversely too few sites may not give a representative signal of a given area. The number of bins needed will likely be slightly different for each dataset (although this requires testing) and be impacted by the spread of RAD sites across the genome.

3.2.6 Comparison with whole genome data

One of the aims of RADiKal is to be able to assess the levels of genetic diversity within a population from raw (demultiplexed) RAD-seq data therefore it is necessary to test whether RADiKal is able to recover the same signal as whole genome data. To produce the data for comparison the data was processed using the same pipeline reported in Chapter 2. First the Chromium 10x reads were processed to remove their barcodes using process_10xReads.py v0.0.2 (https://github.com/ucdavis-bioinformatics/proc10xG.git). The reads from each individual were then treated as 150bp paired end and mapped to the reference genome using Bowtie2 [186] using a very-sensitive local alignment allowing for a single mismatch. Once mapped, duplicates were marked using GATK's MarkedDuplicates v4.6.1.0 [187]. haplocaller combine genotype Next variants were called for each sample using samtools v1.7 mpileup. Each individual's genotype calls were then filtered by depth using to vcftools v0.1.13 [188] where the minimum depth was set to five and the maximum depth was calculated for each individual as two times the mean depth. After each individual had been filtered separately for depth the separate files were merged using vcftools *merge* and a final round of filtering performed that ensured only biallelic SNPS with quality scores of at least Q20 remained.

The global diversity was measured using vcftools v0.1.13 [188] and the number of variants per kilobase in non-overlapping 1MB bins (–SNPdensity 1000000). Regions of interest were examined further using a standard nucleotide basic local alignment search (blastn).

3.3 Results

3.3.1 Spread of scores

RADiKal identified 35,333 putative RAD sites across the reference genome having projected the k-mer coverage from the read ones from every individual. Of these 393 (false positives = 0.011%) scored below 0 and no putatively non-RAD scored above 0. The scores from the sites used to create the classifier were plotted (see Figure 3.6)to show that RADiKal can successfully distinguish RAD from non-RAD . These sites are putatively RAD or non-RAD therefore it is possible that a putative RAD site is not a RAD site and therefore has a score below 0. This may happen if the cut site sequence was present in the genome but was not present in any of the other individuals in the population. Similarly if there was a mutation in a cut site in the reference genome it would be considered a non-RAD site but other individuals may have the non mutated RAD site.

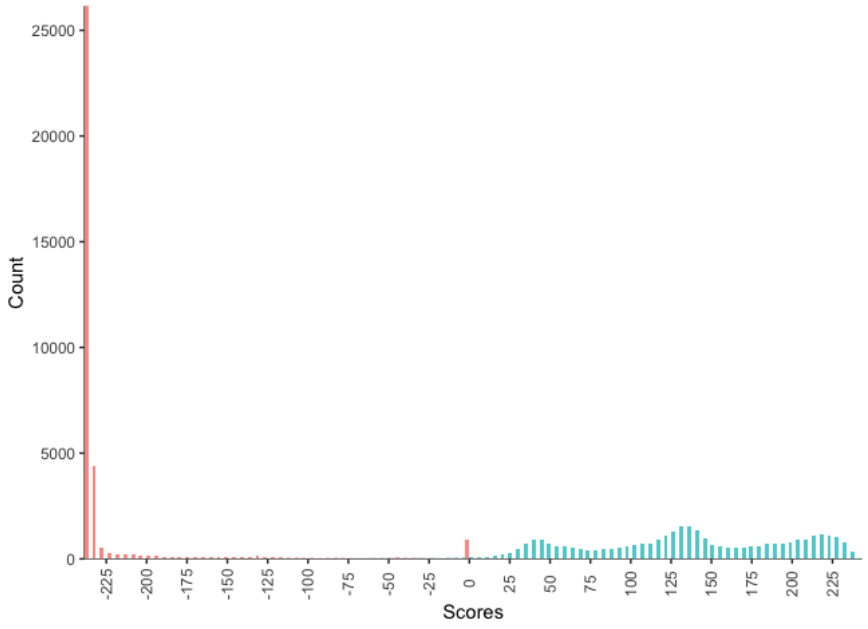


Figure 3.6 Spread of scores for putative RAD (blue) and non-RAD (red) sites across 31 pseudo-chromosomes

There is a three peaked distribution of scores for putative RAD sites, which broadly reflect the three possible states of alleles in comparison to the reference - both alleles homozygous for the reference allele, heterozygous, both alleles homozygous for an alternate allele. By generating heatmaps it is possible to see that the scores themselves provide some information about heterozygosity and similarity of any individual to the reference individual at any given

site. For an example see Figure 3.7 which shows heatmaps of the scoring matrix for sites with scores in each of the three distributions (0 - 88; 88-175, $>$ 175. These scores were chosen based on Figure 3.6). The number of sites in each distribution were counted, see Figure 3.7e, which showed that the majority of the sites fell in distributions that were heterozygous compared to the reference allele. This suggests significant levels of heterozygosity in the wild population.

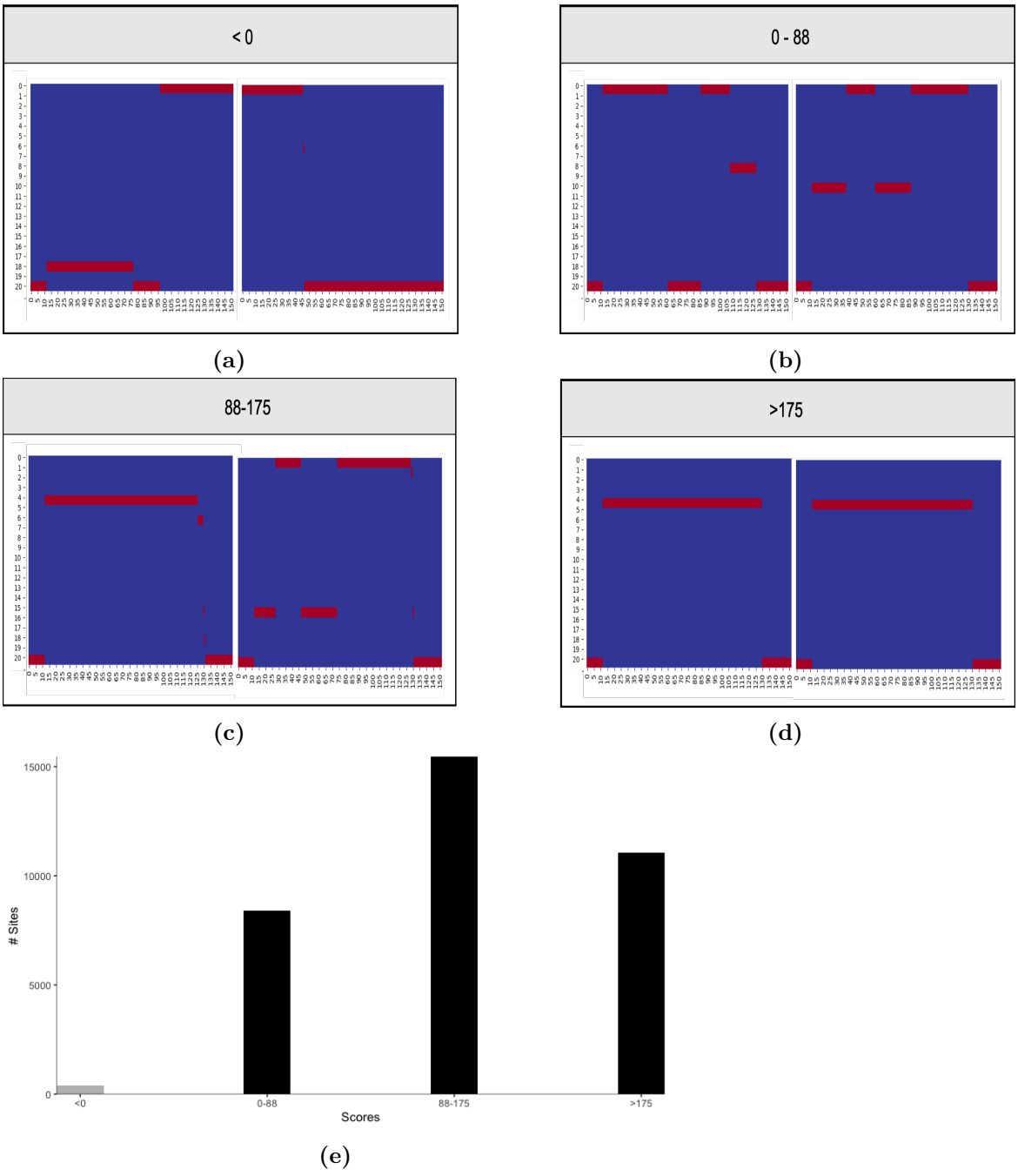


Figure 3.7 Visual representations of the scoring matrices and the approximate scoring that created them from four RAD sites from the pink pigeon genome where red represents the bin that was scored and blue where there was no k-mer coverage, the scoring matrix for the forward and reverse strand are on the left and right respectively. Each of the four sites is an example of different scores **(a)** represents a site where k-mer coverage was high across the site, indicating it is identical to the reference sequence at that site. In image **(b)** the forward strand is identical to the reference genome but the site itself is heterozygous with there being SNPs present in the reverse strand sequence. This can be seen as a reduction in coverage and therefore indicate it is a heterozygous SNP with some individuals having the reference allele and others an alternative allele. Image **(c)** is the most polymorphic site and shows heterozygous SNPs with reduced coverage, about half, and a homozygous SNP in the forward strand where the k-mer coverage drops to nothing as there are no k-mers that match the reference sequence at this site. **(d)** shows a non-RAD site which has a negative score and clearly differs from the other RAD sites. The number of sites in present in each distribution can be seen **(e)** including the number of false positives (putative RAD sites with a score of less than 0) which is shown in grey.

3.3.2 Spread of sites

It is important to examine the spread of RAD sites across the genome to test the commonly held assumption that RAD sites are randomly distributed across the genome. Figure 3.8 shows that the density of RAD sites is not even across the reference genome but that the smaller micro-chromosomes have a higher density of RAD sites compared to the larger chromosomes.

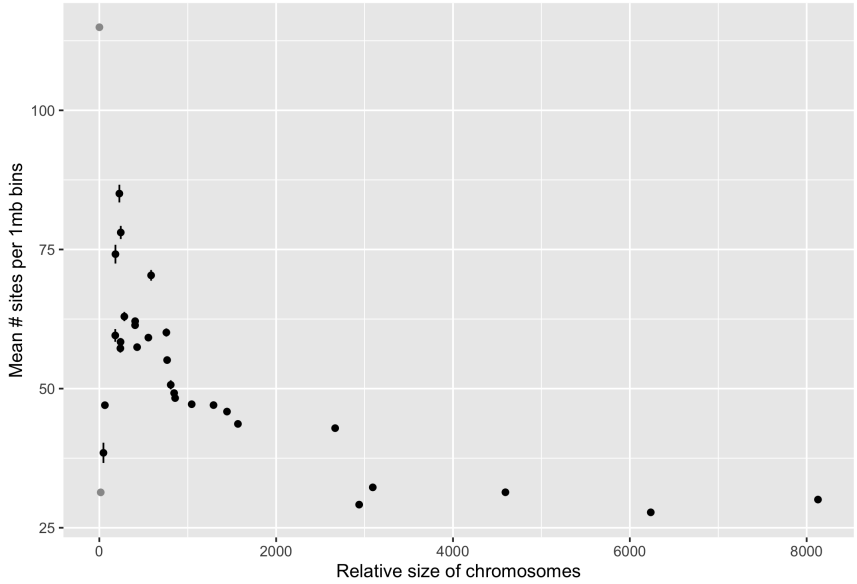


Figure 3.8 Relationship between chromosome length and RAD density. RAD density is measured as the mean number of sites per 1MB non-overlapping bins and sizes are relative to the smallest chromosome (chromosome 32 which is 26,103 bp). Grey points are those whose chromosome size is less than the bin size and the number of mean sites per 1MB was extrapolated from the number of sites present in the chromosome and therefore unlike the other sites no standard errors measures are presented.

3.3.3 Painted chromosomes

The results from RADiKal show that the wild pink pigeon population has a good amount of heterozygosity because the majority of scores represent sites that have some level of heterozygosity compared to the reference genome (see Figure 3.9). RADiKal also shows regional differences, for example some chromosomes have large blocks of homozygosity (represented by blocks of purple or blue) or regions of heterozygosity (regions in red). This already provides users with information about the study species' genome which they can choose to examine in greater detail if it is of interest. For instance, scores from regions of the genome that show extreme variation can be extracted and examined by plotting individual

RAD sites (as in Figure 3.7, which provide a visual means to assess whether a site is a genuine RAD site or not), or by using the coordinates from those regions to extract sequences for further analysis.

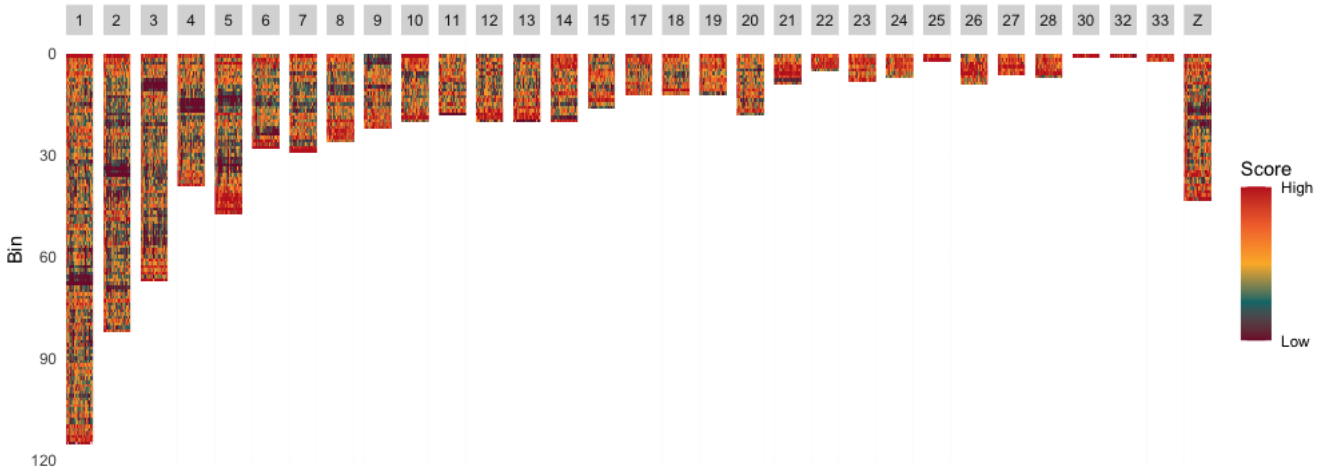


Figure 3.9 31 painted pseudo-chromosomes. Each individual is plotted along the x axis having first been clustered by similarity to ensure patterns in the population are visible, and the y axis shows which bin the scores fall in. The colour represents the amount of variation in any bin ranging from low variation (purple) which represents alleles that are homozygous for the reference allele (purple) to high variation (red) where both alleles are different to the reference.

The number of sites per bin was also painted to examine the number of sites per bin (see Figure 3.10). This was to ensure that any possible patterns were not an artefact of sampling. Overall there is a fairly uniform distribution between and within chromosomes, which is to be expected because of RADiKal's method of plotting the data. There is some variation, most notably the last bin of most chromosomes appears to show a decrease in the number of sites. This is due to the way the chromosomes are binned which means the remainder of sites are always placed in the last bin.

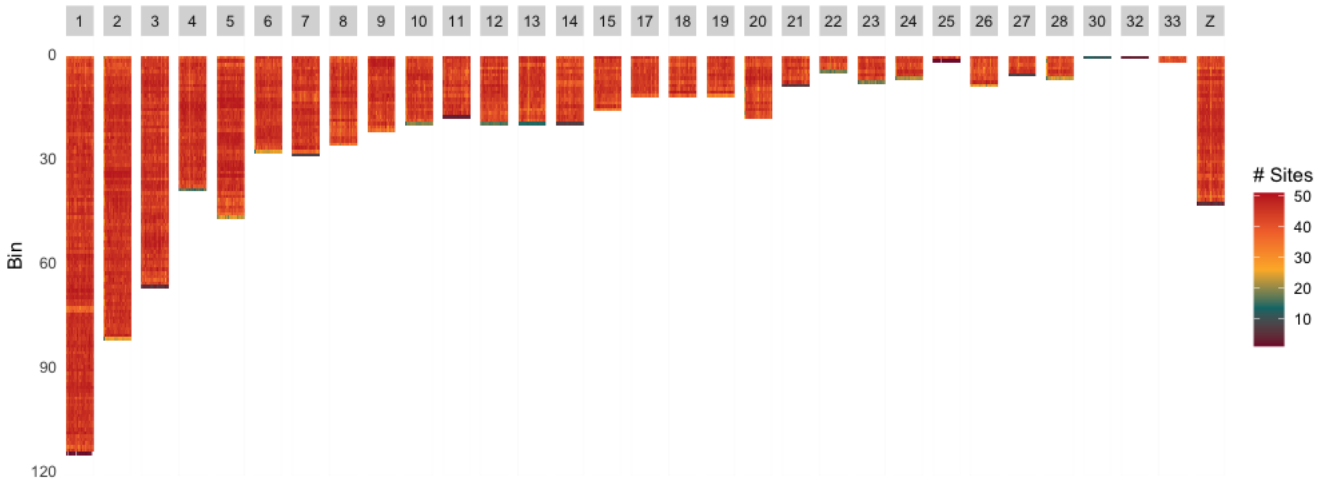
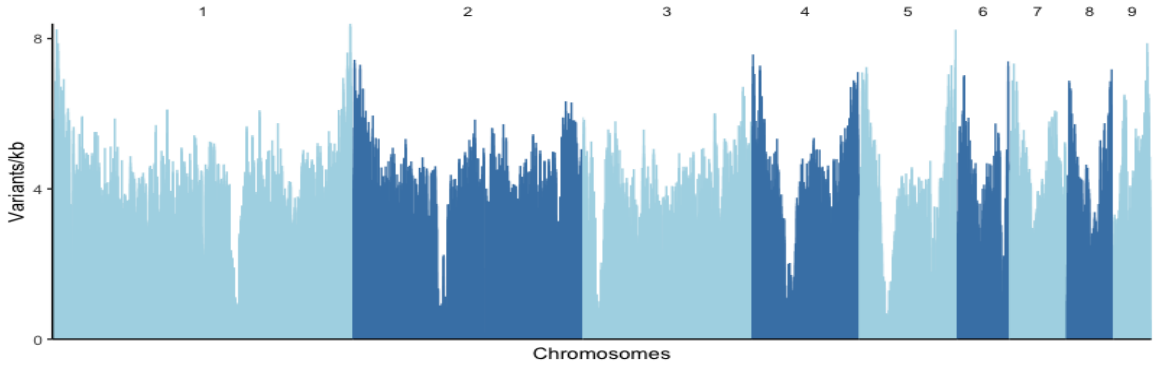


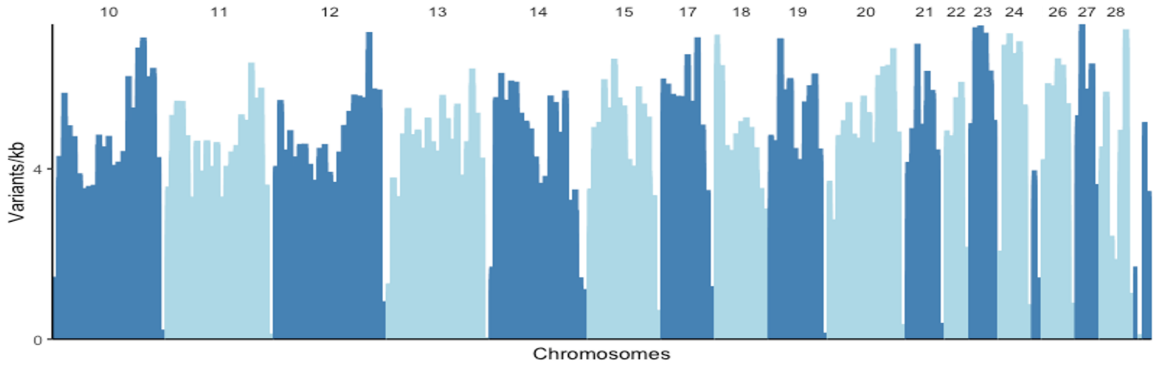
Figure 3.10 31 pseudo-chromosomes painted with the number of sites per bin used to plot RAD scores on pseudo-chromosomes

3.3.4 Comparison to whole genome data

RADiKal identifies the main signals present in the whole genome data (see Figure 3.11) demonstrating that RAD data contains sufficient evidence to detect signals of high and low diversity across the genome. Chromosome five provides a specific example (see Figure 3.12) of how RADiKal detects the defining signals present in the population. In chromosome five there are two distinct regions of low variation. The first region with low variation was blasted (blastn https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) against the chicken genome (*Gallus gallus*) and the best result, with over 90% identity, was the protein coding gene BR serine/threonine kinase 1 (BRSK1).



(a)



(b)

Figure 3.11 Whole genome data from the captive pink pigeon population (n = 6). For clarity the data has been split into nine macro pseudo-chromosomes (**A**) and 21 micro pseudo-chromosomes(**B**) alternating shades of blue have been used to distinguish adjacent chromosomes.

An example of how closely the results produced by RADiKal and whole genome data correlate can be seen by taking a single chromosome as an example (see Figure 3.12). The whole genome data from chromosome five shows two regions of low diversity and this is mirrored by the results from RADiKal.

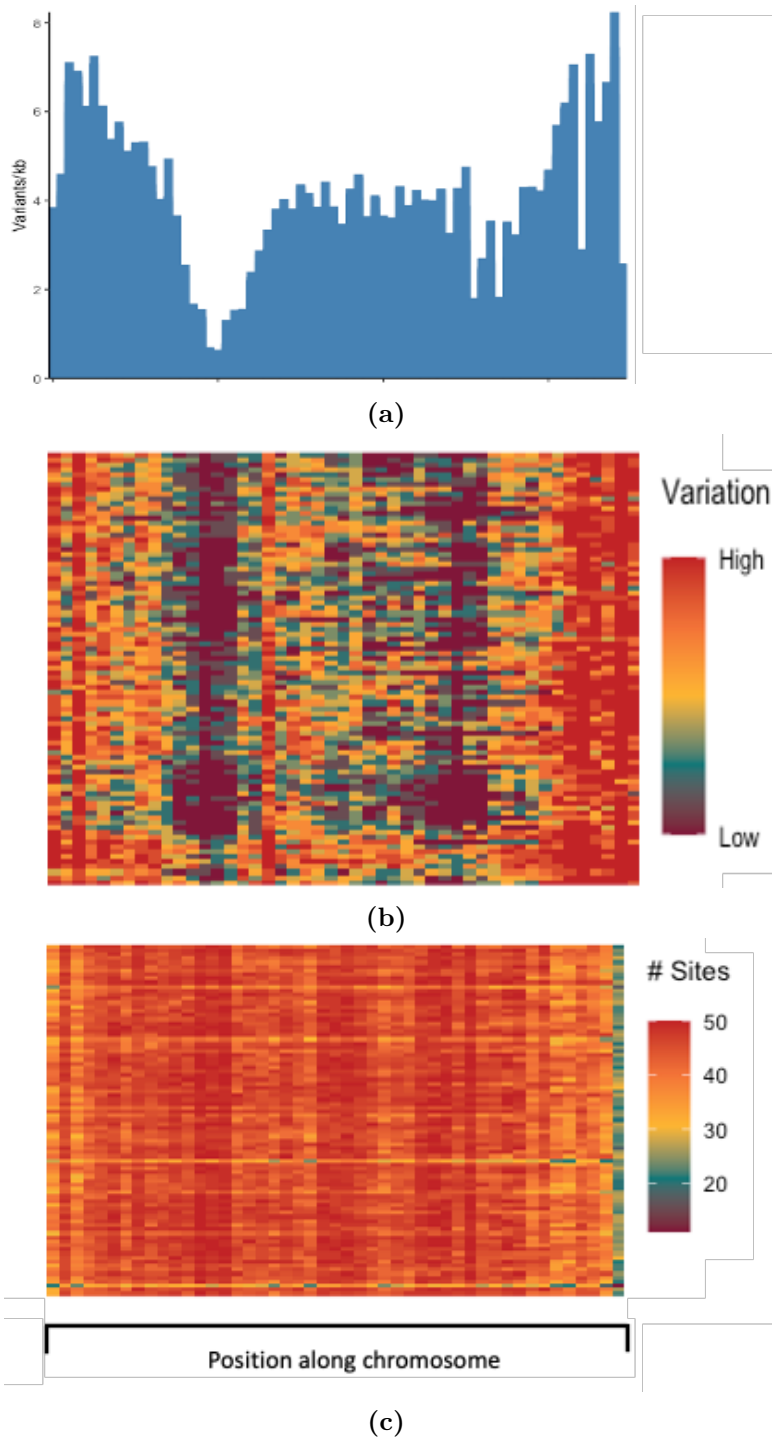


Figure 3.12 (a) Comparing whole genome data for chromosome five with **(b)** RADiKal output. Both graphs show two steep dips in variation shown by a decrease in variants/kb in (a) and the dark purple band in **(b)** which are regions with low variation (high scores). **(c)** shows number RAD sites per bin used to calculate scores to show that the patterns shown by RADiKal represent the biological data and are not just a function of marker density. For all graphs the x axis represents the position along the chromosome and for both **(b)** and **(c)** the y axis represents an individual (n = 87)

3.4 Discussion

RAD-seq is a popular method for generating population level data when studying endangered species but the data processing required to prepare the data for analysis can lead to biased and inaccurate data sets [258]. This problem is compounded by the use of prior assumptions to judge whether or not the data has been processed correctly. RADiKal is the first method that enables researchers, to extract information from raw unprocessed RAD-seq data. Researchers can then use this information to gain an overview of the variation present in a study population and to assess any prior assumptions made about the data.

RADiKal effectively shows the global spread of diversity across the genome and population, this can be clearly seen by comparing the results from RADiKal to those from whole genome resequencing data. The two data sets should never be identical (because they are from different individuals) however RADiKal captures the major patterns of diversity present in the captive population (n=6). In particular it is obvious where there are regions of low diversity which may point to regions of the genome that are highly conserved. Evidence for this was gained by recovering the coordinates of a region of low diversity in chromosome 5 (the first region of low diversity seen in Figure 3.12) and blasting the resulting sequence against the chicken genome which returned the protein coding gene BR serine/threonine kinase 1 (BRSK1). BRSK1 plays a key role in the polarisation of neurons and centrosome duplication and therefore is highly conserved [268]. The identification of this highly conserved genic region may highlight a larger issue in the way RAD data is used, which is that often the information from markers across the genome are condensed into a single value for heterozygosity. Condensing a value like heterozygosity to obtain a single value discards all the information about regions of high or low diversity across a genome that is revealed when whole genome data is displayed. However with RAD data there is the added complication that the marker position in the genome is unknown and the data is not evenly spread throughout the genome therefore it may be possible that many markers fall in a conserved genic regions and therefore may a) be unsuitable for the study of neutral evolution [31] and b) have a low heterozygosity. This would not necessarily reflect the fact that the population is genetically depauperate but rather it would reflect where the markers have fallen. It is assumed that having thousands of markers would decrease the impact of some data points violating the assumptions of the experiment. But this is an assumption that is rarely validated, with very few studies reporting the density or location of markers across the genome. For the pink pigeon SNP density was higher on the smaller micro-chromosomes, this same pattern was reported for another species of bird the Hawaii amakihi (*Hemignathus virens*).

3.4.1 Limitations and future work

RADiKal proves that raw data can be used to produce an overview of the genetic variation found in a population. The results are encouraging but because a reference genome is required there will be studies that are unable to use RADiKal. This should only apply to a small and decreasing number of studies because the number of genome assemblies available is increasing and it is possible to use the genome of a closely related species. The value of a reference genome and its importance for ensuring the accuracy of analyses has been discussed in Chapter 2 however the need for a reference genome does produce a reference bias. This bias may still result in the most polymorphic sites being discarded as non-RAD because they are so dissimilar to the reference genome. Despite this the heatmaps seen in Figure 3.7 demonstrate that RADiKal is able to detect highly polymorphic sites. In future it would be useful to try to quantify the amount of polymorphism in a site that means RADiKal mis-classifies it and compare this with levels of polymorphism found in nature to estimate the amount of information that may be lost.

The results from RADiKal are encouraging nevertheless, there are improvements and possible extensions to the RADiKal software that could be programmed in the future.

1. Firstly RADiKal has only been tested using pink pigeon data. In the future it will be important to test RADiKal on data sets from other organisms with different levels of diversity, that have used different enzymes. However there is nothing in the current design of RADiKal that is specific to the pink pigeon and because it builds a classifier for each data set to reflect the diversity within the raw data it should perform well regardless of the organism and thorough testing should confirm this.

2. Currently RADiKal only analyses RAD sites that are present in the reference genome and therefore will miss sites that may be mutated in the reference but not in the rest of the population. In future versions of RADiKal it is hoped that some of these sites can be recovered by refining the scoring matrix classifier and looking across the whole genome for sites that have k-mer coverage (this is discussed further in point 3.).

3. Currently RADiKal is designed to be used as part of a data processing workflow by providing information useful to experimental design however it does not classify RAD sites. In the future there is the possibility to use RADiKal as a software for classifying RAD sites that could be used in place of existing software such as Stacks. This could be done by using the scoring matrix/classifier to examine every position in the genome (using a sliding window), identify and classify sites and output the resulting data. Much of the infrastructure necessary for this is already present in RADiKal and such an

approach would yield several advantages over existing software. For instance the ability to draw any RAD site (similar to those heatmaps in Figure 3.7) and confirm if it is a RAD site. There would also be minimal parameterisation and therefore, theoretically, minimum information loss resulting in more accurate analyses.

4. The scores produced by RADiKal do not provide information about the difference in base composition or haplotypes. There will be occasions where sites may be heterozygous to the reference but represent a single haplotype in the population. If a population consisted of only a few haplotypes this should be obvious when looking at the painted chromosomes because each individual would be identical you would not get the mosaic pattern you see in the pink pigeon. This assumption could be confirmed by simulating data sites with a different number of haplotypes and examining the patterns produced in the painted chromosomes. It is also possible to retrieve the sequences for each individual at each site, then these sequences could be examined directly. However, there is currently no simple script available in RADiKal to retrieve sequences.

5. Running RADiKal is currently a manual process. Several inputs are hard coded, there is no documentation and the code has not yet been optimised for speed. There are plans to do all of these things in the near future so that the whole RADiKal process runs efficiently from input data to painted chromosomes with minimal user input.

6. RADiKal can produce images of RAD sites and is the first method to provide visual evidence for whether or not something is a RAD site. Currently this has to be done manually with only some images being outputted automatically (those RAD and non-RAD that score above/below the threshold). In the future there will be features to make it much easier for users to visually examine specific RAD sites.

7. RADiKal has been designed assuming a single RAD digest protocol and therefore would need to be adapted to use data from dd-RAD and other protocols. This is a programming challenge as the same philosophy applies in trying to extract RAD site signals from raw data.

3.4.2 Conclusion

RADiKal is the first tool that extracts a global signal from raw RAD-seq data, projects a global view of variation and provides a visual method for assessing whether or not a site is a RAD site. RADiKal is able to identify RAD sites, using a classifier, with very high accuracy (0.011% false positives which are removed from further analysis). RADiKal then extracts a signal from these sites that can be used to examine global patterns of variation at-a-glance, to test prior assumptions and aid in experimental design.

Chapter 4

Genetic rescue recommended: An updated PVA for the pink pigeon *Nesoenas mayeri*

4.1 Introduction

Human activities are causing a widespread decline in global biodiversity and many species now exist in small, fragmented populations [14, 269]. These species may then get caught in an extinction vortex (Figure 4.1) where genetic and non-genetic drivers act synergistically creating a positive feedback loop that places further pressure on endangered species and significantly increases their risk of extinction [18, 270, 19].

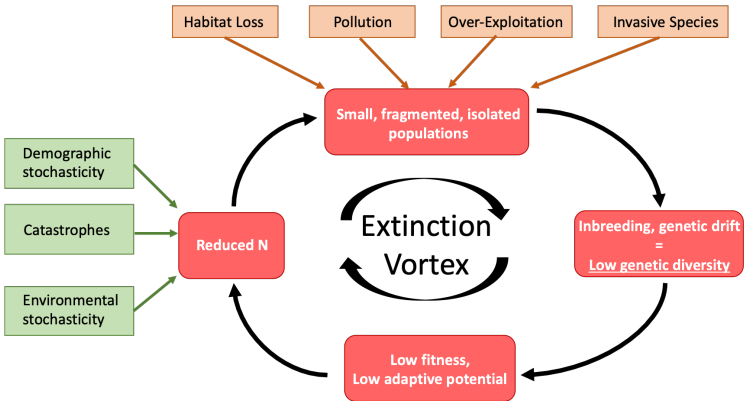


Figure 4.1 The extinction vortex describes a positive feedback loop whereby population size diminishes, ultimately resulting in extinction. Natural impacts are seen in green boxes, anthropogenic influences in yellow boxes and the results of these processes are seen in red.

To effectively disrupt the extinction vortex, management plans must be designed that remove or ameliorate threats and enable the population size to increase. However, given the many and interacting determinants of extinction risk, it is often not possible to decide on the most effective management strategy [271]. The difficulties in implementing effective conservation management plans are further hampered by limited resources [61] and time-scales that prevent scientists from being able to perform replicated experiments and long-term monitoring [272, 103]. Instead, models may be used to represent the study populations and to analyse their response to different demographic, environmental and management variables; these are known as a population viability analyses (PVAs) [101].

4.1.1 Population Viability Analysis

A Population Viability Analysis (PVA), is designed to model the probability of extinction of a population given a set of demographic, environmental, (sometimes) spatial parameters and management strategies [106]. As such, PVAs are an important part of conservation planning and are used as: (1) a conservation triage tool to assess which species are most in need of conservation [273], (2) to quantitatively compare different management strategies [101], or (3) to understand which demographic parameters have the largest impact on a species survival [274]. To construct a PVA either generic software packages are used or custom-built models are created. Custom-built models have been shown to produce higher quality models and analysis than those produced by generic software but they require an expertise in modelling that makes them prohibitive for many studies [275]. Although generic software packages have been created by researchers with an expertise in modelling and population theory [276, 273] they are designed to be user friendly and therefore accessible to researchers from a greater variety of backgrounds [275]. This accessibility is a dual-edge sword; although it may encourage more users their lack of knowledge about how the model is structured may lead to misuse and misinterpretation [275]. For example one of the most popular generic software used to create a PVA is Vortex [106]. Vortex has over 65 variables users can parameterise [273] allowing them to build flexible models tailored to the biology, ecology and threats of their study species. Some of these parameters require data that is difficult to obtain and often these parameters will have default values that can be used. However those values may not accurately reflect the species being studied and if users do not explore the parameter space (as recommended in various guidelines for example see [276]) the output from the models may not accurately reflect their study system [276, 273, 103].

Issues such as this are central to the continued discussion about the reliability, repeatability and effectiveness of PVA in conservation management [273, 275]. There is a concern that reliability and reproducibility, which are crucial to enable researchers to make justifiable and

beneficial decision in conservation, remain low [275]. This is attributed to poor communication, inadequate reporting and the lack of an enforced standardised protocol [273, 275]. Despite these concerns PVAs have been shown to be useful decision-making tools, which when parameterised correctly can accurately reflect and predict the demography of a species [277–279]. Even the process of constructing a PVA can contribute to current knowledge about a species by highlighting what data is lacking and therefore where further research is needed and by encouraging global collaborations [280, 281]. These actions are themselves likely to be beneficial to any conservation effort and more rigorous than the current alternative to PVAs which is human judgement [11].

4.1.2 PVA and pink pigeons

In 1991 a workshop was organised to produce a PVA for the pink pigeon (*Nesoenas mayeri*) [129]. This workshop brought together experts in the field, conservationists, students and used all data, current to that point in time, to produce a model. This original model concluded that in order to survive the wild pink pigeon population needed to be supplemented, and emphasised the importance of the captive population for these reintroductions [129]. Since this original PVA the pink pigeon population has gone through a number of changes including a crash and a recovery [3], the software used (Vortex) has improved, there are more genetic data and the current software accepts genetic data [106]. Therefore, a new PVA has been created using the latest version of Vortex software (v10.3.3.0) to provide an up-to-date baseline model for future studies to use, and to assess the impact of a genetic rescue on the pink pigeon population. In particular whether genetic rescue would significantly reduce the extinction risk of the wild pink pigeon population. The pink pigeon have a wealth of different data that have been collected over the course of more than four decades. These data include behaviour, ecology, DNA and pedigree (studbook) data. The availability of these data provides an opportunity to create a highly accurate Vortex model and to investigate the impact of modelling inbreeding depression using lethal equivalents. The number of lethal equivalents is a parameter that Vortex uses to measure the severity of inbreeding depression [106]; a parameter which is highly relevant for many endangered species. However, to accurately calculate the number of lethal equivalents requires data (inbreeding coefficients, life history trait) that studies are often missing [39]. These data are available for the pink pigeon because of the international studbook kept for captive pink pigeons.

The main aim of this chapter is to produce a PVA model that can be used as a resource to help conservation practitioners when revising management plans for the pink pigeon. In particular this model provides evidence that genetic rescue would be a good management plan that would contribute significantly to the long-term survival of the pink pigeon.

4.2 Methodology

All the Vortex project files, data and scripts used perform the analyses for this chapter can be found in appendix A.3.

4.2.1 Vortex

A PVA was built, using Vortex v10.3.3.0 [282], to assess the impact of genetic rescue on the long-term survival of the wild pink pigeon population. Vortex uses an individual-based, Monte-Carlo model that simulates the effects of stochastic as well as deterministic events on populations [102]. Population dynamics are modelled as discrete sequential events (the order of which can be user determined) that happen according to probabilities that are randomly generated from user specified distributions [107, 106]. Vortex is one of the most commonly used software packages [273, 106] and has been used to run PVAs on over 150 vertebrate species [107, 283]. Vortex was chosen for this project due to its flexibility, the inclusion of stochasticity, the ability to include genetic data and to maintain continuity from the previous PVA [129], which used an earlier version of Vortex (version number not reported). The version of Vortex that was used in 1991 lacked several of the current features specifically in the previous PVA the authors did not specify an inbreeding coefficient, number of lethal equivalents or starting allele frequencies for the population possibly because those functionalities did not exist. To determine if genetic rescue would be a successful and viable conservation strategy three different management scenarios were developed (see Table 4.1: Scenario 1 represents the wild population with no supplementation (baseline model); Scenario 2 simulates the impact of demographic rescue; and Scenario 3 simulates the impact of genetic rescue on the wild population of pink pigeons. The results for each scenario were averaged across 1000 iterations and each scenario was modelled over 100 years which represents 17.88 generations (assuming generation t = 5.6, calculated from international studbook data using PMx v1.4.2 [8]). A 100 year forecast is commonly chosen in studies that use Vortex, particularly when modelling species with shorter life spans (such as the pink pigeon) [284, 106]. The three different management scenarios were tested using the parameters described in appendix A.3. Due to the wealth of data on the pink pigeon, from over four decades of research, the key demographic parameters have good information and therefore do not require further comment. However, several parameters including those involved in the modelling of genetic diversity, inbreeding depression, and supplementation require further explanation and are described below.

Table 4.1 Descriptions of the three scenarios tested using Vortex v10.3.3.0 and the differences in data input between them where *Allele Freq.* refer to which populations were used to calculate the allele frequencies and *Supp.* refers to the supplementation regime and the numbers of females and males supplemented (F:M respectively). Further details describing the choice of parameters are given in the text following the this table, all other parameters are identical between the scenarios (see Appendix for complete list of parameters).

(#)Scenario	Supp.	Allele Freq.	Inbreeding Coefficient (F)
(1) No Supplementation	None	Wild and Captive	Wild =0.15, Captive = 0.10
(2) Demographic rescue	20 (10F:10M)	Wild and Hypothetical	All = 0.15
(3) Genetic rescue	20 (10F:10M)	Wild and Captive	Wild =0.15, Captive = 0.10

4.2.2 Choice of populations to include

Although there are currently eight free-living subpopulations (see Introduction) the genetic data used for this study was collected in previous years (see Table 4.2) when there were only five free-living subpopulations [116]. Therefore, each simulated scenario contained six populations, five of which represented the wild subpopulations (Pigeon Wood (PW), Brise Fer (BF), Bel Ombre (BO), Ile Aux Aigrette (IAA) and Combo (CO). The sixth population in each scenario, which was excluded from metapopulation calculations, represented either the captive population (CA; scenario 1 and 3) or a hypothetical population. This hypothetical population had the same demographic parameters as the captive population but allele frequencies and inbreeding coefficients taken from the wild metapopulation. Having this hypothetical population made it possible to delineate the impact of introducing novel alleles into the population from the effect of increasing the number of individuals through supplementation, i.e. the difference between demographic and genetic rescue.

4.2.3 Supplementation

The supplementation regime used was identical for both genetic and demographic rescue scenarios (scenarios 2 and 3). Ten birds (in equal sex ratio) were supplemented to each wild subpopulation every five years using a staggered supplementation regime whereby each year a different subpopulation is supplemented. When the model was being designed this supplementation regime proved effective, reflected a numbers of birds that appeared feasible based on past reintroductions [3] and recommendations from the previous PVA (10 pairs of birds released every other year) [129]. The supplementation regime chosen was deemed sufficient because the purpose of this PVA was not to design a release programme but rather to provide a baseline model that could be used in future research and to assess the impact of genetic rescue. However during preliminary testing other supplementation regimes were tested and did not appear to significantly alter the results of the model.

4.2.4 Genetic data

Vortex provides the flexibility to include genetic data to allow users to study the interplay between demographic, environmental, stochastic and genetic factors that contribute to extinction risk. The allele frequencies, used in the current model, were calculated in CERVUS 3.0.7 [285] using data from 22 microsatellite markers. These microsatellite markers were generated by a previous study from both captive and wild birds [13], the sampling years and number of individuals are in Table 4.2 and the allele frequency files can be found in appendix A.3. To create the starting allele frequencies for the hypothetical population, the allele frequencies for each of the wild subpopulations were pooled and instead treated as a single "wild" population.

Table 4.2 Information on the samples used to generate the allele frequency files used to calculate the starting genetic diversity of wild, captive and hypothetical populations. Data taken from [13] . **PW** - Pigeon Wood, **BF**- Brise Fer, **BO**- Bel Ombre, **IAA**- Ile Aux Aigrettes, **CO**- Combo, **CA** - Captive

Population	Years sampled	Number of birds
PW	2009, 2011	41
BF	2009-2010	65
BO	2009-2010	24
IAA	2009-2010	45
CO	2009-2010	26
CA	2012-2013	52

4.2.5 Inbreeding

In Vortex an inbreeding coefficient can be assigned to each modelled population but by default Vortex assumes no inbreeding so each individual in a population would start with an inbreeding coefficient of zero (F =0) [286]. For many endangered species that have been living in small populations it is more likely that inbreeding will be present [287, 288]. This is true for the pink pigeon, where inbreeding coefficients have ranged from F = 0.05 to F > 0.25 in wild birds (based on data collected from 1987 - 1998) [120] and from F = 0.0721 - 0.3487 in the the captive population based on international studbook data (see Figure 4.2). The international studbook was curated and downloaded by the studbook keeper, Harriet Whitford of the Durrell Wildlife Conservation Trust, on 16/01/2019 and the information contained within it is current to 21/12/2018.

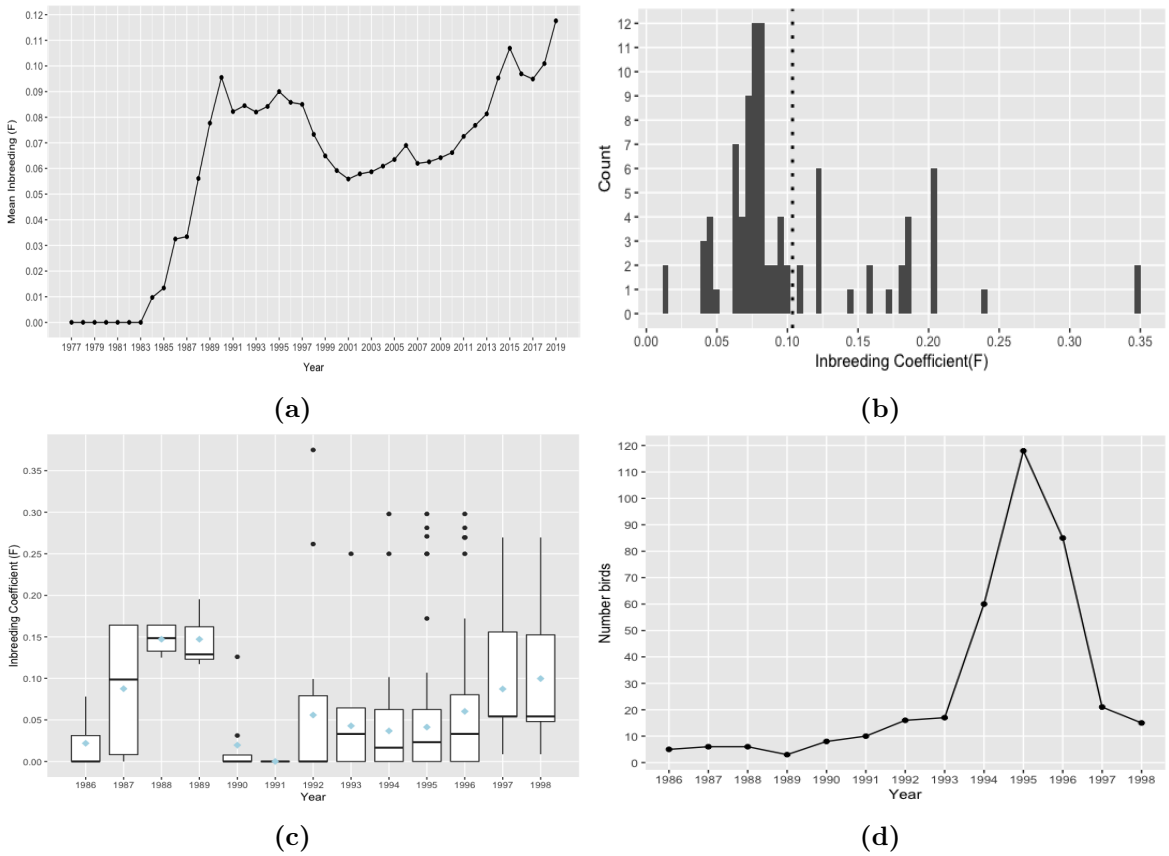


Figure 4.2 Inbreeding in the wild and captive pink pigeon populations. **(a)** Changes in the mean inbreeding levels in the captive population calculated with the genetics module in PMx v1.4.256 [8]. **(b)** Distribution of inbreeding coefficient in the current captive population with the mean inbreeding coefficient represented by a dotted line (F = 0.1036) (n = 91, based on international studbook data from living birds). **(c)** Changes in the levels of inbreeding based on data from fledged birds from the original *in situ* captive population (GDEWS) and wild population the mean inbreeding coefficient for each year is represented by the blue diamond. **(d)** Number of birds fledged (GDEWS) recorded each year. For figure panel **c** and **d**, the data was taken from [9]. Given the demography of the pink pigeon population at that time (see Introduction) the data from GDEWS and any wild population was combined and treated as a single population.

The inbreeding coefficients chosen for the model can be seen in Table 4.1. For the captive population, the mean inbreeding coefficient for the current living population was used (F = 0.10). However the current estimates of inbreeding in the five wild subpopulations are unknown, despite a wild studbook being kept, due to the difficulties of assigning parentage in wild birds [289, 201].

Instead a value of F = 0.15 was chosen to represent a moderately inbred wild population [120]. This value is supported by the following information: (1) the wild population, which is not currently being managed, is likely to have higher levels of inbreeding than the managed

captive population; (2) birds used to represent the captive population in this study (sampled in 2012) have a mean inbreeding coefficient of F = 0.09 (ranging from F = 0.074 – 0.186) birds currently alive in captivity today have an increased mean inbreeding coefficient (F = 0.1036) and contain individuals with an inbreeding coefficient as high as F = 0.3487 (range of F = 0.0721 - 0.3487). This is noteworthy because this increase in inbreeding coefficients occurred whilst the captive population was being managed to reduce the level of inbreeding; (3) the traditional pedigree calculations of inbreeding coefficients assume that all founders are unrelated, yet it is known that the pink pigeon founders were related [129, 120]. Hence, the calculated inbreeding coefficients are likely to be an underestimate; (4) Inbreeding coefficients were calculated for the wild population and *in situ* captive population using individuals in the population in 1987 - 1998 using data from [9]. In 1998 the mean inbreeding was approximately F = 0.10 (but included highly inbred individuals with F = 0.25) since then over two decades have passed with little management aimed at reducing inbreeding. It therefore seems plausible that the inbreeding coefficient of the wild population has increased.

Given these considerations, an inbreeding coefficient of F = 0.15 is a reasonable value but it would aid future research to calculate contemporary inbreeding coefficients for the wild pink pigeon population. It would be preferable to calculate these using molecular methods, which would account for unknown levels of historic inbreeding and reflect demographic processes [204] (also see Chapter 2).

4.2.6 Inbreeding depression and lethal equivalents

Inbreeding depression was modelled for all scenarios to measure the severity of inbreeding depression. Vortex v10.3.3.0 uses the number of lethal equivalents (LE) which represents a group of alleles which if made fully homozygous would be lethal [290]. The number of lethal equivalents was calculated in R v3.5.0 [291] using a logistic regression of (natural log transformed) longevity of birds (expressed in the number of days lived at death + 1) against inbreeding coefficient (F) see Equation 4.1 and Figure 4.3. The inbreeding coefficient was calculated based on international studbook data using PMx v1.4.256 [8] filtered using the Genetic Module to include only captive-born birds with known inbreeding coefficients and known age at death. For this analysis birds that died between 1976 to 21/12/2018 were used, which represented 1112 birds out of a total of 1308 birds present in the international studbook at that time.

$$b = \frac{-(Ln(S))}{F} + A \tag{4.1}$$

S longevity in days; **F** the inbreeding coefficient; **A** is a constant equal to the value of the log of the intercept representing the value of the trait (in this case longevity) in non-inbred individuals; **B** is the gradient of the line and is equal to the number of lethal equivalents in a haploid gamete, therefore 2B is equal to the number of lethal equivalents in a diploid organism [290].

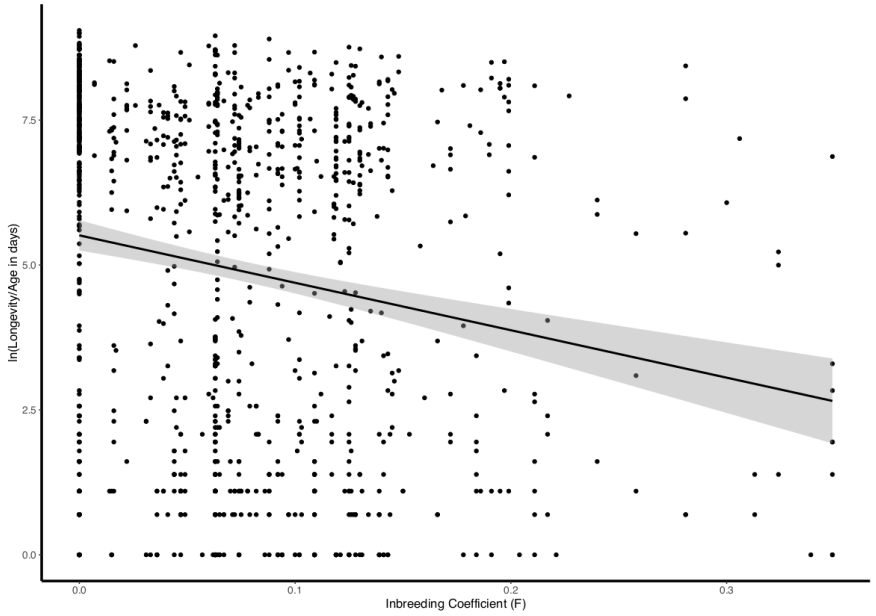


Figure 4.3 Log linear regression of longevity in days against inbreeding coefficient(F) ($F_{1,1111}$ = 33.550; $p < 0.0001$). The slope of the regression line represents the number of lethal equivalents in a haploid individual, resulting in a value of **15.13** lethal equivalents for the diploid pink pigeon.

Vortex v10.3.3.0 models the impact of inbreeding depression as a reduction in first year survival Equation 4.2. Although inbreeding depression can impact individuals, at different life-stages and different life history traits the authors of Vortex show that this can be effectively captured as a decline in first year survival [106]. The total first year mortality for wild, moderately inbred birds based on field observation is 75.3% [120] and this figure represents mortality due to many causes including inbreeding depression. Therefore this value must be adjusted so that the impact of inbreeding depression is not modelled twice and the modelled first year mortality does not end up greater than 75.3%.

$$S = S_{0-1}(e^{-b(1-Pr[Lethals])F}) \tag{4.2}$$

S is the probability of survival, $S_{0\text{-}1}$ is the probability of survival from age 0-1, b is the number of lethal equivalents per haploid genome; Pr[Lethals] is the proportion of inbreeding depression due to lethal alleles (0.5), F the inbreeding coefficient (0.15/0.10). By rearranging

this equation Equation 4.3 and using the values in Table 4.3 the level of first year mortality that needs to be modelled in order for mortality to equal 75.3% can be calculated.

Table 4.3 Values used to calculate a modified survival probability during year 0 - 1

Parameter	Value	Ref
S	0.247	[120]
Pr[Lethals]	0.5	[106]
b	7.565	see above
F	0.10/0.15	wild/captive see above

$$S_{0-1} = \frac{S}{(e^{-b(1-Pr[Lethals])F})} \tag{4.3}$$

This give a first year survival of 0.456 which translated to a first year mortality of 54.4% which will increase to an overall mortality of 75.3% as Vortex increases first year mortality due to inbreeding depression.

4.2.7 Literature search

The pink pigeon has a lot of data that is lacking in other endangered species, for example an almost complete international studbook, which was why it was possible to calculate the number of lethal equivalents in the captive population directly as opposed to using Vortex's default value of 6.29. The number of lethal equivalents calculated was more than twice the default value used by Vortex v10.3.3.0 and research has shown that the number of LE modelled can significantly impact the results of the model [292–294, 105]. Despite this many studies use the default value (due to lack of better data) and do not perform sensitivity testing to examine the impact of LE on the model. A literature search was conducted to see how many studies used the default value and sensitivity testing to explore the parameter space (as recommended in some guidelines [276]). To conduct the literature search all the publications on Vortex's website (http://www.vortex10.org/VortexReferences.aspx [accessed June 2018]) were assessed and those studies which included inbreeding depression (heterosis model) in their simulations were assessed to see what values were used for the number of lethal equivalents and whether sensitivity testing was used.

4.2.8 Sensitivity Testing

Sensitivity testing was used to explore the parameter space of those parameters that were either uncertain and/or of interest to management plans. Vortex has a sensitivity testing module which allows users to specify a baseline model, the parameters they want to alter and the parameter space they want to test. Vortex will use this information to automatically generate all the required sensitivity testing scenarios. In this study the baseline model was Scenario 1 where there was no supplementation therefore any parameters that were not altered during testing were the same as in Scenario 1 (see appendix A.3). Scenario 1 was chosen as the baseline model because it provides the opportunity to explore what parameters (other than supplementation) may have an impact on the survival of the pink pigeon and may inspire alternative management options. Table 4.4 shows the parameters,the range of values used in sensitivity testing and the reason they were included.

A single factor sensitivity analysis was used which changes a single parameter at a time altering the value according to a user specified interval (or list of values). Single factor analyses are useful for identifying variables that may have a significant impact on the model individually [106, 283]. Sensitivity analyses were run for 1000 iterations [106].

Table 4.4 Range of values used in sensitivity analysis. The "Base" scenario copies its values from Scenario 1 (no supplementation).

Parameter	**Base**	**Range**	**Interval**	**Reason for inclusion**
Adult Mortality*	14.78	10 - 30	5	The data used is over 10 years old [116] therefore it is possible that they are not representative of the wild population as it is today. Because the primary causes of death varies between juveniles and adults [125, 3] knowledge about the impact of mortality rates independent of life-stage may indicate where conservation should be focused [283]
Juvenile Mortality	54.4	30- 60	10	These data were generated from a single subpopulation (IAA) several years ago because this is the only subpopulation that these data can be feasibly collected [3]. As such there is a chance that they may not accurately represent the other subpopulations or current mortality rates in IAA.
% of birds that have 1 offspring **	88	50 - 100	10	Inbreeding depression has been linked to reduced fertility in pink pigeons [120], if this parameter proves sensitive it would add support for the need for genetic rescue.
Carrying capacity	200	50 -250	50	It is difficult to estimate carrying capacity that may also have been elevated artificially by supplementary feeding [119]. If it is a significant factor in the pink pigeon's survival further research would be needed to ensure reserves are large enough to maintain viable populations

Range of values used in sensitivity analysis (continued)

Parameter	Base	Range	Interval	Reason for inclusion
Number of LE	15.13	3 - 21	3	LE are often not calculated and instead the default for Vortex is used [89] however several studies have found their models sensitive to the number of LE modelled [292, 105]. The number of LE were calculated using longevity data but the number of LE can be estimated independently for a range of life history traits and then the results summed [39]. By focusing solely on longevity the number of LE in pink pigeons may have been underestimated.

* Adult mortality varies for each subpopulation however during sensitivity testing the mean mortality for all five subpopulations (excluding Captive) was used as a baseline scenario as the aim was to examine the impact of altering adult mortality on the metapopulation but not individual subpopulations.

** pink pigeons can have up to 2 offspring per brood, this parameter alters the percentage of females having a single offspring as opposed to two offspring, therefore 50 represents half the females having two offspring, and the other half of the female pink pigeons having a single offspring.

4.2.9 Statistical analysis

To evaluate the effects of the different management plans (demographic and genetic rescue) and model settings, the mean population size and heterozygosity were calculated at year 50 and 100 to test for any short- or long- term effects. These key outcome variables were compared statistically against the baseline scenario (Scenario 1) using strictly standardised mean differences (SSMD) in the R package VortexR v1.1.5 [283] used in R v3.5.0 [291]. The SSMD is the ratio of mean to standard deviation of the difference between two groups. Unlike other statistical tests SSMD is robust to increasing sample size, with an increase in sample size resulting in the SSMD approaching its true value as opposed to artificially decreasing the p value [295]. Therefore SSMD is recommended when dealing with large sample sizes (such as a large number of iterations) where the object is to test for the difference between two populations [103].

4.3 Results

To test if genetic rescue would be a viable and successful management plan for the pink pigeon three Vortex models were created and the parameter space explored. Because the primary research goal was to assess the impact of management plans on the population as a whole and given the size of Mauritius and the inter-connectedness of the subpopulations all the results described here are for the wild metapopulation (excluding captive or hypothetical populations) not for any single wild subpopulation.

4.3.1 Impact of genetic rescue

The results from the three main scenarios showed that with no supplementation (Scenario 1 – baseline), the pink pigeon is almost certain to go extinct within 100 years Table 4.5. Supplementing the population appeared to increase the chance of pink pigeon survival with the greatest difference observed when genetic rescue was employed and captive individuals with novel alleles (Scenario 3) were used to supplement the population Figure 4.4.

Table 4.5 The probability of extinction and meant time to extinction (measured in years) of the three main scenarios modelled in Vortex.

(#) Scenario	Pr. extinction	Mean time to extinction
(1) Baseline	0.997	54.4
(2) Demographic rescue	0.552	82.9
(3) Genetic rescue	0.005	95

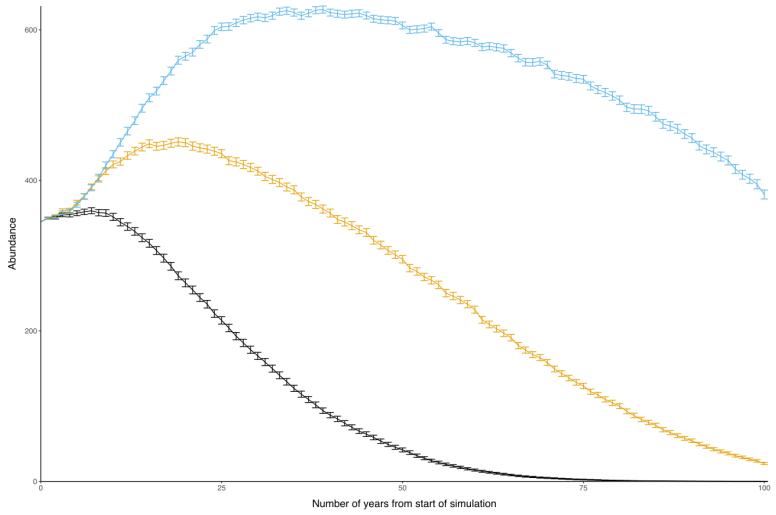


Figure 4.4 Probability of long-term survival of free-living pink pigeons. The mean abundance (±SE) of pink pigeons has been calculated over 1000 iterations for each year of the model. Scenario 1 (no supplementation; black line); Scenario 2 (demographic rescue, i.e. supplementation with a hypothetical genepool similar to the free-living metapopulation; yellow line), and Scenario 3 (genetic rescue, i.e. supplementation with zoo-bred captive birds; light blue line). The supplementation regime for Scenario 2 and 3 was identical; ten birds for each subpopulation every five years).

It is important to be able to quantitatively show that a management strategy produces a statistically significant result compared to the baseline scenario [296]. The results of the pairwise analysis showed that despite demographic rescue (Scenario 2) qualitatively improving the long-term survival of the pink pigeon population, genetic rescue (Scenario 3) was the only model that had a statistically significant impact on the mean number of individuals in the wild metapopulation in both the short-term (50 years) and long-term (100 years) (see Table 4.6a). Genetic rescue was also the only model to significantly increase the level of heterozygosity within the wild population in the long-term compared to the baseline scenario (see Table 4.6b).

Table 4.6 Results from pairwise analysis showing the impact of demographic and genetic rescue on (**a**) the abundance (Nall) of pink pigeons and (**b**) the levels of heterozygosity in the pink pigeon population compared to the baseline scenario (no supplementation).

(a)

Scenario	**50 Years**		**100 Years**	
	SMMD	*P*	SMMD	*P*
2	1.4245	0.0772	0.4970	0.3096
3	3.3373	**0.0004**	2.0185	**0.0218**

(b)

Scenario	**50 Years**		**100 Years**	
	SMMD	*P*	SMMD	*P*
2	0.9559	0.1696	0.9966	0.1595
3	1.1737	0.1203	10.8500	**0.0000**

4.3.2 Sensitivity Testing

The sensitivity testing scenarios that had a significant increase on the mean abundance of birds or their heterozygosity were: decreasing the number of lethal equivalents (3,6,9), decreasing juvenile mortality (30%, 40%) and decreasing the % of birds laying one egg (50%, 60%). The sensitivity testing models which were significantly different from the baseline (Scenario 1) are presented in Table 4.7. There was only one scenario which caused a significant decrease in levels of heterozygosity which was when the number of lethal equivalents were increased to 21.

Table 4.7 Results from single factor sensitivity analysis. Only scenarios which produced a significant difference in the abundance (Nall) **(a)** or heterozygosity **(b)** (in at least one time point) of the pink pigeon compared to the baseline scenario (no supplementation) are shown below. **JM** – juvenile mortality; **LE** – number of lethal equivalents; **%OF** – percentage of birds having a single offspring as opposed to two.

(a)

Parameter(value)	**50 Years**		**100 Years**	
	SMMD	*P*	SMMD	*P*
JM(30%)	4.3734	0.0000	7.9449	0.0000
JM(40%)	3.4252	0.0003	3.4692	0.0003
%OF(50%)	3.0082	0.0013	2.5769	0.0050
%OF(60%)	2.004	0.0225	1.2502	0.1056
LE(3)	5.1780	0.0000	9.1258	0.0000
LE(6)	4.0264	0.0000	6.5468	0.0000
LE(9)	2.6576	0.0039	2.8943	0.0019

(b)

Parameter(value)	**50 Years**		**100 Years**	
	SMMD	*P*	SMMD	*P*
JM(30%)	0.5579	0.2884	7.7836	0.0000
JM(40%)	0.5572	0.2887	7.6731	0.0000
%OF(50%)	0.5558	0.2892	5.9777	0.0000
%OF(60%)	0.5342	0.2966	2.6506	0.0040
LE(3)	0.5379	0.2953	7.6539	0.0000
LE(6)	0.5415	0.2941	7.7026	0.0000
LE(9)	0.5249	0.2998	5.2128	0.0000
LE(21)	-2.3962	0.0083	-0.1704	0.4324

All the sensitivity testing scenarios which caused a significant difference in abundance (Nall) compared to the baseline scenario resulted in the pink pigeon surviving for one hundred years (see Figure 4.5).

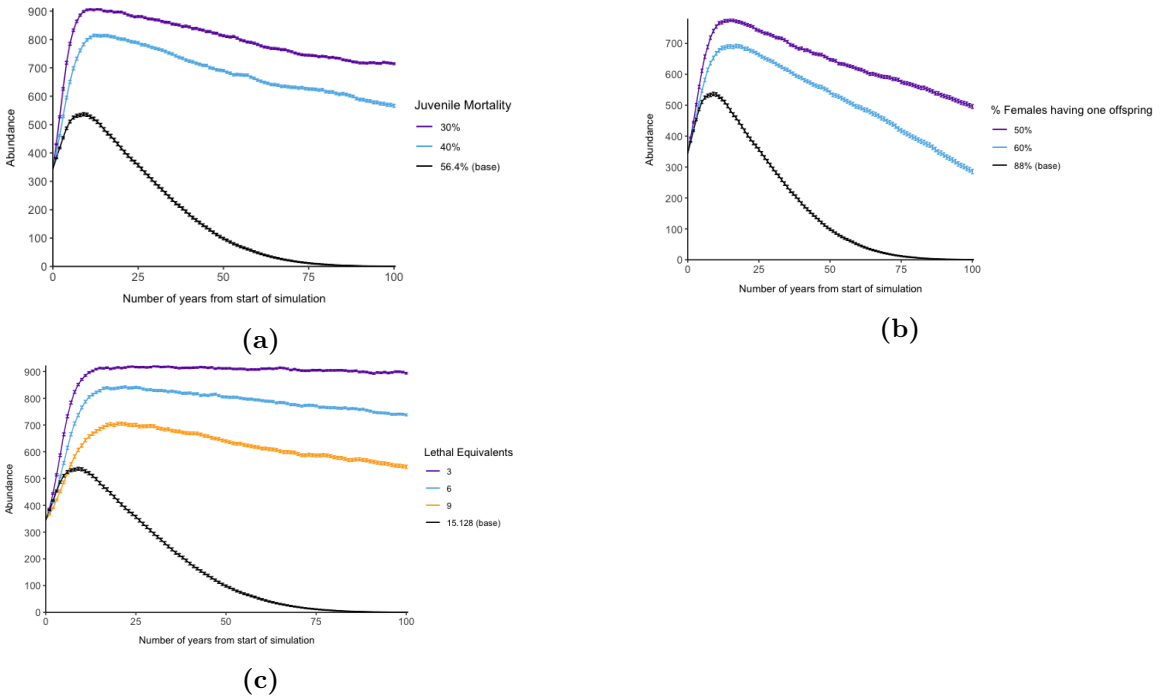


Figure 4.5 Sensitivity scenarios that were significanlty different to the baseline scenarios. Single factor anylysys. **(a)** decreasing juvenile mortlaity **(b)** increasing the number of females producing twp offspring, **(c)** decreasing the number of lethal equiavlants. Legends in each plot show the values that were modelled.

The number of lethal equivalents calculated for the pink pigeon (15.13) was larger than Vortex's default value of 6.29 and sensitivity testing demonstrated that the number of lethal equivalents chosen could have a significant effect on both the abundance and heterozygosity of the pink pigeon. The impact of altering lethal equivalents was explored further by re-running the baseline scenario with the default value of 6.29 LE (which was not initially tested during sensitivity testing) and showed that when LE = 6.29 the pink pigeons would not go extinct during the lifetime of the model (see Table 4.8). Using the default value for LE had a significant effect on the mean abundance of individuals in both the short and long-term both (SSMD = 6.1423 , p =< 0.0001 ; SSMD = 8.5333, p < 0.0001 respectively) and heterozygosity in the long-term (SSMD = 19.9983 , p < 0.0001).

Table 4.8 The impact of using Vortex's default value for lethal equivalents, on the pink pigeon model. All results are taken from the final year of the model (year 100) and averaged over 1000 iterations.

Scenario	**#Lethal equivalents**	**Pr extinction**	**Mean time to extinction**
Baseline	15.13	0.9990	54.4
Default LE	6.29	0.000	NA

4.3.3 Literature search results

Of the fifty publications listed on Vortex's website (at the time of searching in June 2018) 21 studies had studied inbreeding depression, a table describing the studies can be found in the appendix A.3. Of those 21 manuscripts 13 (61.90%) used the default value for the number of lethal equivalents and 6 (28.57%) of those did no sensitivity testing, the possible implications of these findings are discussed below.

4.4 Discussion

A PVA built using Vortex v10.3.3.0 demonstrated that without genetic rescue the pink pigeon would go extinct, thus adding to the growing body of evidence that highlights the importance of genetic rescue is conservation [101, 111, 156]. This PVA also highlights the potential for other management strategies and the hazards of using default values without exploring the uncertainties in parameter choice.

The results of the current pink pigeon PVA indicate that genetic rescue will be an effective management strategy resulting in the long term (100 years) survival of the wild population. This adds to the growing body of evidence that has shown both the positive impact of genetic rescues [81, 89] as well as the utility of a PVA as a tool to design and test management strategies [89, 297]. PVAs are an invaluable tool for testing management strategies in particular for assessing the impact of more unconventional management options. For example a PVA for the critically endangered helmeted honeyeater (*Lichenostomus melanops cassidix*) was used to assess the impact of a genetic rescue through hybridisation with different subspecies, the common yellow-tufted honeyeater (*Lichenostomus melanops gippslandicus*). Although this is not a conventional management strategy the results of the PVA showed that restoring gene flow between the sub species to historic levels was a viable conservation strategy and had positive demographic and genetic outcomes for the helmeted

honeyeater [89]. It could be argued that models may be misleading and not accurately reflect the genetic impacts of supplementation however, a recent study comparing the results of a PVA that modelled population supplementation with empirical data showed that the results predicted by the PVA accurately reflected the changes in genetic diversity (measured before and after supplementation using genetic markers) [298]. Hopefully more studies will be done to provide evidence that a well-implemented model can accurately reflect real world scenarios and this will then encourage more studies to use PVA and inspire greater confidence in their results.

A model is only as good as its parameterisation [284, 299] and in this study every effort was made to ensure that the parameterisation was accurate. The most accurate data available were used and uncertain parameters were subject to sensitivity analysis as recommended by PVA guidelines [276, 103]. Nevertheless, due to the constraints of working with wild animals and the availability of data there are several factors that may impact how accurately the model reflects the current wild population. For instance much of the data have been opportunistically sampled from populations during previous research and because of this only five out of the current eight subpopulations were modelled. Another factor to consider is that some of the data came from studbooks from the managed captive population, which may differ from the wild population [300]. However, the pink pigeon is data rich compared to most endangered species and this meant that an accurately parameterised model could be developed. This model can now be used to inform management decisions for the pink pigeon, and to gain insight into parameters (e.g. the number of lethal equivalents) where data are often lacking in other species [39].

Sensitivity testing was used to test the robustness of the PVA and to assess the impact of parameters on the model. Given the caveats in data collection mentioned above it was especially important to ensure the robustness of the model. If any of the tested parameters proved to be sensitive, for example, if a small deviation from the modelled value would cause a significant effect to the model output, the model may be less robust and more likely to output inaccurate results. Instead, although the majority of tested parameters (3/5) created scenarios that had a significant effect on the model output (here the heterozygosity or abundance of birds) the parameterised values were all more than 25% different to the baseline parameters. Given the quality of the data such a large deviation from the measured parameters seems unlikely and provides evidence for the robustness of the current model.

The three parameters that had scenarios that were significantly different to the baseline model were juvenile mortality, the proportion of birds producing two young and the number of lethal equivalents. The results showed that decreasing juvenile mortality, increasing the proportion of females laying two eggs and decreasing the number of lethal equivalents all

had a significantly positive impact on the model, resulting in more birds and/or greater heterozygosity. Interestingly the impact of increasing the number of females producing two offspring to from 12% to 40% (60% producing a single offspring) only had a significant effect on abundance in the short term (50 years). This may be because this increase in female productivity was not enough to offset the increased juvenile mortality due to the increase in inbreeding that happens over time. The long term impact of increased productivity was only seen when 50% of the females produced two offspring which is markedly different to the current value from field records of 12% of females producing two offspring . Overall the results of the sensitivity testing can be described, in terms of management plans, as increasing productivity and decreasing the genetic load. This is an encouraging finding because pink pigeons can be bred in captivity very efficiently and good captive breeding and release protocols have already been established from previous conservation work[129, 3]. Trying to improve productivity in wild birds is much more difficult especially as there are already several management plans designed to do this, such as predator and disease control [301], however these finding emphasise the importance of continued management.

The sensitivity testing also added support to several studies which have reported that the number of lethal equivalents modelled, or the modelling of inbreeding depression can have a significant impact on the outcome of a PVA [292–294, 105]. A decrease in the number of lethal equivalents (LE = 3,6,9) led to a significant increase in both the abundance (short- and long- term) and heterozygosity (long-term) of the pink pigeon. Increasing the number of lethal equivalents, to LE = 21 only caused a significant decrease in heterozygosity (not abundance) in the short-term (over 50 years), compared to the baseline model. This finding, that an increase to the number of lethal equivalents has relatively little impact on the baseline model, is perhaps reassuring because it is possible that the total number of lethal equivalents in the pink pigeon is greater than the 15.13 calculated. The total number of lethal equivalents in an individual reflects the impact of deleterious variants across different life history stages [290, 38, 287]. In this study the number of lethal equivalents has only been calculated using longevity and not, for example, fecundity due to the availability of appropriate data. Another reason to think that the calculated value for LE may be lower than in the wild population is that the data used was from the international studbook. Given that the captive population has been bred to reduce inbreeding and maintain genetic variation [120] it is possible that the number of LE is greater in the wild population [300]. Although the model may be robust to some increases in the number of lethal equivalents it is important to note that using the baseline (LE = 15.13) the wild pink pigeon metapopulation still went extinct within 100 years.

The pink pigeon is one of the few endangered species with sufficient data to attempt to calculate the number of LE present and so more accurately model the impacts of inbreeding

depression. If inbreeding depression is not modelled studies risk underestimating a populations extinction risk [39]. The inappropriate use of the default value for the number of lethal equivalents (LE = 6.29) could also jeopardise the accuracy of the model (see Results) [302–304]. The default value used by Vortex is based on a study that gathered data from 40 wild taxa and gave an estimates of 12.3 diploid lethal equivalents, from this study Vortex uses a value of 6.29 which is the combined effect of inbreeding on fecundity and first year survival [39, 106]. This study was not exhaustive, used a relatively small sample size which covered a wide range of taxa therefore there is no guarantee that this value will reflect any particular species. Whilst it is understandable that studies want to include inbreeding depression in their models, even if they do not have sufficient data to calculate the number of lethal equivalents for their species, it is worrying that a considerable proportion of studies (28.57%) were found to have done this without sensitivity testing to explore the parameter space. Even though the sample size of studies that were used during the literature search was small (n = 22) it is a worrying finding. If PVA models are used that have been parameterised incorrectly and/or if the uncertainties in the results have not been made clear to practitioners then the use of these models could lead to ill informed management decisions, which could have disastrous consequences for the species involved [39, 27].

4.4.1 Management and the pink pigeon

In 2018 the pink pigeon was downlisted to Vulnerable (see https://www.iucnredlist.org/species/22690392/131665077) despite no improvement in the population size since the last IUCN assessment and no consideration for the genetic health of the population (see Introduction). The results seen by the current PVA suggest that the pink pigeon should not have been downlisted because it still has a high probability of extinction without continued intensive management, which includes genetic rescue. The pink pigeon is an ideal candidate for a genetic rescue because it fulfils the recommended requirements for a genetic rescue listed by Frankham in [91]: (1) it has a source population of the same species, (2) with the same karyotype, (3) that has not been separated for more than 500 years. Not only would a translocation from the *ex situ* captive population conform to recommended guidelines but the translocated birds would join a long running programme led by experts in pink pigeon conservation with experience in successful reintroductions [120, 3].

Whilst the pink pigeon is an ideal candidate for a genetic rescue, and the current model supports this as a viable and necessary management strategy, there are some caveats. For instance, adaptation to captivity can impair fitness in the wild and because this was not modelled it is possible that the improvement seen in the wild population after the introduction of captive-bred animals was overestimated [96, 94]. Although it may be useful to model the

lowered fitness of introduced birds in the future, the negative impact due to adaptation to captivity; (1) is likely far less than the negative impact of not doing a genetic rescue [111, 156] and (2) may not be as much for the pink pigeon compared to other species. This is because decades of experience have led to the creation of soft release protocols which minimise the impact of the release and provides time for acclimatisation [129]. In previous translocations, following the soft release, birds were monitored and managed (e.g. supplementary food, treated for disease) leading to the successful establishment and supplementation of several subpopulations [129, 3]. Having established breeding and release protocols would also aid with the proposed management strategy of continued releases over time which has shown to be effective (see Results and [101, 156]). This continuous re-introduction strategy was based on the number of birds previously released during *in situ* reintroduction attempts and recommendations from the previous PVA (although the previous PVA only considered a single subpopulation, PW) [129, 3]. However the precise strategy of supplementing a different population every year, in a cyclical manner, has not previously been attempted for the pink pigeon therefore whether or not this is practical and possible needs to be discussed with practitioners and a range of supplementation regimes explored. There is also the possibility that continuous reintroductions may lead to diminishing returns when considering the decrease in genetic load [305] which is one reason why genetic rescue must be monitored carefully and every effort should be made to alleviate the threats that led to the populations decline [35] emphasising the importance of continued management of the pink pigeon that includes predator control, disease control, habitat restoration.

One of the main challenges for researchers developing a PVA is their ability to communicate the uncertainties inherent in the model [304, 296]. For this model to be useful it needs to be presented and any potential uncertainties explained to practitioners and potential management options discussed. Any feedback from practitioners should then be added and the model re-run with the understanding that the PVA is a resource that can be added to and developed to form part of an adaptive management framework [81]. For example this model only considered five out of the current eight subpopulations because of a lack of data for these newer subpopulations but collaboration with practitioners could enable the addition of these three extra subpopulations to the model. Collaboration with practitioners will ensure that the PVA developed here can be more confidently used to guide future research efforts [299] and examine the impact of management scenarios on the pink pigeon.

4.5 Conclusion

"All models are wrong but some are useful" - **Dr. George Box**

The recommendations from this current PVA are clear, pink pigeons need genetic rescue and continued management. However, biological systems are complex and even with the best available data there will always exist a level of uncertainty within a model due to uncertainty in the data, the stochasticity in biotic (and abiotic) systems, and the numerous unknowns (e.g. disease outbreaks, historic levels of inbreeding, etc.) [306, 307, 284]. So whilst all models are wrong the question remains how useful is this model? I believe that the model produced here represents a useful resource to help plan current and future management plans for the pink pigeon. The results and insights gained by this model are clear and add to a growing body of evidence about the need for genetic rescue and the negative impact of inbreeding depression [111, 81, 156]. However the models true utility will only be realised after consultation with practitioners.

Chapter 5

I Choose You: A simple tool to help preserve genetic diversity

5.1 Introduction

Captive breeding programmes are playing an increasingly vital role in conservation by acting as source populations for reintroductions, including genetic rescues [289, 308, 66]. Organising a translocation for genetic rescue is a lengthy and complicated process and practitioners have to consider a variety of factors ranging from the bureaucratic to the biological [289, 19]. Perhaps the most important biological factor to consider during a genetic rescue is which individuals should be chosen for translocation [111]. Assuming that pragmatic factors such as age, reproductive potential, adaptation to captivity etc. are equal, the chosen individuals should have low mean kinship and high genetic diversity [309]. The pragmatic considerations are intuitive to practitioners who routinely carry out health checks and use studbook data to maintain a healthy captive population [201]. However, more challenging for practitioners to assess may be the genetic value of an individual for reintroduction [309] but maintaining genetic variation is an important consideration to enable an effective genetic rescue and alleviate the genetic strains caused by a small population size [305, 152]. Nevertheless, the current methods used to decide which individuals should be reintroduced are time consuming, expensive, require specialist knowledge or result in uninformed choices being made [201]. To address these challenges a new tool I Choose You (ICY) was created. ICY is a quick and easy-to-use tool that practitioners can use to help them select individuals that would best fulfil the goals of a genetic rescue - to increase genetic diversity and reduce inbreeding [92, 111].

The goal of genetic rescue is to introduce individuals with novel genetic variation into a small, genetically depauperate recipient population [84, 95]. To select appropriate captive individuals, to release into the wild, managers can use pedigree/studbook data, empirical genetic data [202] or sometimes no data at all [201]. There are a number of reasons why individuals for release may be chosen seemingly at random, for instance managers may not have access to the relevant genetic data, or it may not matter which individuals are chosen - if for example, there was an urgent need to increase census size [120]. Whilst such factors are understandable, if projects determine to undertake a genetic rescue then the most effective rescues will be with those where the choice of individuals for release is informed. The most informed choices will be made using empirical genetic data (markers, genome sequences etc.), from both the captive and wild population [202]. This provides the most accurate information on which lineages have the highest genetic diversity as well as the true relationship between individuals and populations [310, 311, 201, 312]. But genetic data may not be available for many reintroduction attempts because it can be costly and time consuming to produce [48, 26] and many practitioners may not be familiar with analysing molecular markers. Instead most practitioners will be familiar working with pedigree data, which represents a history of a population, and can be used to assess an individual's potential genetic contribution and relationship to other individuals in the population [313].

If pedigrees exist for both the wild and captive populations, and they are of sufficient quality and completeness, a comparison of the pedigrees can show which individuals are most likely to bring novel genetic variation during a reintroduction attempt [314]. However, because of the practical difficulties involved with keeping wild pedigrees these are only available for a few species [315, 316] and even those pedigrees kept for captive species can be largely incomplete [317, 318]. Nevertheless, almost every captive breeding programme will keep a pedigree in the form of a studbook, which also contains sufficient information to calculate useful metrics such as: inbreeding coefficients, mean kinship, founder information, relatedness and (theoretical) genetic diversity. In the absence of empirical genetic data, or complete wild and captive pedigrees, theoretical measures such as founder equivalents and mean kinship, calculated from studbook data, can be used to inform genetic rescue [319].

Selecting individuals with the lowest mean kinship (MK) for captive breeding programmes has been shown theoretically and empirically to retain genetic diversity [320, 64] and is recommended by current reintroduction guidelines [321, 111]. Mean kinship (described in Equation 5.2) is defined as an individual's average kinship to all other individuals in the population, excluding founding individuals but including itself. By minimising mean kinship within a population large amounts of genetic diversity can be retained, it also ensures that the founder alleles with the lowest frequencies are preferentially propagated [322]. Mean kinship has been used successfully to manage captive populations [323, 289, 324] however, when

planning reintroductions there are other metrics that should be considered such as founder genome equivalents (FGE) and founder equivalents (FE) which could be considered proxies for genetic variation [10, 322]. Founder equivalents measures the founder representation in a managed population and account for the loss of genetic variability due to unequal founder contributions [10] (see 5.1) it is defined by Lacy (1989) as "*the number of founders that would produce a population with the same diversity of founder alleles ... if all founders had contributed equally to each descendant generation (while the number of descendants remained the same)*". To maximise genetic diversity it is best that all founders are represented evenly because a skew in representation has been shown to decrease genetic diversity as rare alleles are more likely to be lost [10, 310, 66] (see Figure 5.1). Founder equivalents do not consider the random loss of alleles due to drift and therefore a more accurate reflection of the genetic variability of a population may be obtained by using computer simulations to calculate founder genome equivalents (FGE), which include a measure of allelic retention [10, 319, 322]. However FGE can not be simply calculated from studbook data but requires computer simulations [10, 316].

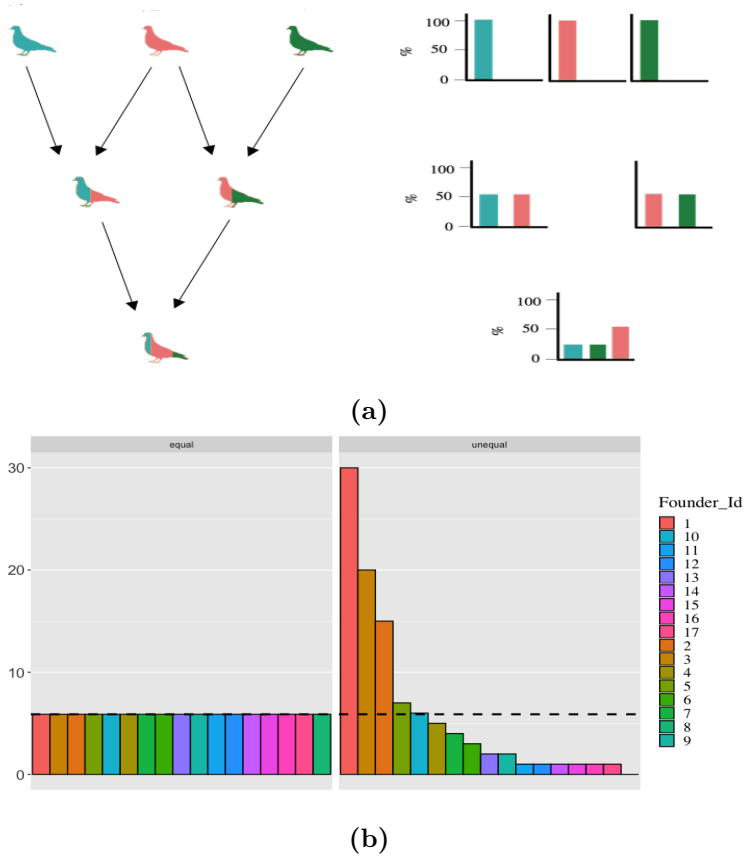


Figure 5.1 (a) How founder representation is calculated, from top to bottom the figure shows a population starting from three founders and how founder representation in their offspring vary from one generation to the next. The bar charts show the percentage of a birds genome that belongs to each founder. The assumption used by all studbook and pedigrees is that all founding individuals are unrelated and genetically distinct therefore each founder bird's genetic makeup is 100% unique [10]. In the first generation of offspring produced by founder pairings the resulting offspring's genome will be made up of 50% of each parental genome. If both the F1s, shown here in the middle of the image, reproduced 25% of the F2's genetic makeup is from the blue founder, 25% is from the green founder and 50% from the red founder because the red founder is represented in both F1 offspring. The individuals which have the greatest number of founders represented in their genome are likely to be the most genetically diverse. However if individuals have equal representations from all founders within their genome **(b)**, the individual with the more even contributions (left) are considered to be more diverse.

When deciding on the best method to use to choose a group of animals for reintroduction both FE and FGE have been shown to result in genetically more diverse populations than choosing groups by random, by fecundity or using allelic diversity [325]. However it is not possible to calculate FE, FGE by hand [10] and they are difficult to calculate using existing software [201]. This inaccessibility may present a barrier, to practitioners, to the use of these metrics to choose individuals for genetic rescue, therefore a novel solution is required. I Choose You (ICY) is a software tool that has been designed to take studbook data, calculate

FE per individual and use this in combination with mean kinship, to recommend a group of genetically diverse and unrelated individuals for genetic rescue. ICY has been designed to be quick and easy for practitioners to use so that they can make decisions about which individuals have the most genetic diversity without the need to consult specialist molecular ecologists or conservation geneticists which can be a further barrier to including genetics in a conservation plan [48, 26].

5.2 Methodology

The studbook data used for all the examples and analysis within this chapter were exported by the pink pigeon's studbook keeper, Harriet Whitford, Durrell Wildlife Conservation Trust, on 16/01/2019 and the information contained within is current to 21/12/2018. A copy of the studbook, the code needed to run ICY and all the scripts used for the analyses in this chapter are available in appendix A.4.

5.2.1 ICY's algorithm

ICY (I Choose You) is an easy-to-use, web-based tool that recommends a group of individuals for reintroduction. The chosen group should have high genetic diversity whilst remaining as distantly related as possible to reduce inbreeding. ICY measures this genetic diversity using the number of founder equivalents (FE) across all individuals in the contemporary population, which can be calculated from studbook data using Equation 5.1.

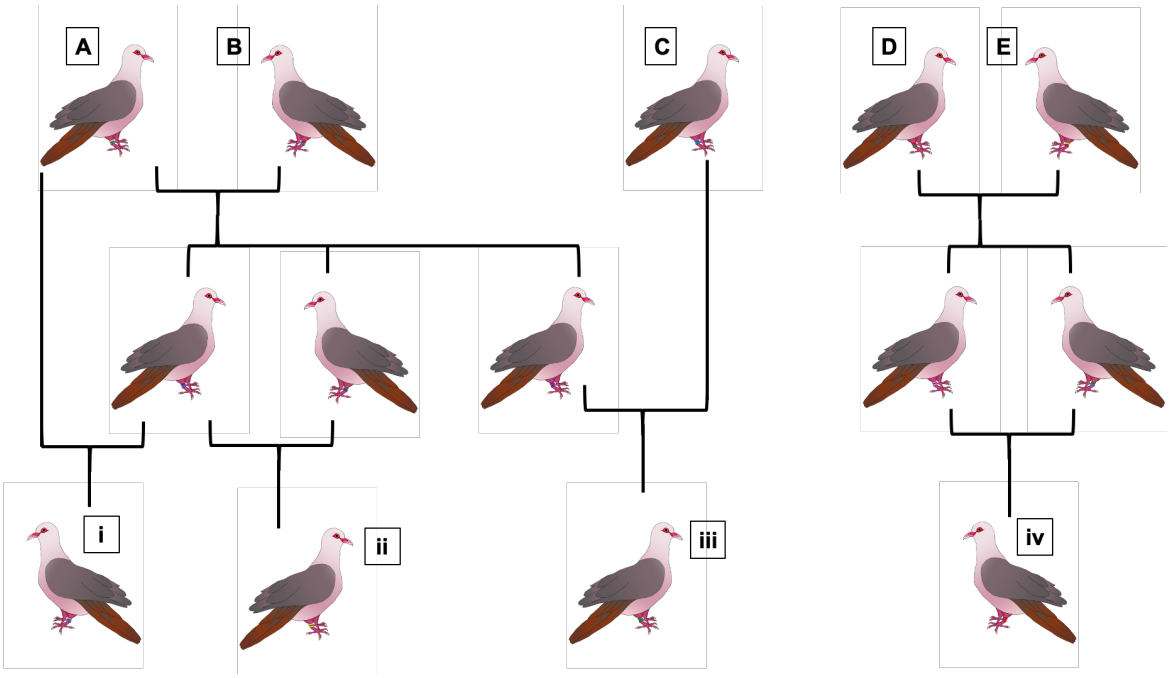
$$FE = \frac{1}{\sum(pi^2)} \tag{5.1}$$

Where FE is the number of founder equivalents and pi is the proportion of genes retained from founder i [10].

An overview ICY's algorithm is described in Table 5.1 and Figure 5.3 shows an example of the process ICY uses to select individuals for reintroduction. Briefly, ICY calculates founder equivalents for each living individual in the population and then ranks individuals first based on FE (the higher the number of founder equivalents the higher the theoretical genetic diversity of that individual) and then based on their MK (see Equation 5.2).

$$MK_i = \frac{\sum_{j=1}^{N} k_g}{N} \tag{5.2}$$

Where K_g is the kinship between i and j and N is the number of individuals in the population, mean kinship is not calculated in ICY but is automatically calculated by PMx using the "Genetics" module. This is a static mean kinship that, unlike a dynamic mean kinship estimate, does not change after individuals are chosen or discarded. Therefore there could be the risk that closely related individuals who are likely to have similar mean kinships may be selected as they are likely to be ranked in similar positions. This would not be an optimal strategy for a genetic rescue because not only would you be limiting the genetic diversity introduced into a population (as related individuals share large proportions of their genomes) but there is also the risk this would lead to increased inbreeding after release. It is for these reasons that the pairwise kinship estimates between each chosen individual and the remaining individuals in the captive population are also used to ensure that none of the individuals selected are more closely related than a user defined threshold. Overall by minimising mean kinship it is hoped to increase the genetic diversity and help retain founder alleles which are present at lower frequencies, for an individual with a given number of founder equivalents (this is illustrated in Figure 5.2).



(a)

Individual	A	B	C	D	E	FE
i	0.75	0.5	0	0	0	1.231
ii	0.5	0.5	0	0	0	2.000
iii	0.25	0.25	0.5	0	0	2.667
iv	0	0	0	0.5	0.5	2.000

Individual	i	ii	iii	iv	MK
i	0.625	0.3125	0.125	0	0.266
ii	0.3125	0.625	0.125	0	0.266
iii	0.125	0.125	0.5	0	0.187
iv	0	0	0	0.625	0.156

(b)

Rank	Individual	MK	FE
1	iii	0.187	2.667
2	iv	0.156	2.000
3	ii	0.266	2.000
4	i	0.266	1.231

(c)

Figure 5.2 An illustration of how ICY uses FE and MK to select the most genetically diverse individuals. Starting from a pedigree of a captive population **A** where individuals A,B,C,D,E represent the founders and individuals i,ii,iii,iv represent the current population both MK and FE values can be calculated for each individual **B**, this population is inbred which is reflected in their self MK values which are >0.5. The combination of MK and FE can then be used to rank the individuals from those with high to low genetic variation **C**. Individual iii ranks highest is represented by three out of the original founders and (taking into account uneven representation) has 2.667 founder equivalents and a low mean kinship. For both individuals iv and ii FE = 2 however there has been a greater amount of inbreeding between individual A,B and their offspring which has resulted in individual ii having a much higher MK than iv and therefore ii is ranked higher. The lowest ranking bird is the highly inbred individual i which has both a high MK and low FE. This example is based on an example using golden lion tamarin (*Leontopithecus rosalia*) from [11, page 440] and does not represent real pink pigeon data. Images courtesy of Dr Rebecca Gooley

ICY then selects the highest ranking individual and, using the pairwise kinship information from the studbook, ensures it is less related to the next highest ranked individuals than the user defined threshold. This is an iterative process with individuals either being added to the group or discarded until all the input parameters have been met or ICY realises it is not possible to suggest, for example, a group of three females and 1 males that are completed unrelated MK = 0.

Table 5.1 Overview of ICY, web-based tool for choosing individuals with relatively high genetic diversity, low relatedness, and low inbreeding for the purpose of genetic rescue.

Input:	Genetic data from studbook exported from PMx Kinship matrix from studbook exported from PMx [optional] Founder information from studbook exported from PMx Number of males and females wanted Maximum mean kinship allowed [optional] IDs of individuals previously released [optional] Filters for age and location of individuals
Output:	Table containing the most suitable individuals for release, their inbreeding coefficient, MK, number of founders and number of FE Graphs for each individual showing founder representation. If founder information is provided this will also include a comparison with the mean founder representation for the population. Downloadable report containing the same information as presented on screen as well as a detailed description about how ICY works, founder information for all individuals currently in the population and all individuals currently living ranked based on MK and then FE.
Steps:	**1.** Load input files **2.** Apply any optional filters (such as age or locations) and create a formatted data-frame **3.** Calculate the number of founder equivalents (FE) for each individual using Equation 5.1. **4.** Rank individuals by sorting the data frame first by FE then by MK

5. Add highest ranking individual to a list containing 'chosen' individuals

6. Compare the chosen individual to every individual in the data set and remove, from the working data-frame, any individuals whose pairwise kinship to the chosen individual is greater than the relatedness threshold.

7. Repeat step 6 until the required number of male and female birds have been chosen or until there are no more birds to choose from - i.e. all have been deleted as they are too related to the birds already chosen. For an example of how this iterative process works see Figure 5.3

Software required: Python3

R v3.6.1

Latex

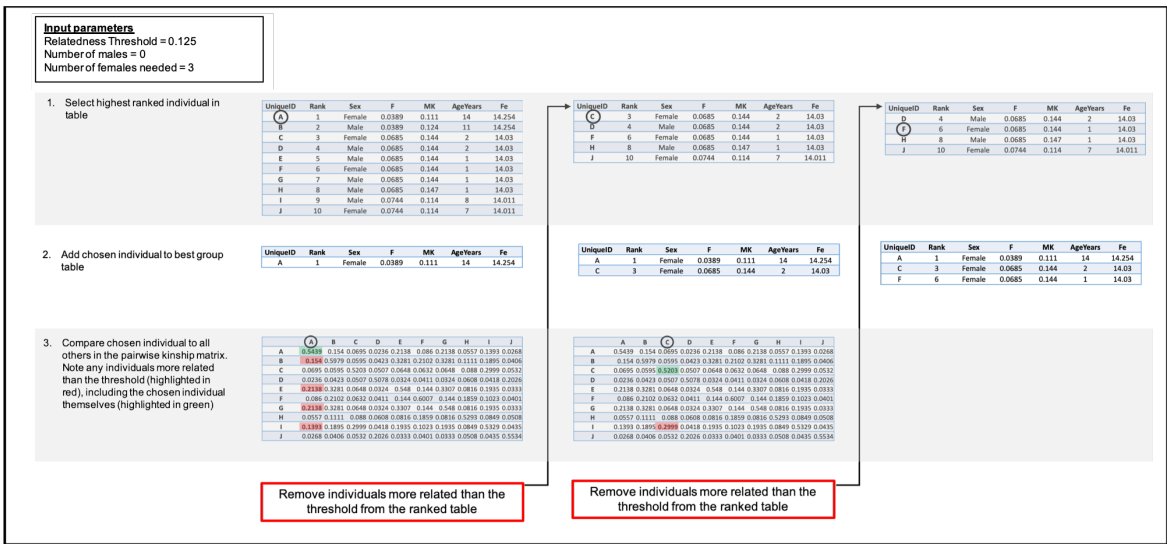


Figure 5.3 An example of the iterative process used by ICY to provide the user with recommendation of a group of individuals for genetic rescue. In this example ICY has been asked to choose a group of three female individuals, all of which must have a pairwise kinship less than 0.125.

5.2.2 Proof of concept

A proof of concept test was designed to assess if ICY would choose a group of birds with greater genetic diversity and a lower relatedness than randomly selecting birds. To compare the results of ICY with randomly selecting individuals, ten groups of three randomly selected males and three randomly selected female birds were created using a bespoke python script (see appendix A.4). Each group of individuals was entered into PMx v1.5.6 [8] as if they were a standalone population of six individuals and the genetic module was used to analyse the genetic metrics (described in Table 5.3) of each group. Without empirical genetic data practitioners rely on the genetic information produced by studbook software [316] therefore it is appropriate to use these same metrics to compare the different methods of selecting individuals for genetic rescue. Note that the purpose of creating random scenarios is to reflect the occasions when practitioners select individuals based on no genetic information (only pragmatic considerations such as age, health, reproductive potential etc.) as initially happened with the pink pigeon.

Although it would be more rigorous to try every possible combination of three males and three females and see if ICY chose the best combination this was not feasible because each possible group has to be manually entered into PMx. Instead another test was designed to try and gauge how well ICY performed compared to randomly choosing groups of birds by comparing the number of FE in the population of six birds chosen by ICY and 1000 randomly chosen groups of three females and three males.

Table 5.3 A description of the metrics produced by PMx and used to compare the groups of birds chosen by ICY to group chosen randomly, group chosen based on mean kinship or mean kinship and pairwise relatedness

Metric	How PMx calculates metric
Gene diversity	Probability that two alleles at the same locus sampled at random in the population are identical by descent. Calculated as the heterozygosity of founder alleles (1 - inbreeding)
Gene Value	The gene diversity weighted for the reproductive value of individuals
Founder genome equivalents (FGE)	The number of unrelated individuals that would represent the same amount of diversity
Founder genomes surviving	Sum of allelic retention of the individual founders. Where allelic retention is the probability that an allele contributed by a founder exists in at least one living descendant
Mean inbreeding	Mean inbreeding coefficient among living individuals weighted for known pedigree

5.2.3 Exploring pink pigeon studbook data

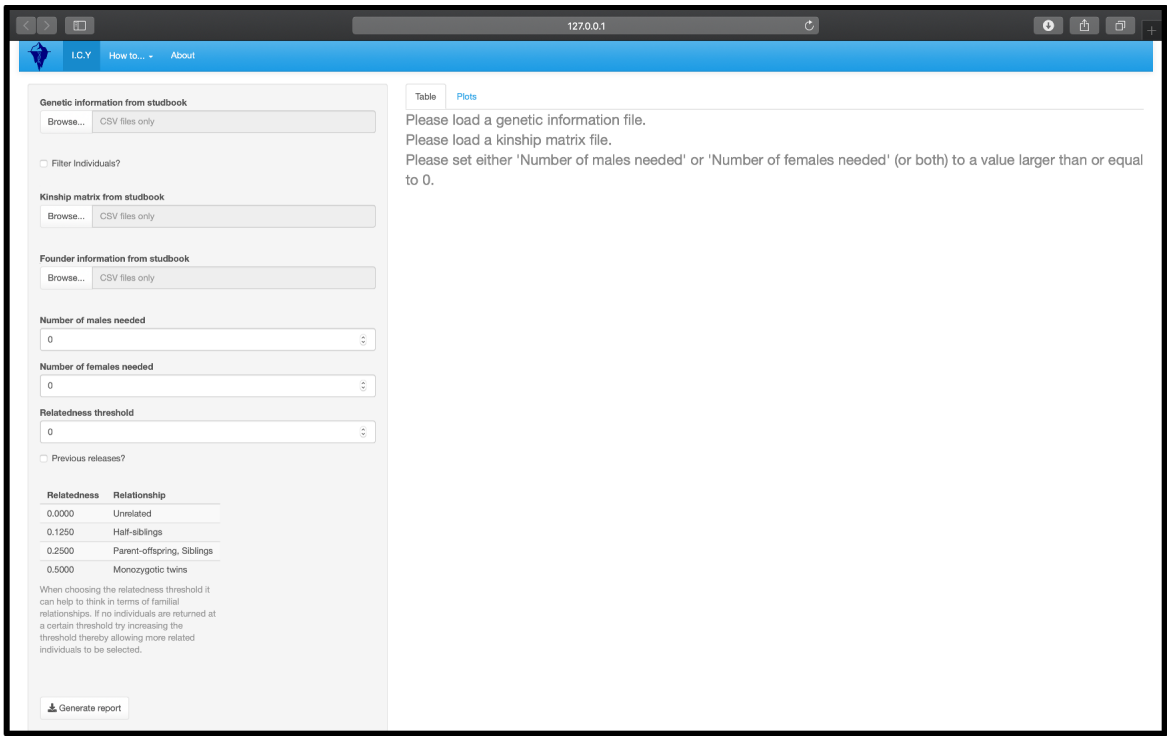
Data from the studbook were used to examine how the pink pigeon captive population had been managed over time and how many of the original founder lineages have been retained. The studbook was first processed in PMx v1.5.6 [8] then further exploration undertaken in R v3.6.1 [291].

5.2.4 Distribution

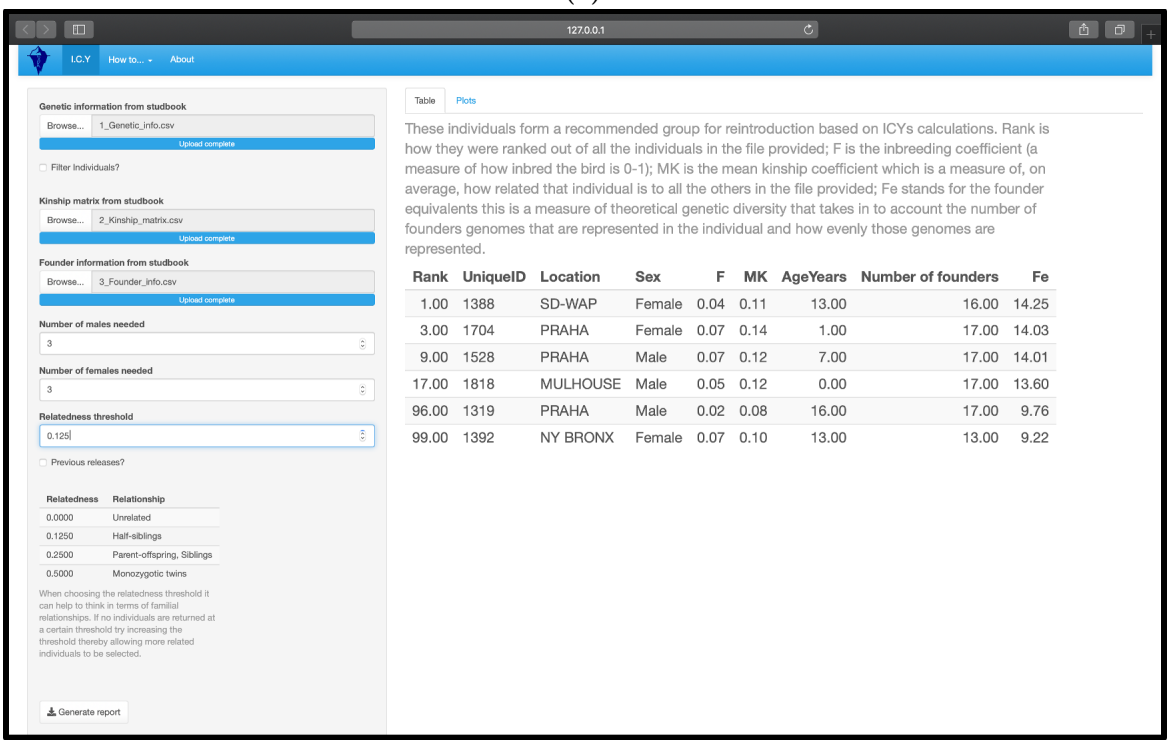
ICY is written in R v3.6.1 (R Core Team 2019) and uses Python v3.7.4 functions to format studbook data, calculate founder equivalents and select the best group of individuals for reintroduction. The app is distributed as a docker image which has a container running at: https://dry-caverns-77518.herokuapp.com/. ICY is open source and available at: https://github.com/Mills33/ICY. This code is also available in appendix A.4.

5.3 Results

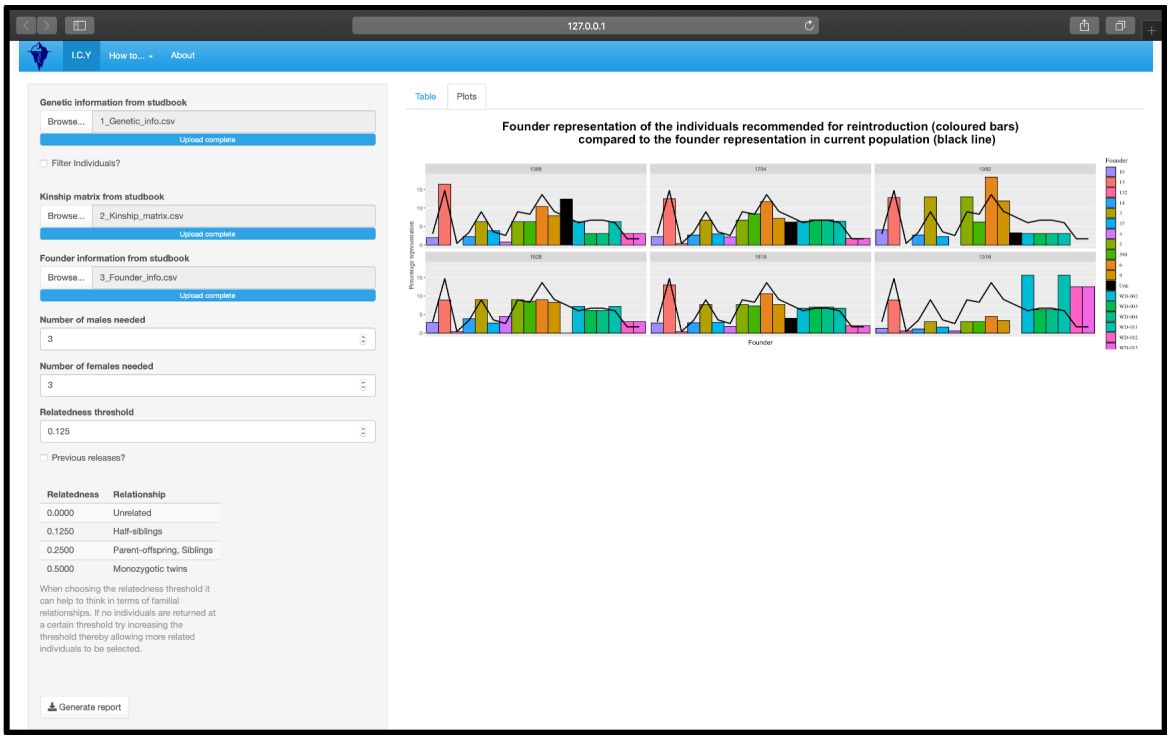
ICY is an easy-to-use web based tool based that was written using R's Shiny architecture. ICY has a clear interface (see Figure 5.4), requires minimal parameterisation, and can be run locally or through a web browser meaning it should be accessible for all practitioners regardless of location or internet facilities. Furthermore ICY's algorithm is computationally efficient giving near instantaneous results, which work well within ICY's reactive framework, and gives consistent results. This means different users, who have access to the same data set, will get the same answer (assuming the same parameters are entered) this is useful for practitioners collaborating in different places.


127.0.0.1
I.C.Y
How to...
About
Genetic information from studbook
Browse...
CSV files only
Filter Individuals?
Kinship matrix from studbook
Browse...
CSV files only
Founder information from studbook
Browse...
CSV files only
Number of males needed
0
Number of females needed
0
Relatedness threshold
0
Previous releases?
Relatedness
Relationship
0.0000
Unrelated
0.1250
Half-siblings
0.2500
Parent-offspring, Siblings
0.5000
Monozygotic twins
When choosing the relatedness threshold it can help to think in terms of familial relationships. If no individuals are returned at a certain threshold try increasing the threshold thereby allowing more related individuals to be selected.
Generate report
Table
Plots
Please load a genetic information file.
Please load a kinship matrix file.
Please set either 'Number of males needed' or 'Number of females needed' (or both) to a value larger than or equal to 0.

(a)


127.0.0.1
I.C.Y
How to...
About
Genetic information from studbook
Browse...
1_Genetic_info.csv
Upload complete
Filter Individuals?
Kinship matrix from studbook
Browse...
2_Kinship_matrix.csv
Upload complete
Founder information from studbook
Browse...
3_Founder_info.csv
Upload complete
Number of males needed
3
Number of females needed
3
Relatedness threshold
0.125
Previous releases?
Relatedness
Relationship
0.0000
Unrelated
0.1250
Half-siblings
0.2500
Parent-offspring, Siblings
0.5000
Monozygotic twins
When choosing the relatedness threshold it can help to think in terms of familial relationships. If no individuals are returned at a certain threshold try increasing the threshold thereby allowing more related individuals to be selected.
Generate report
Table
Plots
These individuals form a recommended group for reintroduction based on ICYs calculations. Rank is how they were ranked out of all the individuals in the file provided; F is the inbreeding coefficient (a measure of how inbred the bird is 0-1); MK is the mean kinship coefficient which is a measure of, on average, how related that individual is to all the others in the file provided; Fe stands for the founder equivalents this is a measure of theoretical genetic diversity that takes in to account the number of founders genomes that are represented in the individual and how evenly those genomes are represented.
Rank UniqueID Location Sex F MK AgeYears Number of founders Fe
1.00 1388 SD-WAP Female 0.04 0.11 13.00 16.00 14.25
3.00 1704 PRAHA Female 0.07 0.14 1.00 17.00 14.03
9.00 1528 PRAHA Male 0.07 0.12 7.00 17.00 14.01
17.00 1818 MULHOUSE Male 0.05 0.12 0.00 17.00 13.60
96.00 1319 PRAHA Male 0.02 0.08 16.00 17.00 9.76
99.00 1392 NY BRONX Female 0.07 0.10 13.00 13.00 9.22

(b)



(c)

Figure 5.4 Images showing ICY's interface, for clarity the main writing has been enlarged. **(a)** Is the main page loaded when ICY is started it has tabs on the top that provide users with guides on how to use ICY. Down the left hand side of the page is where all files are uploaded and parameters entered. As soon as the necessary files have been uploaded, and a number of males/females and relatedness threshold has been entered ICY will automatically calculate the optimal grouping **(b)**. If any of the parameters are changed the table will automatically change. By using the *Plots* tab above the table users are also able to examine the founder representation in each chosen individual **(c)**. In this example a founder information file was uploaded and therefore the black line shows the founder representation for the population as a whole whilst the coloured bars show the founder representation in one of the chosen individuals.

5.3.1 Proof of concept

ICY gave better results for almost every metric measured, when compared to choosing individuals randomly (see 5.5). ICY chose a group which had a higher gene value on average than selecting a random group but there were random combinations that produced higher gene values. Gene value is the gene diversity but weighted for reproductive value which is defined by PMx as the number of same-sex offspring produced this year and in future years by an individual of age x.

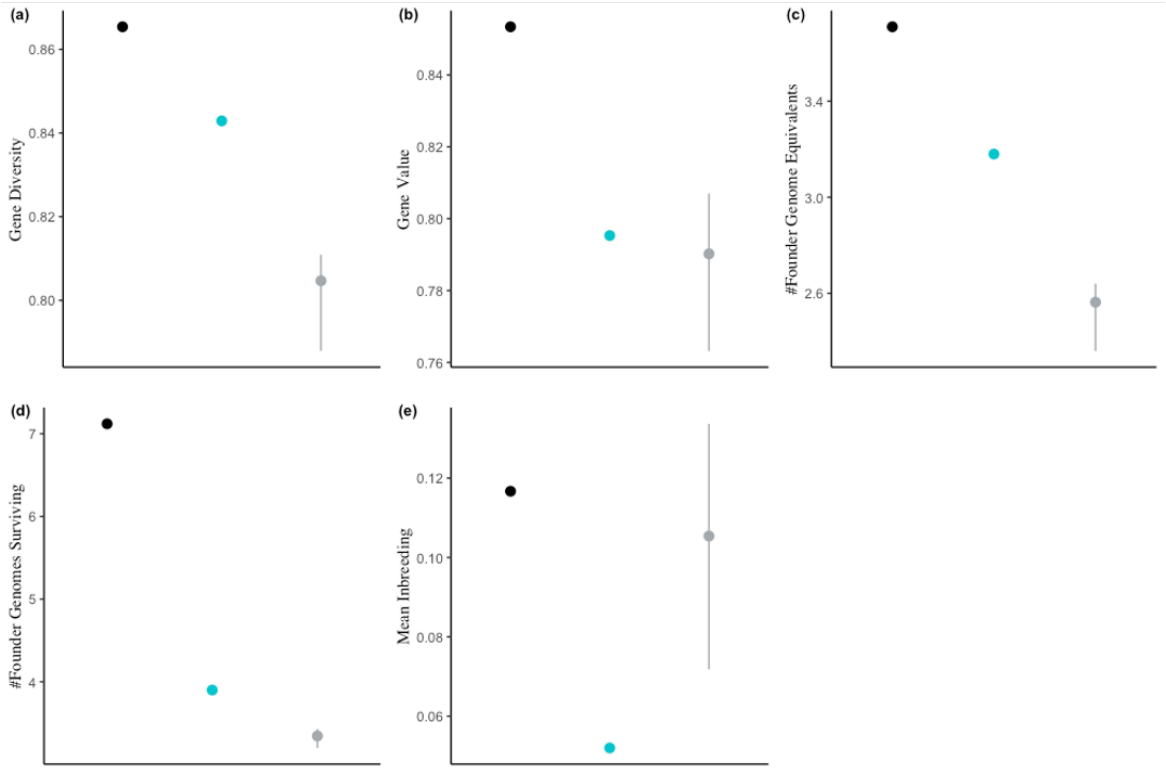


Figure 5.5 The results from the proof of concept experiment using PMx. Included are the values from the whole population (black), ICY (blue) and the mean values of ten sets of randomly chosen individuals (grey dots) where grey bars represent the range of values obtained across all ten groupings. For both ICY and each random set three males and three females were selected.

When only considering the number of FE in the chosen "population" of six individuals ICY performed well compared to 1000 randomly chosen groups (see Figure 5.6) with only a single random group having a marginally higher FE (FE = 14.197) than ICY's suggested group (FE = 14.163))

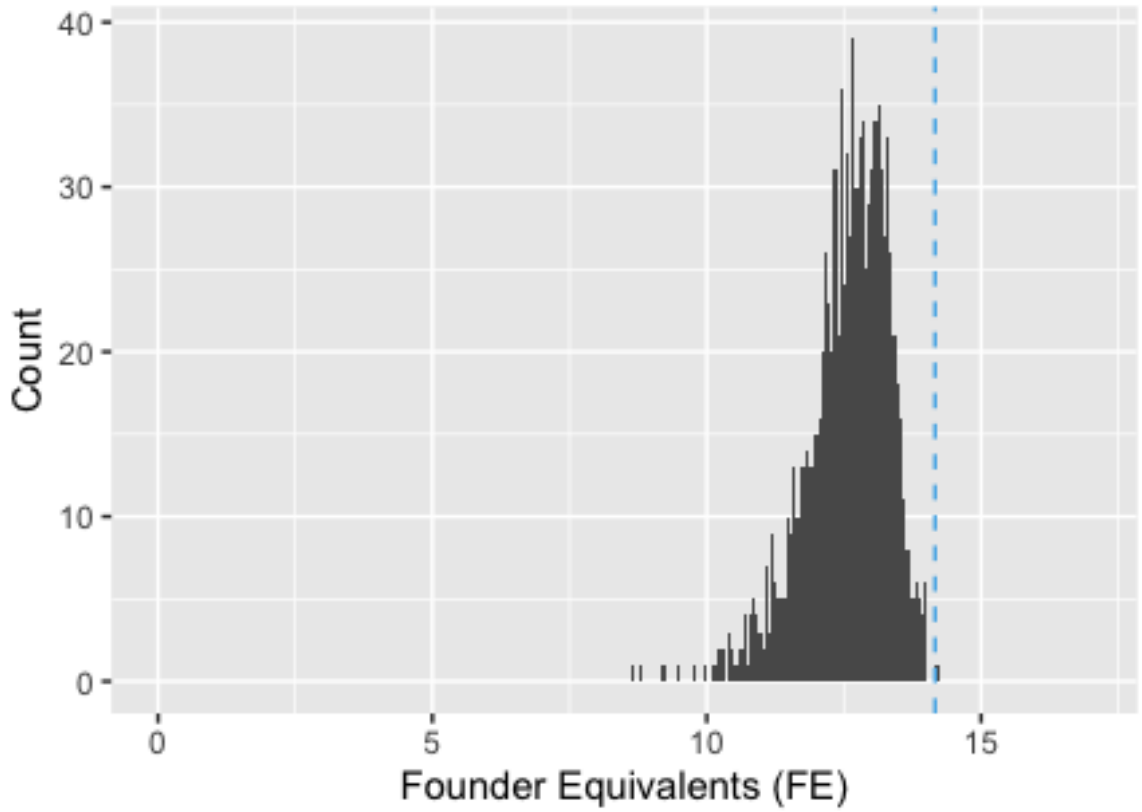
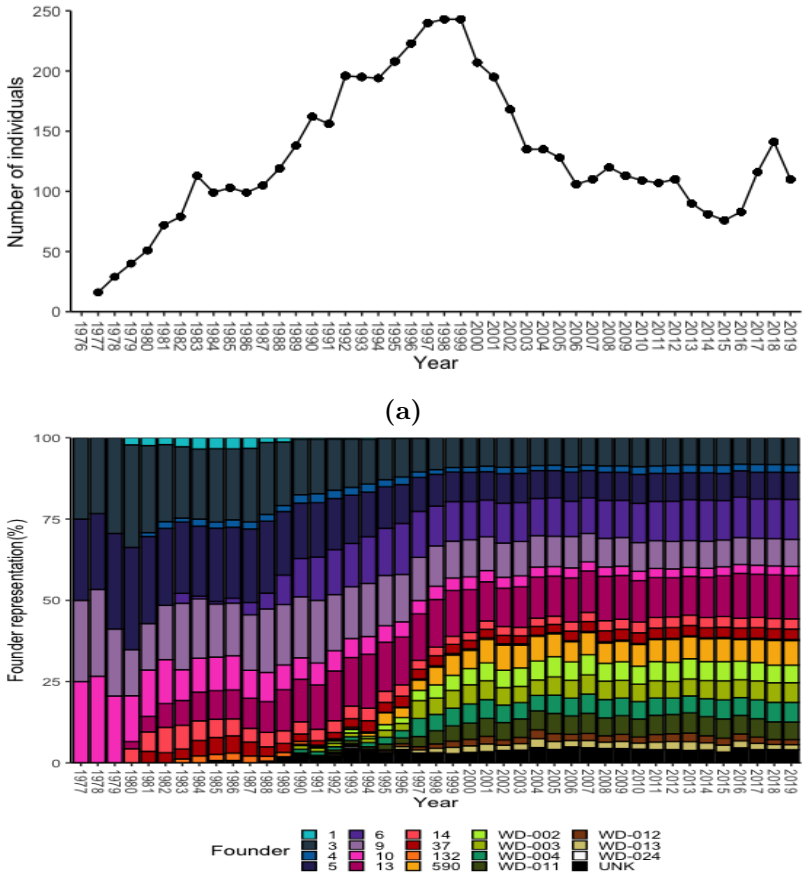


Figure 5.6 The results comparing the number of FE in the population chosen by ICY to 1000 randomly chosen groups. The distribution shows the results of calculating the FE for each of 1000 randomly chosen groups of three male and three female birds, the number of FE in the group chosen by ICY is represented by the dashed blue line (14.163).

5.3.2 Exploring data

Founder information

Extracting information about the founders shows that all original 17 founders are still represented in the current population although some at low proportions (see Figure 5.7). Founders lineages represented by low proportions are in danger of extinction for example, founder 132 (orange) is present in very low proportions in the current population.



(a)

(b)

Figure 5.7 (a) The pink pigeon captive population from founding (1977) until current (2019). The number of pink pigeons in captivity and **(b)** how the founder representation of the population as a whole has changed over time the different colours represent different founders except for black which represents the proportion of the population that had unknown representation. **NB** Although Founder 1 and WD-024 have been shown here as founders they never produced any offspring and therefore are not true founders and are not included as part of the total number of founding birds (n = 17).

5.4 Discussion

With captive breeding programmes playing an increasingly active role in conservation, there exists the need for practical tools, designed for practitioners, that are easy to learn, quick to use and use data that are already available. ICY is an example of such a tool and is designed to recommend a group of individuals (high genetic diversity, low inbreeding and relatedness) for genetic rescue which would retain as much genetic diversity as possible and positively contribute to a reintroduction.

ICY chose a group with higher genetic diversity, lower inbreeding and greater retention of founder alleles than any of the ten randomly chosen groups. Although ten is a small sample size, and more rigorous testing may be required in the future, nevertheless these results are encouraging. There was a single metric, gene value, for which ICY did not perform as well as each of the ten random groups, however this result was likely skewed by ICY choosing young individuals (< 1 year). Gene value represents the heterozygosity expected in progeny produced by random mating, if the population was in Hardy-Weinberg equilibrium, and if each individual were to the produce the number of offspring expected from its age-based reproductive value. However this may not have be the most appropriate metric to use for captive bred individuals like the pink pigeon whose ability to breed is determined by management and therefore data detailing the ages and number of young per reproduction are unlikely to reflect biological capability. As well as being a poor reflection of biological capability, pragmatically young (< 1 year) pink pigeons are preferred for translocations (Carl Jones pers. comms.).

ICY also performed very well when the founder equivalents (FE) in the population of six birds chosen by ICY was compared to the FE of 1000 randomly chosen populations of three males and three females. Only a single random group had a higher number of FE than ICY and the difference between the values was tiny, 14.197 compared to ICY's 14.163. It is possible that better groupings do exist because the design of ICY's algorithm (ranking individuals and considering them individually) does not guarantee that it will pick the most optimal grouping. It may be possible to find the optimal grouping (for a given number of male and female birds) using an exhaustive search algorithm where the FE and MK of every single grouping is considered and then the best combination chosen. Another possible improvement to the algorithm would be to use founder genome equivalents (FGE) (instead of FE). FGE have been used to judge the success of reintroductions [316] and performed the best when testing metrics to be used for selecting groups for reintroduction [319]. FGE is similar to FE but takes into account the random drift of alleles through the pedigree and as such it a more accurate estimate of the current diversity of the population when compared to FE and this is especially true for small populations [10]. This difference can be seen by comparing the number of FE calculated by ICY for the chosen group (14.163) to the FGE calculated by PMx (3.18) (see Figure 5.5). Whilst FGE may be more accurate it can only be estimated and only using gene drop simulations. This is computationally more intensive then calculating FE which has also been proven to be highly effective at selecting grouped for reintroduction [319]. In the future it may be possible to integrate a more exhaustive search algorithm which also computes FGE to select the optimal groupings of individuals for any release however such an algorithm would become more complex and computationally intensive and the challenge would be to integrate an algorithm within the current easy-to-use and reactive framework.

Although ICY may not choose the optimal grouping it performs better than random choice and provides a simple-to-use tools that practitioners can run locally or online. One of the main purposes of ICY was to make it easy for practitioners to extract information on the genetic suitability and diversity of individuals for a release program. When speaking to a practitioner who had used the results of ICY they stated the most useful part was the table that was produced in the downloadable report (see appendix A.4). This table ranks every individual in the population and provides a quick and easy overview of the population as a whole. The current version of ICY represents a large improvement compared to the current options of random choice or the use of more complex software packages and analyses that require more manual involvement and that may not be known to practitioners [201].

5.4.1 Considerations and future improvements

An important consideration for a genetic rescue is that whilst you are augmenting a wild population with novel genetic diversity you are also removing this same diversity from the captive population. Frankham (2012) suggested a framework (see Figure 5.8) that could be used to evaluate the overall merits of reintroducing each individual, based on whether it was genetically valuable to the captive population, the wild population, both or neither. The conclusion was that initially individuals with genotypes over-represented in both populations should be reintroduced. Whilst this strategy may not reintroduce beneficial diversity to the wild population it would allow practitioners to work out effective reintroduction strategies (reintroductions often have high mortality) without the risk of losing genetically valuable individuals [11]. However without pedigree or molecular data from both the wild and captive population such a framework is hard to apply nevertheless these are important considerations. ICY directs the users attentions to these principles in the documentation.

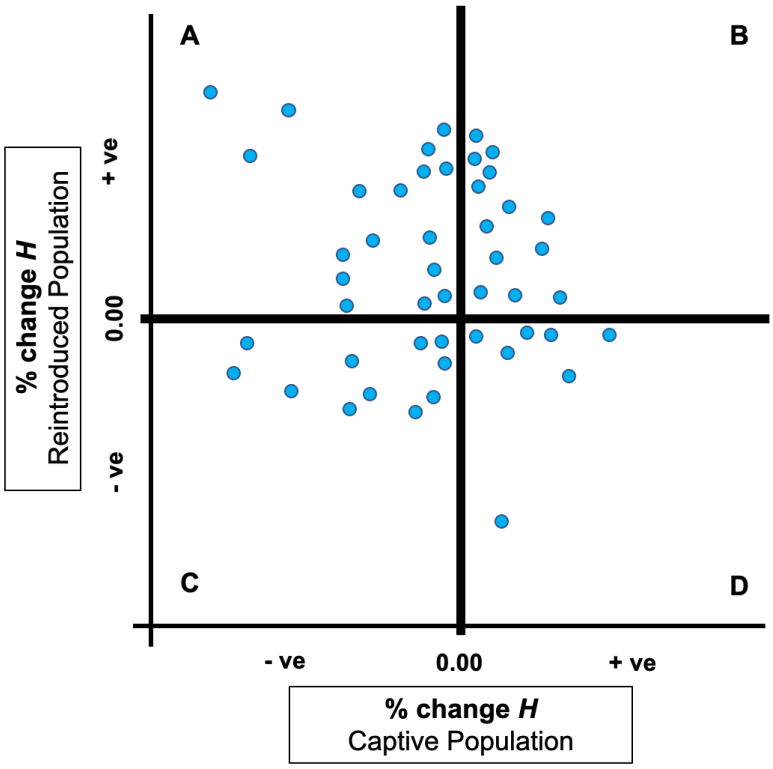


Figure 5.8 An illustration of the possible conflict between choosing individuals to release into the wild and maintaining a healthy captive breeding program [adapted from [11, pg.463]]. Each point represents an individual and how choosing that individual would impact the captive population (x axis) or the reintroduced population (y axis). Choosing individuals from quadrant **A** would be beneficial for the reintroduced population but detrimental to the captive population, these would be individuals that are genetically valuable and have few wild relatives. Choosing individuals in **B** would be beneficial to both populations likely representing genetically over-represented captive individuals who have few reintroduced relatives. Choosing individuals in **C** would be detrimental to both populations as they are under-represented in the captive population and have many wild relatives. Whereas individuals that fall in quadrant **D** are over-represented in both populations so would be detrimental to the reintroduced population but their release is good for the captive population. It is recommended that initially individuals in **D** are used because of the high mortality associated with the release of captive-bred individuals at the beginning of a reintroduction project. Although this framework is useful for visualising the impact of choosing different individuals it may not be directly applicable to every situation. For example if an entirely new population is being formed then there will be no wild relatives and the choice then will be selecting individuals that are not genetically valuable and under-represented in the captive population.

Currently ICY has only been tested using captive data from the pink pigeon, which has been bred in captivity for more then 40 years. During that time, despite the fluctuations in the size of the captive population, the founder representation has stayed relatively constant and lineages from all the original founders are represented Figure 5.7 . Because of this most birds have a reasonably high number of founder equivalents (see report in appendix A.4) which could explain the narrow spread of values seen when selecting groups of individuals for

release Figure 5.6. Therefore it is possible that the empirical gains from using ICY could be much greater for a species that has a greater skew in founder representation than the pink pigeon. In the future ICY should be tested using studbook data from a number of different species to test this and confirm the results presented here. Another difference between the pink pigeon and other captive animals is that their studbook records and pedigree, are also almost complete (96%). Many captive animals, particularly group animals, have a greater proportion of missing data in their pedigree because of the difficulty in assigning parentage [201], pink pigeons are monogamous and paired by keepers therefore assigning parentage is usually trivial. The calculation of founder equivalents by ICY relies on pedigree data and whilst ICY will calculate the percentage of unknown representation in an individual it does not currently have a method to weight the amount of missing data in an individual. Users wishing to use ICY with a highly uncertain pedigree are recommended first to use some of the analytical methods in PMx to further complete their pedigree. The ability to weigh individual unknown representation is something that may have to be considered in future versions of ICY.

Many of the considerations required to use ICY most appropriately (for example percentage pedigree known) could be empirically answered with molecular data. Whilst ICY was designed specifically to allow practitioners without molecular data to gain greater access to their data and infer genetic diversity the methods presented here are theoretical. As the price of genetic data continues to decrease its use is to be encouraged to remove the ambiguity present in many of the theoretical methods. However as long as it remains unfeasible for any practitioners to use genetic data, there exists the need for tools such as ICY that can consider genetic diversity.

ICY was created to help practitioners choose genetically suitable animals for reintroductions. Reintroduction and genetic rescues require a huge amount of work from people working in different roles (often countries) and time is usually limited [111]. Therefore any tools aimed at practitioners should be quick, easy-to-use and understand and provide solutions to specific problems. ICY was designed specifically for practitioners in these situations (such as studbook holders and keepers), it is free and easy-to-use with extensive 'How to' documentation. However ICY's simple design means that it does not have some of the flexibility of the more complex programs such as PMx. Specifically there are currently only two filter that alter the selection of individuals from the studbook data provided. Instead ICY provides detailed explanations in the 'How To' guides about to how to use PMx and what considerations may be needed to ensure a suitable selection of individuals are used as input for ICY. For example there is no point considering individuals that you have no intention of reintroducing such as those with known adaptation to captivity or an illness.

As input ICY takes data generated from an animal's studbook, three or four user-defined parameters and up to three filters (see Table 5.1). ICY's algorithm ensures only data from living individuals are used but does not, currently, take raw studbook data from Sparks as input. Instead ICY requires at least two specific comma separated files which can be easily exported from software like PMx with details of how to do this provided in ICY's documentation. However in the future it would be best if ICY were capable of directly parsing a Sparks file this would again make it simpler for practitioners to use and may encourage more people to use ICY.

5.5 Conclusion

ICY is an easy-to-use, open source tool to help practitioners select individuals for genetic rescue and reintroductions using widely available studbook data. Using studbook data from the pink pigeon as an example it was demonstrated that ICY provides a more genetically suitable group of birds (low relatedness, high diversity) than selecting groups randomly. Whilst the use of molecular data will increase in zoos and captive populations ICY is the first tool to provide practitioners, without access to such data, an understandable tool to consider genetic diversity when planning genetic rescue.

Chapter 6

General discussion

The current loss of biodiversity is so great that it may surpass that of ancient mass extinction events [14, 326, 269]. The defining feature of the current mass extinction is that the major drivers are anthropogenic therefore [327], in theory, society has the ability to slow, stop and maybe reverse the declines in biodiversity. Not only can it be argued this is a moral imperative but it is also in human-kind's best interest to conserve biodiversity and preserve ecosystems for example, to ameliorate the impacts of climate change or reduce the incidence of pandemics [269]. The effective conservation of biodiversity means establishing self-sustaining populations that are able to persist and adapt to future challenges [54] to try and achieve this goal genomic data can be used to study a species adaptive potential and inbreeding and the results integrated into management [328, 54]. To contribute to effective conservation the results produced from genomic data must accurately represent the biology of the study species rather than the data processing pipelines used to create them. This requires careful analysis and novel approaches that focus on recovering the maximum amount of unbiased information from the available genetic data. Therefore better tools will lead to better genomic resources and more accurate analyses which should enable more insightful conservation practices.

Within this thesis two novel tools (RADiKal and ICY) are presented and two resources created (pseudo-chromosome assembly and Vortex model) that can be used by researchers studying the pink pigeon (*Nesoenas mayeri*) or other endangered species. The key findings from each chapter are discussed below as well the implications for pink pigeon management and more broadly the role of data analysis, and programming in the field of conservation genetics.

6.0.1 Chapter 2

High quality, contiguous reference genomes are a valuable resource for conservation genetic projects because they provide a coordinate system for comparison between individuals, increase the accuracy of analyses and enable the study of mechanisms such as adaptive potential and inbreeding depression [63]. A high quality pseudo-chromosome level genome assembly was produced for the pink pigeon and used to examine levels of variation and inbreeding in the current captive population. The primary finding was the captive population has relatively high levels of variation across its genome and has similar levels of heterozygosity to non-threatened species (e.g. humans). This is an encouraging, if surprising, result given the history of the pink pigeon. One explanation for the high genetic diversity of the captive population is that they have been managed well as part of the captive breeding programme. This observation is supported by the results of ICY in chapter 5 which showed that the pink pigeon still retains contributions from all founders (n = 17) and that most individuals have a relatively high number of founder equivalents and therefore are predicted to have high levels of genetic diversity. Genetic diversity is only one measure of genetic health and although the former has been used as a proxy for adaptive potential this work adds support to an increasing number of studies that have indicated that the links between genome-wide variation and adaptive potential are not necessarily straightforward [146, 147]. However access to a high quality genome will enable studies of adaptive potential especially if the genome could be annotated, this could be done automatically using annotations from the chicken (*Gallus gallus*) genome. This may allow researchers to better understand if the pink pigeon has the capacity to adapt to future challenges, for example diseases such as avian malaria.

Although I found comparably high levels of variation in the pink pigeon genome, it also showed a fairly high inbreeding coefficient. Remarkably I could not detect any signs of recent inbreeding at the molecular level, as measured by runs of homozygosity. This apparent lack of recent inbreeding may reflect the excellent management of the pigeon, although it also may reflect a potentially high recombination rate in the micro-chromosomes [177], which is expected to break up the runs of homozygosity, making the signature of inbreeding look older than it genuinely is [5]. Indeed, in the past the pink pigeon has displayed signs associated with inbreeding depression such as low fecundity and skeletal deformities [120]. These may represent isolated incidences, may not be representative of long term inbreeding and/or may be because the pink pigeon has a high genetic load. A high genetic load, is also supported by the calculations of lethal equivalents (one measure of genetic load) in Chapter 4 and could increases the frequency of inbreeding depression [124]. The precise mechanism of inbreeding depression in the pink pigeon may be clarified by further research using this

genome assembly, for example the identification of deleterious variants in ultra-conserved elements, and loss-of-function mutations in annotated genes. However, although the study of deleterious variants is enabled by the presence of a pseudo-chromosome genome assembly, it also requires phenotypic data which is difficult to obtain from wild pink pigeons.

The conclusions drawn from these whole genome analyses differ from those from previous studies which found low levels of genetic variation and inbreeding [129, 9, 2]. These discrepancies may be because there is a significant difference temporally (many of the studies were conducted 20 or more years ago), spatially (Mauritius versus *ex situ* populations) or is the result of different data-types (whole genome data, versus molecular markers). Further sampling is needed not only from the captive population (to increase the sample size of this study) but also from the current wild population to clarify the reason for these discrepancies.

6.0.2 Chapter 3

In a previous study individuals from the wild pink pigeon population were sequenced using a reduced representation sampling (RAD-seq) and the results from that study concluded that the pink pigeon had low levels of diversity [2] however current methods of processing RAD-seq data are known to underestimate diversity [240, 258]. This underestimation is frequently overlooked especially if the study species is believed to have low diversity. Although such a prior assumption may seem sensible the results of processing should not be justified using prior assumptions, but need to be validated based on the properties of the raw data. RADiKal was created to access the signal in the raw data and present it so that users could easily see an overview of the variation within their study population. The results from RADiKal show that there is variation within the wild pink pigeon population and the patterns of variation are comparable with what has been found in the whole genome sequencing from captive individuals (Chapter 2).

The ability to assess the levels of genome wide variation in a population at-a-glance without the need for extensive, processing and parameterisation is unique to RADiKal and arguably its greatest strength. Furthermore RADiKal offers a solution to a fundamental problem with current processing, which is: what is a RAD site? RADiKal uses the properties of the RAD sites themselves which can be visually seen as heat plots. In the future these properties could be used to classify RAD sites directly from raw data circumventing the need for complex, parameterisation and processing that discards real information and results in biased data sets.

6.0.3 Chapter 4

If accurate information can be obtained from the genome of an endangered species, the next challenge is to bridge the gap between the academic study of conservation genetics and tools that can be used in policy and management. Vortex is an example of such a tool, not only does a framework already exist to include the results of Vortex analyses into IUCN Redlist classifications (category E) but it has the support of the Conservation Planning Specialist Group (CPSG) who are trying to increase the number of species' action plans that include a Vortex model (Jamie Copsey pers. comms.). Vortex models provide a valuable resource as a platform to test management options and hypotheses, however they also require a lot of data and are sensitive to parameterisation. Many parameters can be accurately obtained from either wild or captive individuals but mechanisms such as inbreeding depression are much harder to model accurately and can have a profound effect on the model output. This means that it is important to explore the parameter space of your model to ensure the model is robust and to understand what parameters have the greatest effect on the chosen response variable (for example census size).

The model produced for the pink pigeon was robust and indicated that genetic rescue was the best management option to ensure the long term survival (at least 100 years) of the pink pigeon. An exploration of the parameter space using sensitivity testing showed that increasing productivity would also significantly improve the probability of survival for the pink pigeon. The practical measures that could be taken to improve productivity would be to eradicate predators and disease and increase the number of chicks produced, perhaps using the captive breeding techniques that were developed in the 1970s. Predator and disease control are already a part of the current management plan but a complete eradication of either, seems unfeasible. Increased juvenile mortality and infertility can also be the results of inbreeding depression. This is likely because of the high genetic load (LE = 15.128) present in the pink pigeon (as measured by lethal equivalents). Not only does the pink pigeon have a high genetic load but the proportion of their genome that is identical-by-descent is higher when measured empirically (using ZooRoH in Chapter 2) than was expected using the pedigree. This indicates that the extinction risk may have been underestimated and increases the support for a genetic rescue to help reduce the genetic load and the negative consequences of inbreeding depression.

The results from the Vortex model highlight the importance of the on-the-ground conservation measures such as predator and disease control and advocate the use of genetic rescue as a conservation plan. This adds to the growing body of evidence that calls for genetic rescues to be routinely considered as a conservation management strategy [90, 329, 91, 95, 86, 330, 152, 81**?** , 100, 156].

6.0.4 Chapter 5

Genetic rescues are often logistically challenging and one of the primary considerations is which population or individuals to release [19]. ICY is a software tool that contrives to help practitioners choose which individuals will be the most genetically valuable to reintroduce based on their perceived genetic diversity (measured as founder equivalents) and relatedness (mean kinship). ICY is designed as a simple-to-use web application but can also be downloaded and run on a local machine making it accessible to managers all around the world. Similar to RADiKal, ICY increases the amount of information that can be extracted from a data set and presents its findings in an clear visual manner. As well as getting near instantaneous results on screen, the report produced by ICY is a valuable resource for practitioners. This report allows practitioners to assess the genetic health of a population, see the differences between individuals and have a ranked list containing every individual in the population. The population results in the report for the pink pigeon (see appendix A.4) emphasised the patterns seen in Chapter 2, that the pink pigeon has been managed well. This is evident because contributions from all the original founders have been retained resulting in high levels of diversity and most individuals having a high number of founder equivalents.

ICY is an example of how conservation geneticists can use their skill-set (population genetics theory, bioinformatics, programming) to help practitioners. ICY was created after a biennial meeting organised by the Mauritius Wildlife Foundation (MWF) which brought together practitioners, managers and academics who were working on projects involving Mauritian wildlife. During this meeting the pink pigeon studbook keeper, Harriet Whitford, gave an update on the organisation of the planned pink pigeon genetic rescue using birds from the *ex situ* captive population. The individuals that were to be translocated had already been chosen however other than practical requirements (such as sex, age, location) she was unsure how best to assess which individuals would offer the most valuable genetic contribution. This led to the development of ICY. The focus of ICY was to provide a practical tool that was easy-to-use and presented the necessary information in a clear and concise fashion. This represents an example of how the skills learnt from programming and bioinformatics can be used to help solve real world problems and how effective collaboration can help bridge the "research-implementation gap" [114].

Bridging the "research-implementation gap" in conservation genetics is an ongoing challenge. Collaborative efforts between conservation biologists, bioinformaticians, population geneticists and software developers will help to bridge this gap by producing tools, analyses and resources that are useful for practitioners, thereby contributing to the conservation of endangered species [114, 26]. Effective collaboration must be a priority if the goal of conservation geneticists is to contribute meaningfully to conservation however these priorities

may conflict with the academic impetus of "publish or perish" [24]. This conflict may result from academics producing resources or performing analyses which may not result in a paper or contribute directly to academic research. Importantly, these resources must be completed on a time-scale that is meaningful to practitioners which may differ from academic priorities. For example, if a practitioner approaches a scientist with a current issue it may not be viable for them to for them to wait for four years for the results of a PhD thesis, or the duration of a grant. Nevertheless such collaborations are vital, particularly in conservation genetics/genomics, to ensure that academics are contributing meaningfully to the conservation of endangered species and that genetic data are being integrated into management plans.

6.0.5 Closing remarks

Much of the research in this PhD was opportunistic based on what questions could be answered by the data that was available at any given time throughout the project. I believe this adaptability is common to many conservation projects and the ability to adapt is vital so that when opportunities arise (for example extra funding or sampling) the project is able to capitalise on them. However, even allowing for the opportunistic nature of much of this thesis, with hindsight there are several things I would do differently. Firstly I would have prioritised the genome assembly and included some long read technology to produce a high quality, chromosome level, *de novo* genome assembly that would be a more complete resource for future studies on the pink pigeon. This resource would ideally have also included basic annotation of the genome using sytneny. I would also have tried to collect samples from birds currently in the wild population to use for both RAD sequencing and whole genome analysis. The latter would have made the most difference to the current analyses and would have allowed a direct comparison between the genomes of captive and wild birds. The ability to collect and sequence samples was inhibited by both the logistics of sampling from endangered animals and funding however with prior planning and applications for further funding this may have been possible. Finally at a recent (November 2020) meeting about the conservation of the pink pigeon and other Mauritian wildlife I realised that I had made an error in not speaking directly to the current field team about the pink pigeon. Although I had spoken with many people involved in pink pigeon conservation in Mauritius it would have been useful to get the information first hand from the current field team whose up-to-date knowledge may have been invaluable. As more academics get involved in conservation projects around the world, in places they may never physically visit, this communication is invaluable. Not only to make sure models and analyses are accurate but also to ensure that the academics understand the practicalities of the conservation project and the intricacies of the environment they are

conducting research on. This communication and understanding will hopefully ensure that an academic's research contributes meaningfully to practical management.

The analysis of genome-wide data from the pink pigeon produced some interesting results but revealed little of immediate applied conservation value. Nevertheless the genomic resources created for the pink pigeon provide a foundation for future studies of adaptive potential and inbreeding. However, the ability to actually impact these phenomena directly, as opposed to broadly via breeding management or genetic rescue, is not possible. In the future gene editing, such as clustered regularly interspaced short palindromic repeats (CRISPR) technology, may be used to increase adaptive potential or remove deleterious variants. For instance the organisation Revive and Restore (https://reviverestore.org/) are dedicated to understanding how bio-technology, like gene-editing, can be successfully implemented in conservation projects. One project is trying to gain approval to introduce plague resistance into the twice extinct black-footed ferret (*Mustela nigripes*) [331]. If any immune genes could be identified in the pink pigeon, or even another bird, that conferred immunity to avian trichomonosis, this would increase productivity (by increasing juvenile survival) and the adaptive potential of the pink pigeon. Another possible application would be using gene-editing to directly remove deleterious variants thereby reducing both the genetic load of the pink pigeon and the incidence of inbreeding depression. These methods may have the potential to revolutionise conservation genetics in the future but currently are not widely applicable. Therefore, for the pink pigeon, the best management strategy is sustained genetic rescue to try and decrease the genetic load in wild birds and ensure high levels of variation. As important to the survival of the pink pigeon (as genetic rescue) is the continued predator and disease control, as well as the continued management of the captive population to maintain high levels of genetic variation.

Although we are currently living in an age of rapid technological growth combined with a corporate impetus to protect biodiversity [332] conservation remains an underfunded field often unable to take advantage of the latest technologies. Many technological advances come from fields with better funding and a more direct impact on human beings for example medicine and in my opinion to look at recent advances in medicine or human genetics is to understand what technologies may be of use to conservation within the next decade. However, funding isn't the only obstacle to the uptake of new technologies by conservation, often new technologies are highly specialised and practitioners may be unaware of them or how they may be applied to conservation. If awareness is the primary obstacle then this may be rectified by increased communication with academics and industry. The current pandemic (2020 for people reading in the future) has been devastating and yet informative with one of the major changes being an increase in video calling and an acceptance that you do not have to physically work somewhere to work there. As global internet connectivity

improves this will only facilitate communication between experts in the field and academics and this should enable a better understanding of both the challenges faced by practitioners and academics as well as the opportunity to utilise an academic's knowledge about novel technology and how it may be applied to conservation problems. Increased communication may also allow practitioners to make full use of an academics skill set. For example, much of bioinformatics is file manipulation and automation this may not be directly related to conservation however conservation projects produce vast amounts of data, often over the course of decades and there are many data related house keeping tasks - data entry, databases, collating data that could be automated by scripts. These sorts of tasks could make very meaningful contributions to conservation projects but are, by themselves, unlikely to be publishable work and therefore may conflict with the academic impetus of "publish or perish". An increase in collaboration and understanding between academics and practitioners would hopefully allow opportunities for researchers to provide those less publishable skills perhaps alongside other more publishable research projects. Ultimately it is hoped that increased communication and collaborative efforts will help to bridge the "research-implementation gap" enabling researchers to produce tools, analyses and resources that are useful for practitioners and contribute to the conservation of endangered species [114, 26].

6.0.6 Conclusion

Ideally conservation plans should include genetic data to ensure that the most effective management plans can be designed to fulfil the ultimate goal of conservation - enabling self-sustaining populations. This thesis has demonstrated how genome-wide data can reveal surprising insights even about well studied species like the pink pigeon. However, current tools for processing genomic data often rely on complex paramterisation and prior assumptions that reduce the information available from data and result in biased and inaccurate data sets. Methods such as RADiKal that can extract information from raw unprocessed sequencing reads can significantly reduce these biases and improve the accuracy of these data sets. This information can then be used to inform models which help integrate the results of genetic analyses into management. Ultimately better tools should lead to better resources which should enable better (more effective) conservation.

References

[1] A. S. L. Rodrigues, J. D. Pilgrim, J. F. Lamoreux, et al. The value of the IUCN Red List for conservation. *Trends in Ecology & Evolution*, 21(2):71–76, 2006. doi: 10.1016/j.tree.2005.10.010.

[2] M. Albeshr. *Conservation biology and genomics of a flagship endangered species: the Mauritian pink pigeon Nesoenas mayeri*. THESIS.DOCTORAL, 2016.

[3] L. Concannon. *Managing threatened species: understanding the factorslimiting the recovery of the endangered pink pigeon*. THESIS.DOCTORAL, University of Reading, 2014.

[4] J. A. Robinson, J. Räikkönen, L. M. Vucetich, et al. Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Science Advances*, 5(5):eaau0757, 2019. ISSN 2375-2548. doi:10.1126/sciadv.aau0757.

[5] F. C. Ceballos, P. K. Joshi, D. W. Clark, et al. Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews. Genetics*, 19(4):220–234, 2018. ISSN 1471-0056. doi:10.1038/nrg.2017.109.

[6] S. Li, B. Li, C. Cheng, et al. Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. *Genome Biology*, 15(12):557, 2014. ISSN 1465-6906. doi:10.1186/PREACCEPT-2051958771459391.

[7] J. R. Paris, J. R. Stevens, and J. M. Catchen. Lost in parameter space: a road map forstacks. *Methods in ecology and evolution / British Ecological Society*, 8(10):1360–1373, 2017. ISSN 2041210X. doi:10.1111/2041-210X.12775.

[8] J. Ballou, R. Lacy, and J. Pollak. *PMx: software for demographic and genetic analysis and management of pedigreed populations*. Chicago Zoological Society, Brookfield, Illinois, USA., 2010.

[9] K. Swinnerton. *The ecology and conservation of the pink pigeon Columba mayeri in Mauritius*. THESIS.DOCTORAL, University of Kent at Canterbury, 2001.

[10] R. Lacy. Analysis of Founder Representation in Pedigrees: Founder Equivalents and Founder Genome Equivalents. *Zoo Biology*, 8:111–123, 1989.

[11] R. Frankham, B. Jonathan, and B. David. *Introduction To Conservation Genetics*. Cambridge University Press, New York, 2nd edition, 2012.

[12] K. R. Andrews, J. M. Good, M. R. Miller, et al. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews. Genetics*, 17(2):81–92, 2016. doi:10.1038/nrg.2015.28.

[13] H. Jackson, L. Percival-Alwyn, C. Ryan, et al. A blind spot in the IUCN Red List criteria; not all is rosy for the pink pigeon. *in prep.*, 2020.

[14] A. D. Barnosky, N. Matzke, S. Tomiya, et al. Has the Earth's sixth mass extinction already arrived? *Nature*, 471(7336):51–57, 2011. doi:10.1038/nature09678.

[15] C. N. Johnson, A. Balmford, B. W. Brook, et al. Biodiversity losses and conservation responses in the Anthropocene. *Science*, 356(6335):270–275, 2017. ISSN 0036-8075. doi:10.1126/science.aam9317.

[16] R. Frankham. Inbreeding and extinction: island populations. *Conservation Biology*, 12(3):665–675, 2008. ISSN 08888892. doi:10.1111/j.1523-1739.1998.96456.x.

[17] A. Finger, C. J. Kettle, C. N. Kaiser-Bunbury, et al. Back from the brink: potential for genetic rescue in a critically endangered tree. *Molecular Ecology*, 20(18):3773–3784, 2011. doi:10.1111/j.1365-294X.2011.05228.x.

[18] M. Gilpin and M. Soule. Minimum viable populations: processes of species extinction. In M. Soule, editor, *Conservation Biology: the science of Scarcity and Diversity*, pages 19–34. Sinauer, Sunderland, 1986.

[19] S. M. Love Stowell, C. A. Pinzone, and A. P. Martin. Overcoming barriers to active interventions for genetic diversity. *Biodiversity and conservation*, 26(8):1–13, 2017. ISSN 0960-3115. doi:10.1007/s10531-017-1330-z.

[20] A. R. Hoelzel, M. W. Bruford, and R. C. Fleischer. Conservation of adaptive potential and functional diversity. *Conservation genetics (Print)*, 20(1):1–5, 2019. ISSN 1566-0621. doi:10.1007/s10592-019-01151-x.

[21] S. Hendricks, E. C. Anderson, T. Antao, et al. Recent advances in conservation and population genomics data analysis. *Evolutionary applications*, 11(8):1197–1211, 2018. ISSN 17524571. doi:10.1111/eva.12659.

[22] I. Strandén, J. Kantanen, I.-R. M. Russo, et al. Genomic selection strategies for breeding adaptation and production in dairy cattle under climate change. *Heredity*, 2019.

[23] G. J. Doherty, M. Petruzzelli, E. Beddowes, et al. Cancer treatment in the genomic era. *Annual Review of Biochemistry*, 88:247–280, 2019. doi:10.1146/annurev-biochem-062917-011840.

[24] B. A. Garner, B. K. Hand, S. J. Amish, et al. Genomics in Conservation: Case Studies and Bridging the Gap between Data and Application. *Trends in Ecology & Evolution*, 31(2):81–83, 2016. doi:10.1016/j.tree.2015.10.009.

[25] E. A. McLennan, B. R. Wright, K. Belov, et al. Too much of a good thing? Finding the most informative genetic dataset to answer conservation questions. *Molecular ecology resources*, 19(3):659–671, 2019. doi:10.1111/1755-0998.12997.

[26] M. Britt, S. E. Haworth, J. B. Johnson, et al. The importance of non-academic coauthors in bridging the conservation genetics gap. *Biological Conservation*, 218:118–123, 2018. ISSN 00063207. doi:10.1016/j.biocon.2017.12.019.

[27] S. T. Turvey, R. L. Pitman, B. L. Taylor, et al. First human-caused extinction of a cetacean species? *Biology Letters*, 3(5):537–540, 2007. doi:10.1098/rsbl.2007.0292.

[28] A. B. A. Shafer, J. B. W. Wolf, P. C. Alves, et al. Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, 30(2):78–87, 2015. doi:10.1016/j.tree.2014.11.009.

[29] L. Berec, E. Angulo, and F. Courchamp. Multiple Allee effects and population management. *Trends in Ecology & Evolution*, 22(4):185–191, 2007. ISSN 01695347. doi: 10.1016/j.tree.2006.12.002.

[30] M. J. Wittmann, H. Stuis, and D. Metzler. Genetic Allee effects and their interaction with ecological Allee effects. *The Journal of Animal Ecology*, 87(1):11–23, 2018. doi: 10.1111/1365-2656.12598.

[31] A. D. Yoder, J. W. Poelstra, G. P. Tiley, and R. C. Williams. Neutral theory is the foundation of conservation genetics. *Molecular Biology and Evolution*, 35(6):1322–1326, 2018. doi:10.1093/molbev/msy076.

[32] M. Kimura. Evolutionary Rate at the Molecular Level. *Nature*, 217(5129):624–626, 1968. ISSN 0028-0836, 1476-4687. doi:10.1038/217624a0.

[33] D. L. Gilroy, C. van Oosterhout, J. Komdeur, and D. S. Richardson. Toll-like receptor variation in the bottlenecked population of the endangered Seychelles warbler. *Animal Conservation*, 20(3):235–250, 2017. ISSN 13679430. doi:10.1111/acv.12307.

[34] M. E. Roelke, J. S. Martenson, and S. J. O'Brien. The consequences of demographic reduction and genetic depletion in the endangered Florida panther. *Current Biology*, 3(6):340–350, 1993. doi:10.1016/0960-9822(93)90197-V.

[35] P. W. Hedrick, J. A. Robinson, R. O. Peterson, and J. A. Vucetich. Genetics and extinction and the example of Isle Royale wolves. *Animal Conservation*, 2019. ISSN 13679430. doi:10.1111/acv.12479.

[36] B. Charlesworth. Mutational load, inbreeding depression and heterosis in subdivided populations. *BioRxiv*, 2018. doi:10.1101/352146.

[37] P. W. Hedrick and S. T. Kalinowski. Inbreeding Depression in Conservation Biology. *Annual review of ecology and systematics*, 31(1):139–162, 2000. ISSN 0066-4162. doi: 10.1146/annurev.ecolsys.31.1.139.

[38] K. Ralls, J. D. Ballou, and A. Templeton. Estimates of lethal equivalents and the cost of inbreeding in mammals. *Conservation Biology*, 2(2):185–193, 1988. ISSN 0888-8892. doi:10.1111/j.1523-1739.1988.tb00169.x.

[39] J. J. O'Grady, B. W. Brook, D. H. Reed, et al. Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biological Conservation*, 133(1):42–51, 2006. ISSN 00063207. doi:10.1016/j.biocon.2006.05.016.

[40] M. A. Supple and B. Shapiro. Conservation of biodiversity in the genomics era. *Genome Biology*, 19(1):131, 2018. doi:10.1186/s13059-018-1520-3.

[41] O. Frankel. Genetic Conservation: Our Evolutionary Responsibility. *Genetics*, 78, 1974.

[42] J. Greig. Principles of genetic conservation in relation to wildlife management in Southern Africa. *South African Journal of Wildlife Research*, 9(3-4):57–78, 1979.

[43] N. Cortes-Rodriguez, M. G. Campana, L. Berry, et al. Population Genomics and Structure of the Critically Endangered Mariana Crow (Corvus kubaryi). *Genes*, 10(3), 2019. doi:10.3390/genes10030187.

[44] D. W. G. Stanton, P. Frandsen, R. K. Waples, et al. More grist for the mill? Species delimitation in the genomic era and its implications for conservation. *Conservation genetics (Print)*, 20(1):101–113, 2019. ISSN 1566-0621. doi:10.1007/s10592-019-01149-5.

[45] C. Vernesi, M. W. Bruford, G. Bertorelle, et al. Where's the conservation in conservation genetics? *Conservation Biology*, 22(3):802–804, 2008. doi:10.1111/j.1523-1739.2008.00911.x.

[46] W. C. Funk, B. R. Forester, S. J. Converse, et al. Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists. *Conservation Genetics*, 20(1):115–134, 2019. ISSN 1566-0621, 1572-9737. doi:10.1007/s10592-018-1096-1.

[47] A. Sandström, C. Lundmark, K. Andersson, et al. Understanding and bridging the conservation-genetics gap in marine conservation. *Conservation Biology*, 33(3):725–728, 2019. doi:10.1111/cobi.13272.

[48] H. R. Taylor, N. Dussex, and Y. van Heezik. Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Global ecology and conservation*, 10:231–242, 2017. ISSN 23519894. doi:10.1016/j.gecco.2017.04.001.

[49] J. M. Miller, R. M. Malenfant, P. David, et al. Estimating genome-wide heterozygosity: effects of demographic history and marker type. *Heredity*, 112(3):240–247, 2014. doi:10.1038/hdy.2013.99.

[50] J. Bohling, M. Small, J. Von Bargen, et al. Comparing inferences derived from microsatellite and RADseq datasets: a case study involving threatened bull trout. *Conservation genetics (Print)*, 20(2):329–342, 2019. ISSN 1566-0621. doi:10.1007/s10592-018-1134-z.

[51] R. Frankham, J. Ballou, K. Ralls, et al. *Genetic Management of Fragmented Animal and Plant Populations*. Oxford University Press, Oxford, 2017.

[52] H. Kirk and J. R. Freeland. Applications and implications of neutral versus non-neutral markers in molecular ecology. *International Journal of Molecular Sciences*, 12(6):3966–3988, 2011. doi:10.3390/ijms12063966.

[53] J. Sunde, Y. Yıldırım, P. Tibblin, and A. Forsman. Comparing the Performance of Microsatellites and RADseq in Population Genetic Studies: Analysis of Data for Pike (Esox lucius) and a Synthesis of Previous Studies. *Frontiers in genetics*, 11:218, 2020. ISSN 1664-8021. doi:10.3389/fgene.2020.00218.

[54] B. K. Mable. Conservation of adaptive potential and functional diversity: integrating old and new approaches. *Conservation genetics (Print)*, 20(1):1–12, 2018. ISSN 1566-0621. doi:10.1007/s10592-018-1129-9.

[55] Z. L. Fuller, V. J. Mocellin, L. Morris, et al. Population genetics of the coral Acropora millepora : Towards a genomic predictor of bleaching. *BioRxiv*, 2019. doi:10.1101/867754.

[56] S. C. Hammerly, D. A. de la Cerda, H. Bailey, and J. A. Johnson. A pedigree gone bad: increased offspring survival after using DNA-based relatedness to minimize inbreeding in a captive population. *Animal Conservation*, 19(3):296–303, 2016. ISSN 13679430. doi:10.1111/acv.12247.

[57] T. Dejean, A. Valentini, C. Miquel, et al. Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrogLithobates catesbeianus. *Journal of Applied Ecology*, 49(4):953–959, 2012. ISSN 00218901. doi: 10.1111/j.1365-2664.2012.02171.x.

[58] C. S. Baker, D. Steel, Y. Choi, et al. Genetic evidence of illegal trade in protected whales links Japan with the US and South Korea. *Biology Letters*, 6(5):647–650, 2010. doi:10.1098/rsbl.2010.0239.

[59] J. Szymanski, C. Pollack, L. Ragan, et al. Species status assessment for the eastern massasauga rattlesnake (Sistrurus catenatus). \SSA Report, U.S. Fish and Wildlife Service, 2016.

[60] E. Banguera-Hinestroza, H. Cárdenas, M. Ruiz-García, et al. Molecular identification of evolutionarily significant units in the Amazon River dolphin Inia sp. (Cetacea: Iniidae). *The Journal of Heredity*, 93(5):312–322, 2002. doi:10.1093/jhered/93.5.312.

[61] W. Morrison, J. Lohr, P. Duchen, et al. The impact of taxonomic change on conservation: Does it kill, can it save, or is it just irrelevant? *Biological Conservation*, 142(12):3201–3206, 2009. ISSN 00063207. doi:10.1016/j.biocon.2009.07.019.

[62] M. Kardos, G. Luikart, R. Bunch, et al. Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Molecular Ecology*, 24(22):5616–5632, 2015. doi:10.1111/mec.13415.

[63] P. Brandies, E. Peel, C. J. Hogg, and K. Belov. The value of reference genomes in the conservation of threatened species. *Genes*, 10(11), 2019. doi:10.3390/genes10110846.

[64] J. A. Ivy and R. C. Lacy. A comparison of strategies for selecting breeding pairs to maximize genetic diversity retention in managed populations. *The Journal of Heredity*, 103(2):186–196, 2012. doi:10.1093/jhered/esr129.

[65] Mussmann, Douglas, Anthonysamy, et al. prairiechickenandthe problem of conservation reliance in the Anthropocene.

[66] J. M. Miller, M. C. Quinzin, E. H. Scheibe, et al. Genetic Pedigree Analysis of the Pilot Breeding Program for the Rediscovered Galapagos Giant Tortoise from Floreana Island. *The Journal of Heredity*, 109(6):620–630, 2018. doi:10.1093/jhered/esy010.

[67] Z. W. Bateson, S. C. Hammerly, J. A. Johnson, et al. Specific alleles at immune genes, rather than genome-wide heterozygosity, are related to immunity and survival in the critically endangered Attwater's prairie-chicken. *Molecular Ecology*, 25(19):4730–4744, 2016. doi:10.1111/mec.13793.

[68] S. J. Oyler-McCance, K. P. Oh, K. M. Langin, and C. L. Aldridge. A field ornithologist's guide to genomics: Practical considerations for ecology and conservation. *The Auk*, 133(4):626–648, 2016. ISSN 0004-8038. doi:10.1642/AUK-16-49.1.

[69] L. M. Benestan, A.-L. Ferchaud, P. A. Hohenlohe, et al. Conservation genomics of natural and managed populations: building a conceptual and practical framework. *Molecular Ecology*, 25(13):2967–2977, 2016. doi:10.1111/mec.13647.

[70] Z. Yang, Y. Sun, Y. Ma, et al. A comprehensive analysis of the germline and expressed TCR repertoire in White Peking duck. *Scientific Reports*, 7:41426, 2017. doi:10.1038/srep41426.

[71] P. A. Larsen, C. R. Campbell, and A. D. Yoder. Next-generation approaches to advancing eco-immunogenomic research in critically endangered primates. *Molecular ecology resources*, 14(6):1198–1209, 2014. doi:10.1111/1755-0998.12274.

[72] G. Tan, L. Opitz, R. Schlapbach, and H. Rehrauer. Long fragments achieve lower base quality in Illumina paired-end sequencing. *Scientific Reports*, 9(1):2856, 2019. doi:10.1038/s41598-019-39076-7.

[73] S. J. O'Leary, J. B. Puritz, S. C. Willis, et al. These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, 2018. ISSN 09621083. doi:10.1111/mec.14792.

[74] N. Díaz-Arce and N. Rodríguez-Ezpeleta. Selecting RAD-Seq Data Analysis Parameters for Population Genetics: The More the Better? *Frontiers in genetics*, 10:533, 2019. doi:10.3389/fgene.2019.00533.

[75] D. L. J. Vendrami, J. Forcada, and J. I. Hoffman. Experimental validation of in silico predicted RAD locus frequencies using genomic resources and short read data from a model marine mammal. *BMC Genomics*, 20(1):72, 2019. doi:10.1186/s12864-019-5440-8.

[76] R. Holderegger, N. Balkenhol, J. Bolliger, et al. Conservation genetics: Linking science with practice. *Molecular Ecology*, 28(17):3848–3856, 2019. ISSN 1365-294X. doi:10.1111/mec.15202.

[77] E. McCartney-Melstad, M. Gidiş, and H. B. Shaffer. Population genomic data reveal extreme geographic subdivision and novel conservation actions for the declining foothill yellow-legged frog. *Heredity*, 121(2):112–125, 2018. ISSN 0018-067X, 1365-2540. doi:10.1038/s41437-018-0097-7.

[78] V. R. Titus, R. C. Bell, C. G. Becker, and K. R. Zamudio. Connectivity and gene flow among Eastern Tiger Salamander (Ambystoma tigrinum) populations in highly modified anthropogenic landscapes. *Conservation genetics (Print)*, 15(6):1447–1462, 2014. ISSN 1566-0621. doi:10.1007/s10592-014-0629-5.

[79] E. McCartney-Melstad, J. K. Vu, and H. B. Shaffer. Genomic Data from an Endangered Amphibian Reveal Unforeseen Consequences of Fragmentation by Roads. *BioRxiv*, 2018. doi:10.1101/306340.

[80] A. J. Lind, P. Q. Spinks, G. M. Fellers, and H. B. Shaffer. Rangewide phylogeography and landscape genetics of the Western U.S. endemic frog Rana boylii (Ranidae): implications for the conservation of frogs and rivers. *Conservation Genetics*, 12(1):269–284, 2011. ISSN 1566-0621, 1572-9737. doi:10.1007/s10592-010-0138-0.

[81] C. Pacioni, C. Rafferty, K. Morley, et al. Augmenting the conservation value of rehabilitated wildlife by integrating genetics and population modeling in the post-rehabilitation decision process. *Current zoology*, 64(5):593–601, 2018. doi:10.1093/cz/zox065.

[82] A. R. Weeks, J. Stoklosa, and A. A. Hoffmann. Conservation of genetic uniqueness of populations may increase extinction likelihood of endangered species: the case of Australian mammals. *Frontiers in zoology*, 13(1):31, 2016. doi:10.1186/s12983-016-0163-z.

[83] W. E. Johnson, D. P. Onorato, M. E. Roelke, et al. Genetic restoration of the Florida panther. *Science*, 329(5999):1641–1645, 2010. doi:10.1126/science.1192891.

[84] P. W. Hedrick and R. Fredrickson. Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. *Conservation genetics (Print)*, 11(2):615–626, 2010. ISSN 1566-0621. doi:10.1007/s10592-009-9999-5.

[85] C. Vilà, A.-K. Sundqvist, Flagstad, et al. Rescue of a severely bottlenecked wolf (Canis lupus) population by a single immigrant. *Proceedings. Biological Sciences / the Royal Society*, 270(1510):91–97, 2003. doi:10.1098/rspb.2002.2184.

[86] M. \AAkesson, O. Liberg, H. Sand, et al. Genetic rescue in a severely inbred wolf population. *Molecular Ecology*, 25(19):4745–4756, 2016. doi:10.1111/mec.13797.

[87] R. J. Fredrickson, P. Siminski, M. Woolf, and P. W. Hedrick. Genetic rescue and inbreeding depression in Mexican wolves. *Proceedings. Biological Sciences / the Royal Society*, 274(1623):2365–2371, 2007. doi:10.1098/rspb.2007.0785.

[88] K. B. Beauclerc, B. Johnson, and B. N. White. Genetic rescue of an inbred captive population of the critically endangered Puerto Rican crested toad (Peltophryne lemur) by mixing lineages. *Conservation genetics (Print)*, 11(1):21–32, 2010. ISSN 1566-0621. doi:10.1007/s10592-008-9782-z.

[89] K. A. Harrisson, A. Pavlova, A. Gonçalves da Silva, et al. Scope for genetic rescue of an endangered subspecies though re-establishing natural gene flow with another subspecies. *Molecular Ecology*, 25(6):1242–1258, 2016. doi:10.1111/mec.13547.

[90] Y. Willi, M. van Kleunen, S. Dietrich, and M. Fischer. Genetic rescue persists beyond first-generation outbreeding in small populations of a rare plant. *Proceedings. Biological Sciences / the Royal Society*, 274(1623):2357–2364, 2007. doi:10.1098/rspb.2007.0768.

[91] R. Frankham. Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology*, 24(11):2610–2618, 2015. doi:10.1111/mec.13139.

[92] R. Frankham. Genetic rescue benefits persist to at least the F3 generation, based on a meta-analysis. *Biological Conservation*, 195:33–36, 2016. ISSN 00063207. doi:10.1016/j.biocon.2015.12.038.

[93] K. G. Mansfield and E. D. Land. Cryptorchidism in Florida panthers: prevalence, features, and influence of genetic restoration. *Journal of Wildlife Diseases*, 38(4):693–698, 2002. doi:10.7589/0090-3558-38.4.693.

[94] D. M. Waller. Genetic rescue: a safe or risky bet? *Molecular Ecology*, 24(11):2595–2597, 2015. doi:10.1111/mec.13220.

[95] A. R. Whiteley, S. W. Fitzpatrick, W. C. Funk, and D. A. Tallmon. Genetic rescue to the rescue. *Trends in Ecology & Evolution*, 30(1):42–49, 2015. doi:10.1016/j.tree.2014.10.009.

[96] R. Frankham, J. D. Ballou, M. D. B. Eldridge, et al. Predicting the probability of outbreeding depression. *Conservation Biology*, 25(3):465–475, 2011. doi:10.1111/j.1523-1739.2011.01662.x.

[97] S. Edmands. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology*, 16(3):463–475, 2007. doi:10.1111/j.1365-294X.2006.03148.x.

[98] T. L. Goldberg, E. C. Grant, K. R. Inendino, et al. Increased Infectious Disease Susceptibility Resulting from Outbreeding Depression. *Conservation Biology*, 19(2):455–462, 2005. ISSN 0888-8892. doi:10.1111/j.1523-1739.2005.00091.x.

[99] E. McClelland and K. Naish. What is the fitness outcome of crossing unrelated fish populations? A meta-analysis and an evaluation of future research directions. *Conservation genetics (Print)*, 8(2):397–416, 2007. ISSN 1566-0621. doi:10.1007/s10592-006-9178-x.

[100] S. W. Fitzpatrick, G. S. Bradburd, C. T. Kremer, et al. Genomic and fitness consequences of genetic rescue in wild populations. *Current Biology*, 2020. ISSN 09609822. doi:10.1016/j.cub.2019.11.062.

[101] C. Pacioni, S. Trocini, A. F. Wayne, et al. Integrating population genetics in an adaptive management framework to inform management strategies. *Biodiversity and conservation*, 2019. ISSN 0960-3115. doi:10.1007/s10531-019-01920-7.

[102] S. Hamilton and H. Moller. Can PVA models using computer packages offer useful conservation advice? Sooty shearwaters Puffinus griseus in New Zealand as a case study. *Biological Conservation*, 73(2):107–117, 1995. ISSN 00063207. doi:10.1016/0006-3207(95)90033-0.

[103] C. Pacioni, M. R. Williams, R. C. Lacy, et al. Predators and genetic fitness: key threatening factors for the conservation of a bettong species. *Pacific Conservation Biology*, 23(2):200, 2017. ISSN 1038-2097. doi:10.1071/PC17002.

[104] D. Licht. Restoration of Bison (Bison bison) to Agate Fossil Beds National Monument. Technical report, National Park Service, Fort Collins, CO, 2014.

[105] D. S. Licht, R. A. Moen, and M. Romanski. Modeling viability of a potential canada lynx reintroduction to isle royale national park. *Natural Areas Journal*, 37(2):170–177, 2017. ISSN 0885-8608. doi:10.3375/043.037.0206.

[106] R. Lacy and K. Traylor-Holzer. *Vortex 10 User's Manual*. IUCN \SSC Conservation Breeding Specialist Group, and Chicago Zoological Society, Apple Valley, Minnesota, USA, 2017.

[107] J. D. Ballou, K. Traylor-Holzer, A. Turner, et al. Simulation model for contraceptive management of the Assateague Island feral horse population using individual-based data. *Wildlife Research*, 35(6):502, 2008. ISSN 1035-3712. doi:10.1071/WR07124.

[108] B. Dayananda, S. Gray, D. Pike, and J. K. Webb. Communal nesting under climate change: fitness consequences of higher incubation temperatures for a nocturnal lizard. *Global change biology*, 22(7):2405–2414, 2016. doi:10.1111/gcb.13231.

[109] R. Heinsohn, R. C. Lacy, D. B. Lindenmayer, et al. Unsustainable harvest of dugongs in Torres Strait and Cape York (Australia) waters: two case studies using population viability analysis. *Animal Conservation*, 7(4):417–425, 2004. ISSN 1367-9430. doi:10.1017/S1367943004001593.

[110] D. Lidenmayer, T. Clark, R. Lacy, and V. Thomas. Population Viability Analysis as a Tool in Wildlife Conservation Policy: With Reference to Australia. *Environmental Management*, 17(6):745–758, 1993.

[111] K. Ralls, J. D. Ballou, M. R. Dudash, et al. Call for a paradigm shift in the genetic management of fragmented populations. *Conservation letters*, 11(2), 2017. ISSN 1755263X. doi:10.1111/conl.12412.

[112] IUCN/\SSC. Guidelines for Reintroductions and Other Conservation Translocation. Technical report, IUCN Species Survival Commission„ Gland, Switzerland, 2013.

[113] B. Collen, N. K. Dulvy, K. J. Gaston, et al. Clarifying misconceptions of extinction risk assessment with the IUCN Red List. *Biology Letters*, 12(4), 2016. doi:10.1098/rsbl.2015.0843.

[114] C. J. Hogg, C. E. Grueber, D. Pemberton, et al. "devil tools & tech": A synergy of conservation research and management practice. *Conservation letters*, 10(1):133–138, 2015. ISSN 1755263X. doi:10.1111/conl.12221.

[115] A. Cheke and J. Hume, editors. *Lost Land of the Dodo: The Ecological History of Mauritius, Reunion and Rodrigues.* A&C Black, London, UK, 2008.

[116] N. Bunbury. *Parasitic disease in the endangered Mauritian pink pigeon.* THESIS.DOCTORAL, University of East Anglia, 2006.

[117] N. Bunbury, E. Barton, C. G. Jones, et al. Avian blood parasites in an endangered columbid: Leucocytozoon marchouxi in the Mauritian Pink Pigeon Columba mayeri. *Parasitology*, 134(Pt 6):797–804, 2007. doi:10.1017/S0031182006002149.

[118] FOWLER, GANESHAN, MAUREMOOTOO, and MUNGROO. Biological Control of Weeds in Mauritius: Past Successes Revisited and Present Challenges. pages 43–50. Montana State University, Bozeman, Montana, USA, 2000.

[119] C. Jones. The larger land-birds of Mauritius. In A. Diamond, editor, *Studies of Mascarene Island Birds*, pages 208–300. Cambridge University Press, Cambridge, 1987.

[120] K. J. Swinnerton, J. J. Groombridge, C. G. Jones, et al. Inbreeding depression and founder diversity among captive and free-living populations of the endangered pink pigeon Columba mayeri. *Animal Conservation*, 7(4):353–364, 2004. ISSN 1367-9430. doi:10.1017/S1367943004001556.

[121] K. J. Swinnerton, M. A. Peirce, A. Greenwood, et al. Prevalence of Leucocytozoon marchouxi in the endangered Pink Pigeon Columba mayeri. *Ibis*, 147(4):725–737, 2005. ISSN 00191019. doi:10.1111/j.1474-919X.2005.00454.x.

[122] K. Edmunds, N. Bunbury, S. Sawmy, et al. Restoring avian island endemics: use of supplementary food by the endangered Pink Pigeon (Columba mayeri). *Emu*, 108(1):74, 2008. ISSN 0158-4197. doi:10.1071/MU06056.

[123] C. Jones, N. Bunbury, N. Zuel, et al. Pink pigeon Nesoenas mayeri . In R. Safford and F. Hawkins, editors, *Birds of Africa Volume 8: the Malagasy region.* Bloomsbury Publishing„ London, UK, 2013.

[124] D. Charlesworth and J. H. Willis. The genetics of inbreeding depression. *Nature Reviews. Genetics*, 10(11):783–796, 2009. doi:10.1038/nrg2664.

[125] N. Bunbury, C. G. Jones, A. G. Greenwood, and D. J. Bell. Trichomonas gallinae in Mauritian columbids: implications for an endangered endemic. *Journal of Wildlife Diseases*, 43(3):399–407, 2007. doi:10.7589/0090-3558-43.3.399.

[126] D. Gaspar da Silva, E. Barton, N. Bunbury, et al. Molecular identity and heterogeneity of Trichomonad parasites in a closed avian population. *Infection, Genetics and Evolution*, 7(4):433–440, 2007. doi:10.1016/j.meegid.2007.01.002.

[127] N. Bunbury, C. Jones, A. Greenwood, and D. Bell. Epidemiology and conservation implications of Trichomonas gallinae infection in the endangered Mauritian pink pigeon. *Biological Conservation*, 141(1):153–161, 2008. ISSN 00063207. doi:10.1016/j.biocon.2007.09.008.

[128] N. Bunbury, D. Bell, C. Jones, et al. Comparison of the InPouch TF culture system and wet-mount microscopy for diagnosis of Trichomonas gallinae infections in the pink pigeon Columba mayeri. *Journal of Clinical Microbiology*, 43(2):1005–1006, 2005. doi:10.1128/JCM.43.2.1005-1006.2005.

[129] U. Seal and M. Bruford. Columba (Nesoenas) mayeri pink pigeon conservation viability assessment workshop. Technical report, Captive breeding specialist group, Jersey Wildlife Preservation Trust, 1991.

[130] L. D. Shepherd and D. M. Lambert. Nuclear microsatellite DNA markers for New Zealand kiwi (Apteryx spp.). *Molecular ecology notes*, 6(1):227–229, 2006. ISSN 1471-8278. doi:10.1111/j.1471-8286.2005.01201.x.

[131] C.-H. Chan, K. Ballantyne, D. Lambert, and G. Chambers. Characterization of variable microsatellite loci in Forbes' parakeet (Cyanoramphus forbesi) and their use in other parrots. *Conservation Genetics*, 6:651–654, 2005. doi:10.1007/s10592-005-9021-9.

[132] P. E. A. Hoeck, J. L. Bollmer, P. G. Parker, and L. F. Keller. Differentiation with drift: a spatio-temporal genetic analysis of Galapagos mockingbird populations (Mimus spp.). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 365(1543):1127–1138, 2010. doi:10.1098/rstb.2009.0311.

[133] E. P. Akst, P. D. Boersma, and R. C. Fleischer. A comparison of genetic diversity between the Galapagos Penguin and the Magellanic Penguin. *Conservation genetics (Print)*, 3:375 – 383, 2002.

[134] Y.-J. Ji, Y.-D. Liu, C.-Q. Ding, and D.-X. Zhang. Eight polymorphic microsatellite loci for the critically endangered crested ibis, Nipponia nippon (Ciconiiformes: Threskiornithidae). *Molecular ecology notes*, 4(4):615–617, 2004. ISSN 1471-8278. doi:10.1111/j.1471-8286.2004.00754.x.

[135] S. Wirtz, C. Böhm, J. Fritz, et al. Isolation of Microsatellite Loci by Next-Generation Sequencing of the Critically Endangered Northern Bald Ibis, Geronticus eremita. *The Journal of Heredity*, 107(4):363–366, 2016. doi:10.1093/jhered/esw013.

[136] B. R. Wright, K. A. Farquharson, E. A. McLennan, et al. A demonstration of conservation genomics for threatened species management. *Molecular ecology resources*, 2020. doi:10.1111/1755-0998.13211.

[137] K. Grigorev, S. Kliver, P. Dobrynin, et al. Innovative assembly strategy contributes to understanding the evolution and conservation genetics of the endangered Solenodon paradoxus from the island of Hispaniola. *GigaScience*, 7(6), 2018. ISSN 2047-217X. doi:10.1093/gigascience/giy025.

[138] M. Todesco, G. L. Owens, N. Bercovich, et al. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, 584(7822):602–607, 2020. ISSN 0028-0836. doi: 10.1038/s41586-020-2467-6.

[139] A. P. Fuentes-Pardo and D. E. Ruzzante. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, 26(20):5369–5406, 2017. doi:10.1111/mec.14264.

[140] R. Ekblom and J. B. W. Wolf. A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary applications*, 7(9):1026–1042, 2014. doi:10.1111/eva.12178.

[141] K.-P. Koepfli, B. Paten, G. K. C. o. Scientists, and S. J. O'Brien. The Genome 10K Project: a way forward. *Annual review of animal biosciences*, 3:57–111, 2015. doi:10.1146/annurev-animal-090414-014900.

[142] M. C. Fitzpatrick and S. R. Keller. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, 18(1):1–16, 2015. doi:10.1111/ele.12376.

[143] R. A. Bay, R. J. Harrigan, V. L. Underwood, et al. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, 359(6371):83–86, 2018. doi:10.1126/science.aan4380.

[144] N. Wu, S. Zhang, X. Li, et al. Fall webworm genomes yield insights into rapid adaptation of invasive species. *Nature Ecology & Evolution*, 3(1):105–115, 2019. doi: 10.1038/s41559-018-0746-5.

[145] J. A. Robinson, D. Ortega-Del Vecchyo, Z. Fan, et al. Genomic flatlining in the endangered island fox. *Current Biology*, 26(9):1183–1189, 2016. doi:10.1016/j.cub.2016.02.062.

[146] S. J. O'Brien, W. E. Johnson, C. A. Driscoll, et al. Conservation genetics of the cheetah: lessons learned and new opportunities. *The Journal of Heredity*, 108(6):671–677, 2017. doi:10.1093/jhered/esx047.

[147] P. A. Morin, F. I. Archer, C. D. Avila, et al. Reference genome and demographic history of the most endangered marine mammal, the vaquita. *BioRxiv*, 2020. doi: 10.1101/2020.05.27.098582.

[148] H. Weigand and F. Leese. Detecting signatures of positive selection in non-model species using genomic data. *Zoological journal of the Linnean Society*, 184(2):528–583, 2018. ISSN 0024-4082. doi:10.1093/zoolinnean/zly007.

[149] A. C. Beichman, K.-P. Koepfli, G. Li, et al. Aquatic Adaptation and Depleted Diversity: A Deep Dive into the Genomes of the Sea Otter and Giant Otter. *Molecular Biology and Evolution*, 36(12):2631–2655, 2019. doi:10.1093/molbev/msz101.

[150] J. R. Willoughby, A. M. Harder, J. A. Tennessen, et al. Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. *Molecular Ecology*, 27(20):4041–4051, 2018. doi:10.1111/mec.14726.

[151] G. G. R. Murray, A. E. R. Soares, B. J. Novak, et al. Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science*, 358(6365):951–954, 2017. doi:10.1126/science.aao0960.

[152] M. Kardos, M. \AAkesson, T. Fountain, et al. Genomic consequences of intensive inbreeding in an isolated wolf population. *Nature Ecology & Evolution*, 2(1):124–131, 2017. ISSN 2397-334X. doi:10.1038/s41559-017-0375-4.

[153] Y. Xue, J. Prado-Martinez, P. H. Sudmant, et al. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science*, 348(6231):242–245, 2015. doi:10.1126/science.aaa3952.

[154] M. Kirin, R. McQuillan, C. S. Franklin, et al. Genomic runs of homozygosity record population history and consanguinity. *Plos One*, 5(11):e13996, 2010. doi:10.1371/journal.pone.0013996.

[155] P. W. Hedrick. Conservation genetics and the persistence and translocation of small populations: bighorn sheep populations as examples. *Animal Conservation*, 17(2):106–114, 2014. ISSN 13679430. doi:10.1111/acv.12064.

[156] K. Ralls, P. Sunnucks, R. C. Lacy, and R. Frankham. Genetic rescue: A critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful genetic variation. *Biological Conservation*, 251:108784, 2020. ISSN 00063207. doi:10.1016/j.biocon.2020.108784.

[157] C. C. Kyriazis, R. K. Wayne, and K. E. Lohmueller. High genetic diversity can contribute to extinction in small populations. *BioRxiv*, 2019. doi:10.1101/678524.

[158] J. A. Kronenberger, W. C. Funk, J. W. Smith, et al. Testing the demographic effects of divergent immigrants on small populations of Trinidadian guppies. *Animal Conservation*, 20(1):3–11, 2017. ISSN 13679430. doi:10.1111/acv.12286.

[159] C. E. Grueber. Making the best of a bad situation: genetic rescue in the absence of an ideal source population. *Animal Conservation*, 20(1):14–15, 2017. ISSN 13679430. doi:10.1111/acv.12337.

[160] S. Ballouz, A. Dobin, and J. A. Gillis. Is it time to change the reference genome? *Genome Biology*, 20(1):159, 2019. doi:10.1186/s13059-019-1774-4.

[161] J. L. Bouzat. Conservation genetics of population bottlenecks: the role of chance, selection, and history. *Conservation genetics (Print)*, 11(2):463–478, 2010. ISSN 1566-0621. doi:10.1007/s10592-010-0049-0.

[162] A. M. Phillippy, M. C. Schatz, and M. Pop. Genome assembly forensics: finding the elusive mis-assembly. *Genome Biology*, 9(3):R55, 2008. doi:10.1186/gb-2008-9-3-r55.

[163] B. Clavijo, G. Garcia Accinelli, J. Wright, et al. W2RAP: a pipeline for high quality, robust assemblies of large complex genomes from short read data. *BioRxiv*, 2017. doi:10.1101/110999.

[164] L. Yanes, G. Garcia Accinelli, J. Wright, et al. A Sequence Distance Graph framework for genome assembly and analysis. *F1000Research*, 8:1490, 2019. doi:10.12688/f1000research.20233.1.

[165] X. Zhang, R. Wu, Y. Wang, et al. Unzipping haplotypes in diploid and polyploid genomes. *Computational and structural biotechnology journal*, 18:66–72, 2020. doi:10.1016/j.csbj.2019.11.011.

[166] S. C. Clark, R. Egan, P. I. Frazier, and Z. Wang. ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. *Bioinformatics*, 29(4):435–443, 2013. doi:10.1093/bioinformatics/bts723.

[167] S. L. Salzberg, A. M. Phillippy, A. Zimin, et al. GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Research*, 22(3):557–567, 2012. doi:10.1101/gr.131383.111.

[168] A. M. Phillippy. New advances in sequence assembly. *Genome Research*, 27(5):xi–xiii, 2017. doi:10.1101/gr.223057.117.

[169] T. Günther and C. Nettelblad. The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genetics*, 15(7):e1008302, 2019. doi:10.1371/journal.pgen.1008302.

[170] R. E. O'Connor, M. Farré, S. Joseph, et al. Chromosome-level assembly reveals extensive rearrangement in saker falcon and budgerigar, but not ostrich, genomes. *Genome Biology*, 19(1):171, 2018. ISSN 1474-760X. doi:10.1186/s13059-018-1550-x.

[171] J. Kim, D. M. Larkin, Q. Cai, et al. Reference-assisted chromosome assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 110(5):1785–1790, 2013. doi:10.1073/pnas.1220349110.

[172] M. Kolmogorov, B. Raney, B. Paten, and S. Pham. Ragout-a reference-assisted assembly tool for bacterial genomes. *Bioinformatics*, 30(12):i302–9, 2014. doi:10.1093/bioinformatics/btu280.

[173] B. Clavijo. *Genome assembly and quality control for non-model organisms*. 2020. Type: THESIS.DEGREE.

[174] M. Kolmogorov, J. Armstrong, B. J. Raney, et al. Chromosome assembly of large and complex genomes using multiple references. *Genome Research*, 28(11):1720–1732, 2018. doi:10.1101/gr.236273.118.

[175] G. Zhang, C. Li, Q. Li, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, 346(6215):1311–1320, 2014. doi:10.1126/science.1251385.

[176] G. Zhang, E. D. Jarvis, and M. T. P. Gilbert. A flockof genomes. *Science*, pages 1308–1309, 2014.

[177] R. H. S. Kraus. *Avian Genomics in Ecology and Evolution: From the Lab into the Wild*. 2019. ISBN 978-3-030-16477-5. OCLC: 1132058386.

[178] N. I. Weisenfeld, S. Yin, T. Sharpe, et al. Comprehensive variation discovery in single human genomes. *Nature Genetics*, 46(12):1350–1355, 2014. doi:10.1038/ng.3121.

[179] R. Luo, B. Liu, Y. Xie, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 1(1):18, 2012. doi:10.1186/2047-217X-1-18.

[180] M. Alonge, S. Soyk, S. Ramakrishnan, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology*, 20(1):224, 2019. ISSN 1474-760X. doi:10.1186/s13059-019-1829-6.

[181] N. I. Weisenfeld, V. Kumar, P. Shah, et al. Direct determination of diploid genome sequences. *Genome Research*, 27(5):757–767, 2017. doi:10.1101/gr.214874.116.

[182] D. Mapleson, G. Garcia Accinelli, G. Kettleborough, et al. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*, 33(4):574–576, 2017. doi:10.1093/bioinformatics/btw663.

[183] R. M. Waterhouse, M. Seppey, F. A. Simão, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, 35(3):543–548, 2018. doi:10.1093/molbev/msx319.

[184] F. A. Simão, R. M. Waterhouse, P. Ioannidis, et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19):3210–3212, 2015. doi:10.1093/bioinformatics/btv351.

[185] J. T. Simpson, K. Wong, S. D. Jackman, et al. ABy\SS: a parallel assembler for short read sequence data. *Genome Research*, 19(6):1117–1123, 2009. doi:10.1101/gr.089532.108.

[186] B. Langmead and S. L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359, 2012. doi:10.1038/nmeth.1923.

[187] A. McKenna, M. Hanna, E. Banks, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297–1303, 2010. doi:10.1101/gr.107524.110.

[188] P. Danecek, A. Auton, G. Abecasis, et al. The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158, 2011. doi:10.1093/bioinformatics/btr330.

[189] A. Auton, K. Bryc, A. R. Boyko, et al. Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Research*, 19(5):795–803, 2009. doi:10.1101/gr.088898.108.

[190] A. R. Bertrand, N. K. Kadri, L. Flori, et al. RZooRoH: An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. *Methods in ecology and evolution / British Ecological Society*, 10(6):860–866, 2019. ISSN 2041-210X. doi:10.1111/2041-210X.13167.

[191] S. Purcell, B. Neale, K. Todd-Brown, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3):559–575, 2007. ISSN 0002-9297. doi:10.1086/519795.

[192] H. Li and R. Durbin. Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357):493–496, 2011. ISSN 1476-4687. doi:10.1038/nature10231.

[193] H. Li, B. Handsaker, A. Wysoker, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, 2009. doi:10.1093/bioinformatics/btp352.

[194] K. Nadachowska-Brzyska, C. Li, L. Smeds, et al. Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. *Current Biology*, 25(10):1375–1380, 2015. doi:10.1016/j.cub.2015.03.047.

[195] K. Nadachowska-Brzyska, R. Burri, L. Smeds, and H. Ellegren. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Molecular Ecology*, 25(5):1058–1072, 2016. doi:10.1111/mec.13540.

[196] G. Yue and L. Wang. Current status of genome sequencing and its applications in aquaculture. *Aquaculture*, 468:337–347, 2017. ISSN 00448486. doi:10.1016/j.aquaculture.2016.10.036.

[197] M. M. Miller and R. L. Taylor. Brief review of the chicken Major Histocompatibility Complex: the genes, their distribution on chromosome 16, and their contributions to disease resistance. *Poultry Science*, 95(2):375–392, 2016. doi:10.3382/ps/pev379.

[198] S. Zhan, W. Zhang, K. Niitepõld, et al. The genetics of monarch butterfly migration and warning colouration. *Nature*, 514(7522):317–321, 2014. doi:10.1038/nature13812.

[199] S. Kurtz, A. Phillippy, A. L. Delcher, et al. Versatile and open software for comparing large genomes. *Genome Biology*, 5(2):R12, 2004. doi:10.1186/gb-2004-5-2-r12.

[200] M. Nattestad, R. Aboukhalil, C.-S. Chin, and M. Schatz. Ribbon: Intuitive visualization for complex genomic variation. *Bioinformatics*, 2020.

[201] J. Sandoval-Castillo, C. R. M. Attard, S. Marri, et al. swinger: a user-friendly computer program to establish captive breeding groups that minimize relatedness without pedigree information. *Molecular ecology resources*, 17(2):278–287, 2017. doi:10.1111/1755-0998.12609.

[202] P. A. Ayala-Burbano, P. M. Galetti Junior, D. Wormell, et al. Studbook and molecular analyses for the endangered black-lion-tamarin; an integrative approach for assessing genetic diversity and driving management in captivity. *Scientific Reports*, 10(1):6781, 2020. doi:10.1038/s41598-020-63542-2.

[203] P. W. Hedrick, U. Hellsten, and D. Grattapaglia. Examining the cause of high inbreeding depression: analysis of whole-genome sequence data in 28 selfed progeny of Eucalyptus grandis. *The New Phytologist*, 209(2):600–611, 2016. doi:10.1111/nph.13639.

[204] E. A. McLennan, R. M. Gooley, P. Wise, et al. Pedigree reconstruction using molecular data reveals an early warning sign of gene diversity loss in an island population of Tasmanian devils (Sarcophilus harrisii). *Conservation genetics (Print)*, 19(2):1–12, 2017. ISSN 1566-0621. doi:10.1007/s10592-017-1017-8.

[205] T. Druet and M. Gautier. A model-based approach to characterize individual inbreeding at both global and local genomic scales. *Molecular Ecology*, 26(20):5820–5841, 2017. doi:10.1111/mec.14324.

[206] T. Callicrate, R. Dikow, J. W. Thomas, et al. Genomic resources for the endangered Hawaiian honeycreepers. *BMC Genomics*, 15:1098, 2014. doi:10.1186/1471-2164-15-1098.

[207] D. Paul, W. M. White, and J. Blichert-Toft. Geochemistry of Mauritius and the origin of rejuvenescent volcanism on oceanic island volcanoes. *Geochemistry, Geophysics, Geosystems*, 6(6), 2005. ISSN 15252027. doi:10.1029/2004GC000883.

[208] J. S. Strijk, R. D. Noyes, D. Strasberg, et al. In and out of Madagascar: dispersal to peripheral islands, insular speciation and diversification of Indian Ocean daisy trees (Psiadia, Asteraceae). *Plos One*, 7(8):e42932, 2012. doi:10.1371/journal.pone.0042932.

[209] Y. Zhang, Y. Zhong, Y. Hong, et al. RAD-Seq data advance captive-based conservation of wild bactrian camels (Camelus ferus). *Conservation genetics (Print)*, pages 1–8, 2019. ISSN 1566-0621. doi:10.1007/s10592-019-01173-5.

[210] L. Laikre, S. Hoban, M. W. Bruford, et al. Post-2020 goals overlook genetic diversity. *Science*, 367(6482):1083–1085, 2020. ISSN 0036-8075. doi:10.1126/science.abb2748.

[211] A. Barbanti, H. Torrado, E. Macpherson, et al. Helping decision making for reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species. *Molecular ecology resources*, 2020. doi:10.1111/1755-0998.13144.

[212] A. Mastretta-Yanes, N. Arrigo, N. Alvarez, et al. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular ecology resources*, 15(1):28–41, 2015. doi: 10.1111/1755-0998.12291.

[213] M. R. Miller, J. P. Dunham, A. Amores, et al. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, 17(2):240–248, 2007. ISSN 1088-9051. doi:10.1101/gr.5681207.

[214] N. A. Baird, P. D. Etter, T. S. Atwood, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *Plos One*, 3(10):e3376, 2008. ISSN 1932-6203. doi:10.1371/journal.pone.0003376.

[215] B. K. Peterson, J. N. Weber, E. H. Kay, et al. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One*, 7(5):e37135, 2012. doi:10.1371/journal.pone.0037135.

[216] S. Wang, E. Meyer, J. K. McKay, and M. V. Matz. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods*, 9(8):808–810, 2012. doi:10.1038/nmeth.2023.

[217] M. R. Miller, T. S. Atwood, B. F. Eames, et al. RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biology*, 8(6):R105, 2007. doi:10.1186/gb-2007-8-6-r105.

[218] J. W. Davey, J. L. Davey, M. L. Blaxter, and M. W. Blaxter. RADSeq: next-generation population genetics. *Briefings in functional genomics*, 9(5-6):416–423, 2010. doi: 10.1093/bfgp/elq031.

[219] J. Buckley, E. B. Holub, M. A. Koch, et al. Restriction associated DNA-genotyping at multiple spatial scales in Arabidopsis lyrata reveals signatures of pathogen-mediated selection. *BMC Genomics*, 19(1):496, 2018. doi:10.1186/s12864-018-4806-7.

[220] B. P. Wiedmann and G. A. Sargeant. Ecotypic variation in recruitment of reintroduced bighorn sheep: Implications for translocation. *The Journal of wildlife management*, 78(3):394–401, 2014. ISSN 0022541X. doi:10.1002/jwmg.669.

[221] S. Yu, W. Chu, L. Zhang, et al. Identification of Laying-Related SNP Markers in Geese Using RAD Sequencing. *Plos One*, 10(7):e0131572, 2015. doi:10.1371/journal.pone.0131572.

[222] R. Ogden, K. Gharbi, N. Mugue, et al. Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. *Molecular Ecology*, 22(11):3112–3123, 2013. doi:10.1111/mec.12234.

[223] J. D. Lozier. Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. *Molecular Ecology*, 23(4):788–801, 2014. doi:10.1111/mec.12636.

[224] T. Suchan, C. Pitteloud, N. S. Gerasimova, et al. Hybridization capture using RAD probes (hyrad), a new tool for performing genomic analyses on collection specimens. *Plos One*, 11(3):e0151651, 2016. doi:10.1371/journal.pone.0151651.

[225] L. A. Lecaudey, U. K. Schliewen, A. G. Osinov, et al. Inferring phylogenetic structure, hybridization and divergence times within Salmoninae (Teleostei: Salmonidae) using RAD-sequencing. *Molecular Phylogenetics and Evolution*, 124:82–99, 2018. doi:10.1016/j.ympev.2018.02.022.

[226] T. A. Maigret. Snake scale clips as a source of high quality DNA suitable for RAD sequencing. *Conservation genetics resources*, pages 1–3, 2018. ISSN 1877-7252. doi:10.1007/s12686-018-1019-y.

[227] K. M. Lee, S. M. Kivelä, V. Ivanov, et al. Information Dropout Patterns in RAD Phylogenomics and a Comparison with Multilocus Sanger Data in a Species-rich Moth Genus. *Systematic Biology*, 67(6):925–939, 2018. doi:10.1093/sysbio/syy029.

[228] M. M.-Y. Tin, E. P. Economo, and A. S. Mikheyev. Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. *Plos One*, 9(5):e96793, 2014. doi:10.1371/journal.pone.0096793.

[229] E. B. Linck, Z. R. Hanna, A. Sellas, and J. P. Dumbacher. Evaluating hybridization capture with RAD probes as a tool for museum genomics with historical bird specimens. *Ecology and evolution*, 7(13):4755–4767, 2017. doi:10.1002/ece3.3065.

[230] Y. Zhong, A. Yang, S. Liu, et al. RAD-Seq Data Point to a Distinct Split in Liriodendron (Magnoliaceae) and Obvious East–West Genetic Divergence in L. chinense. *Forests*, 10(1):13, 2018. ISSN 1999-4907. doi:10.3390/f10010013.

[231] H. Yamashita, H. Katai, L. Kawaguchi, et al. Analyses of single nucleotide polymorphisms identified by ddRAD-seq reveal genetic structure of tea germplasm and Japanese landraces for tea breeding. *Plos One*, 14(8):e0220981, 2019. doi:10.1371/journal.pone.0220981.

[232] L. C. Orantes, C. Monroy, P. L. Dorn, et al. Uncovering vector, parasite, blood meal and microbiome patterns from mixed-DNA specimens of the Chagas disease vector Triatoma dimidiata. *PLoS Neglected Tropical Diseases*, 12(10):e0006730, 2018. doi:10.1371/journal.pntd.0006730.

[233] L. M. Komoroske, M. R. Miller, S. M. O'Rourke, et al. A versatile Rapture (RAD-Capture) platform for genotyping marine turtles. *Molecular ecology resources*, 19(2):497–511, 2019. doi:10.1111/1755-0998.12980.

[234] S. Mynhardt, N. C. Bennett, and P. Bloomer. New insights from RADseq data on differentiation in the Hottentot golden mole species complex from South Africa. *Molecular Phylogenetics and Evolution*, 143:106667, 2020. doi:10.1016/j.ympev.2019.106667.

[235] F. Maroso, R. Franch, G. Dalla Rovere, et al. RAD SNP markers as a tool for conservation of dolphinfish Coryphaena hippurus in the Mediterranean Sea: Identification of subtle genetic structure and assessment of populations sex-ratios. *Marine genomics*, 28:57–62, 2016. doi:10.1016/j.margen.2016.07.003.

[236] E. G. Dierickx, A. J. Shultz, F. Sato, et al. Morphological and genomic comparisons of Hawaiian and Japanese Black-footed Albatrosses (Phoebastria nigripes) using double digest RADseq: implications for conservation. *Evolutionary applications*, 8(7):662–678, 2015. doi:10.1111/eva.12274.

[237] M. R. Campbell, N. V. Vu, A. P. LaGrange, et al. Development and application of single-nucleotide polymorphism (snp) genetic markers for conservation monitoring of burbot populations. *Transactions of the American Fisheries Society*, 148(3):661–670, 2019. ISSN 0002-8487. doi:10.1002/tafs.10157.

[238] D. B. Lowry, S. Hoban, J. L. Kelley, et al. Responsible RAD: Striving for best practices in population genomic studies of adaptation. *Molecular ecology resources*, 17(3):366–369, 2017. doi:10.1111/1755-0998.12677.

[239] K. Robasky, N. E. Lewis, and G. M. Church. The role of replicates for error mitigation in next-generation sequencing. *Nature Reviews. Genetics*, 15(1):56–62, 2014. doi: 10.1038/nrg3655.

[240] B. Arnold, R. B. Corbett-Detig, D. Hartl, and K. Bomblies. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, 22(11):3179–3190, 2013. doi:10.1111/mec.12276.

[241] M. Gautier, K. Gharbi, T. Cezard, et al. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, 22(11):3165–3178, 2013. doi:10.1111/mec.12089.

[242] D. B. Lowry, S. Hoban, J. L. Kelley, et al. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular ecology resources*, 17(2):142–152, 2017. doi:10.1111/1755-0998.12635.

[243] J. M. DaCosta and M. D. Sorenson. Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *Plos One*, 9(9):e106713, 2014. doi: 10.1371/journal.pone.0106713.

[244] D. A. R. Eaton. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13):1844–1849, 2014. doi:10.1093/bioinformatics/btu121.

[245] N. C. Rochette, A. G. Rivera-Colón, and J. M. Catchen. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, 28(21):4737–4754, 2019. doi:10.1111/mec.15253.

[246] O. A. Ali, S. M. O'Rourke, S. J. Amish, et al. RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. *Genetics*, 202(2):389–400, 2016. doi:10.1534/genetics.115.183665.

[247] R. J. Toonen, J. B. Puritz, Z. H. Forsman, et al. ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ*, 1:e203, 2013. doi:10.7717/peerj.203.

[248] J. W. Davey, T. Cezard, P. Fuentes-Utrilla, et al. Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology*, 22(11):3151–3164, 2013. doi: 10.1111/mec.12084.

[249] M. G. Harvey, B. T. Smith, T. C. Glenn, et al. Sequence Capture versus Restriction Site Associated DNA Sequencing for Shallow Systematics. *Systematic Biology*, 65(5):910–924, 2016. doi:10.1093/sysbio/syw036.

[250] P. T. Euclide, G. J. McKinney, M. Bootsma, et al. Attack of the PCR clones: Rates of clonality have little effect on RAD-seq genotype calls. *Molecular ecology resources*, 20(1):66–78, 2020. doi:10.1111/1755-0998.13087.

[251] S. P. Flanagan, B. R. Forester, E. K. Latch, et al. Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary applications*, 11(7):1035–1052, 2018. ISSN 17524571. doi:10.1111/eva.12569.

[252] D. A. R. Eaton and I. Overcast. ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, 2020. doi:10.1093/bioinformatics/btz966.

[253] A. G. Rivera-Colón, N. C. Rochette, and J. M. Catchen. Simulation with RADinitio Improves RADseq Experimental Design and Sheds Light on Sources of Missing Data. *BioRxiv*, 2019. doi:10.1101/775239.

[254] L. Bresadola, V. Link, C. A. Buerkle, et al. Estimating and accounting for genotyping errors in RAD-seq experiments. *Molecular ecology resources*, 2020. doi:10.1111/1755-0998.13153.

[255] N. C. Rochette and J. M. Catchen. Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, 12(12):2640–2659, 2017. doi:10.1038/nprot.2017.123.

[256] P. Nadukkalam Ravindran, P. Bentzen, I. R. Bradbury, and R. G. Beiko. RADProc: A computationally efficient de novo locus assembler for population studies using RADseq data. *Molecular ecology resources*, 19(1):272–282, 2019. doi:10.1111/1755-0998.12954.

[257] R. Nielsen, J. S. Paul, A. Albrechtsen, and Y. S. Song. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews. Genetics*, 12(6):443–451, 2011. doi:10.1038/nrg2986.

[258] M. Cariou, L. Duret, and S. Charlat. How and how much does RAD-seq bias genetic diversity estimates? *BMC Evolutionary Biology*, 16(1):240, 2016. doi:10.1186/s12862-016-0791-0.

[259] J. R. Willoughby, M. Sundaram, B. K. Wijayawardena, et al. The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings. *Biological Conservation*, 191:495–503, 2015. ISSN 00063207. doi:10.1016/j.biocon.2015.07.025.

[260] C. F. Graham, D. R. Boreham, R. G. Manzon, et al. How "simple" methodological decisions affect interpretation of population structure based on reduced representation library DNA sequencing: A case study using the lake whitefish. *Plos One*, 15(1):e0226608, 2020. doi:10.1371/journal.pone.0226608.

[261] J. M. Catchen, A. Amores, P. Hohenlohe, et al. Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda, Md.)*, 1(3):171–182, 2011. doi:10.1534/g3.111.000240.

[262] J. Catchen, P. A. Hohenlohe, S. Bassham, et al. Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11):3124–3140, 2013. doi:10.1111/mec.12354.

[263] C. Li, M. Hofreiter, N. Straube, et al. Capturing protein-coding genes across highly divergent species. *Biotechniques*, 54(6):321–326, 2013. doi:10.2144/000114039.

[264] H. Huang and L. L. Knowles. Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences. *Systematic Biology*, 65(3):357–365, 2016. doi:10.1093/sysbio/syu046.

[265] B. Wright, K. A. Farquharson, E. A. McLennan, et al. From reference genomes to population genomics: comparing three reference-aligned reduced-representation sequencing pipelines in two wildlife species. *BMC Genomics*, 20(1):453, 2019. doi: 10.1186/s12864-019-5806-y.

[266] C. Wang, K. B. Schroeder, and N. A. Rosenberg. A maximum-likelihood method to correct for allelic dropout in microsatellite data with no replicate genotypes. *Genetics*, 192(2):651–669, 2012. doi:10.1534/genetics.112.139519.

[267] J. G. Schraiber and J. M. Akey. Methods and models for unravelling human evolutionary history. *Nature Reviews. Genetics*, 16(12):727–740, 2015. ISSN 1471-0056. doi: 10.1038/nrg4005.

[268] A. Rodríguez-Asiain, G. Ruiz-Babot, W. Romero, et al. Brain specific kinase-1 BRSK1/SAD-B associates with lipid rafts: modulation of kinase activity by lipid environment. *Biochimica et Biophysica Acta*, 1811(12):1124–1135, 2011. doi: 10.1016/j.bbalip.2011.10.004.

[269] G. Ceballos, P. R. Ehrlich, and P. H. Raven. Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proceedings of the National Academy of Sciences of the United States of America*, 2020. ISSN 0027-8424. doi: 10.1073/pnas.1922686117.

[270] N. Myers, R. A. Mittermeier, C. G. Mittermeier, et al. Biodiversity hotspots for conservation priorities. *Nature*, 403(6772):853–858, 2000. doi:10.1038/35002501.

[271] L. W. Traill, B. W. Brook, R. R. Frankham, and C. J. Bradshaw. Pragmatic population viability targets in a rapidly changing world. *Biological Conservation*, 143(1):28–34, 2010. ISSN 00063207. doi:10.1016/j.biocon.2009.09.001.

[272] S. A. Black, J. J. Groombridge, and C. G. Jones. Leadership and conservation effectiveness: finding a better way to lead. *Conservation letters*, 4(5):329–339, 2011. ISSN 1755263X. doi:10.1111/j.1755-263X.2011.00184.x.

[273] C. Morrison, C. Wardle, and J. G. Castley. Repeatability and reproducibility of population viability analysis (PVA) and the implications for threatened species management. *Frontiers in ecology and evolution*, 4, 2016. ISSN 2296-701X. doi: 10.3389/fevo.2016.00098.

[274] R. Heinsohn, G. Olah, M. Webb, et al. Sex ratio bias and shared paternity reduce individual fitness and population viability in a critically endangered parrot. *The Journal of Animal Ecology*, 88(4):502–510, 2018. doi:10.1111/1365-2656.12922.

[275] V. Chaudhary and M. K. Oli. A critical appraisal of population viability analysis. *Conservation Biology*, 34(1):26–40, 2020. doi:10.1111/cobi.13414.

[276] G. Pe'er, Y. G. Matsinos, K. Johst, et al. A protocol for better design, application, and communication of population viability analyses. *Conservation Biology*, 27(4):644–656, 2013. doi:10.1111/cobi.12076.

[277] B. W. Brook, J. J. O'Grady, A. P. Chapman, et al. Predictive accuracy of population viability analysis in conservation biology. *Nature*, 404(6776):385–387, 2000. doi: 10.1038/35006050.

[278] W. Morris, P. Bloch, B. Hudgens, et al. Population viability analysis in endangered species recovery plans: past use and future improvements. *Ecological Applications*, 2002.

[279] M. A. McCarthy, S. J. Andelman, and H. P. Possingham. Reliability of relative predictions in population viability analysis. *Conservation Biology*, 17(4):982–989, 2003. ISSN 08888892. doi:10.1046/j.1523-1739.2003.01570.x.

[280] Y. Matamoros, R. Lacy, O. Byers, et al. Population and Habitat Viability Analysis Workshop for the Galapagos Penguin. Technical report, Galapagos National Park, Galapagos, Ecuador, 2005.

[281] A. Desbiez, K. Taylor-Holzer, R. Lacy, et al. Population Viability Analysis of jaguar populations in Brazil. 7:35–37, 2012.

[282] R. Lacy and J. Pollak. *A stochastic simulation of the extinction process*. Chicago Zoological society, Brookfield, Illinois, USA, 2014.

[283] C. Pacioni and F. Mayer. vortex R: an R package for post Vortex simulation analysis. *Methods in ecology and evolution / British Ecological Society*, 2017. ISSN 2041210X. doi:10.1111/2041-210X.12786.

[284] L. H. Andersen, P. Sunde, I. Pellegrino, et al. Using population viability analysis, genomics, and habitat suitability to forecast future population patterns of Little Owl Athene noctua across Europe. *Ecology and evolution*, 7(24):10987–11001, 2017. ISSN 20457758. doi:10.1002/ece3.3629.

[285] S. T. Kalinowski, M. L. Taper, and T. C. Marshall. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, 16(5):1099–1106, 2007. doi:10.1111/j.1365-294X.2007.03089.x.

[286] R. Lacy. Structure of the VORTEX simulation model for population viability analysis. *Ecological Bulletins*, 48:191–203, 2000.

[287] C. E. Grueber, R. J. Laws, S. Nakagawa, and I. G. Jamieson. Inbreeding depression accumulation across life-history stages of the endangered Takahe. *Conservation Biology*, 24(6):1617–1625, 2010. doi:10.1111/j.1523-1739.2010.01549.x.

[288] A. Caballero, I. Bravo, and J. Wang. Inbreeding load and purging: implications for the short-term survival and the conservation management of small populations. *Heredity*, 118(2):177–185, 2017. doi:10.1038/hdy.2016.80.

[289] K. Ralls and J. D. Ballou. Captive Breeding and Reintroduction. In S. Levin, editor, *Encyclopedia of Biodiversity,*, volume 1, pages 662–667. Academic Press, Waltham, MA, 2nd edition, 2013.

[290] N. E. Morton, J. F. Crow, and H. J. Muller. 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):855–863, 1956. doi:10.1073/pnas.42.11.855.

[291] R. C. Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, VIenna, Austria, 2019. Type: SOFTWARE.COMPUTER_SOFTWARE.

[292] T. King, C. Chamberlan, and A. Courage. Assessing reintroduction success in long-lived primates through population viability analysis: western lowland gorillas Gorilla gorilla gorilla in Central Africa. *Oryx*, 48(02):294–303, 2014. ISSN 0030-6053. doi: 10.1017/S0030605312001391.

[293] D. S. Licht, R. A. Moen, D. P. Brown, et al. The Canada Lynx (Lynx canadensis) of Isle Royale: over-harvest, climate change, and the extirpation of an island population. *The Canadian Field-Naturalist*, 129(2):139, 2015. ISSN 0008-3550. doi:10.22621/cfn.v129i2.1694.

[294] D. S. Licht. Bison conservation in northern great plains national parks: no need to panic. *Great Plains Research*, 27(2):83–92, 2017. ISSN 2334-2463. doi:10.1353/gpr.2017.0017.

[295] X. D. Zhang. A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. *Genomics*, 89(4):552–561, 2007. doi:10.1016/j.ygeno.2006.12.014.

[296] O. J. Robinson, J. L. Lockwood, O. C. Stringham, and N. H. Fefferman. A Novel Tool for Making Policy Recommendations Based on PVA: Helping Theory Become Practice. *Conservation letters*, 8(3):190–198, 2015. ISSN 1755263X. doi:10.1111/conl.12146.

[297] Y. Zhang, F. Wang, Z. Cui, et al. Can we reestablish a self-sustaining population? A case study on reintroduced Crested Ibis with population viability analysis. *Avian Research*, 12(1):14, 2021. ISSN 2053-7166. doi:10.1186/s40657-021-00250-z.

[298] C. Pacioni, A. Atkinson, S. Trocini, et al. Is supplementation an efficient management action to increase genetic diversity in translocated populations? *Ecological Management & Restoration*, 2020. ISSN 1442-7001. doi:10.1111/emr.12411.

[299] I. Jarić, T. Ebenhard, and M. Lenhardt. Population viability analysis of the Danube sturgeon populations in a Vortex simulation model. *Reviews in fish biology and fisheries*, 20(2):219–237, 2010. ISSN 0960-3166. doi:10.1007/s11160-009-9151-0.

[300] J. A. Jiménez, K. A. Hughes, G. Alaks, et al. An experimental study of inbreeding depression in a natural habitat. *Science*, 266(5183):271–273, 1994. doi:10.1126/science.7939661.

[301] N. Bunbury, M. Stidworthy, A. Greenwood, et al. Causes of mortality in free-living Mauritian pink pigeons Columba mayeri, 2002–2006. *Endangered species research*, 9:213–220, 2008. ISSN 1863-5407. doi:10.3354/esr00088.

[302] O. R. W. Pergams, R. C. Lacy, and M. V. Ashley. Conservation and management of anacapa island deer mice. *Conservation Biology*, 14(3):819–832, 2000. ISSN 08888892. doi:10.1046/j.1523-1739.2000.98524.x.

[303] D. Lindenmayer and R. Lacy. Small mammals, habitat patches and PVA models: a field test of model predictive ability. *Biological Conservation*, 103(3):247–265, 2002. ISSN 00063207. doi:10.1016/S0006-3207(01)00134-3.

[304] T. Nilsson. Population viability analyses of the Scandinavian populations of bear (Ursus arctos), lynx (Lynx lynx) and wolverine (Gulo gulo). Technical report, Swedish Environmental Protection Agency, Stockholm, 2013.

[305] P. W. Hedrick, P. E. A. Hoeck, R. C. Fleischer, et al. The influence of captive breeding management on founder representation and inbreeding in the ʻAlalā, the Hawaiian crow. *Conservation genetics (Print)*, 17(2):369–378, 2016. ISSN 1566-0621. doi:10.1007/s10592-015-0788-z.

[306] T. Coulson, G. M. Mace, E. Hudson, and H. Possingham. The use and abuse of population viability analysis. *Trends in Ecology & Evolution*, 16(5):219–221, 2001. doi:10.1016/S0169-5347(01)02137-1.

[307] M. T. Olsen, L. W. Andersen, R. Dietz, et al. Integrating genetic data and population viability analyses for the identification of harbour seal (Phoca vitulina) populations and management units. *Molecular Ecology*, 23(4):815–831, 2014. doi:10.1111/mec.12644.

[308] K. A. Witzenberger and A. Hochkirch. Evaluating ex situ conservation projects: Genetic structure of the captive population of the Arabian sand cat. *Mammalian Biology - Zeitschrift für Säugetierkunde*, 78(5):379–382, 2013. ISSN 16165047. doi: 10.1016/j.mambio.2013.03.001.

[309] R. Frankham. How closely does genetic diversity in finite populations conform to predictions of neutral theory? Large deficits in regions of low recombination. *Heredity*, 108(3):167–178, 2012. doi:10.1038/hdy.2011.66.

[310] M. C. Milinkovitch, D. Monteyne, J. P. Gibbs, et al. Genetic analysis of a successful repatriation programme: giant Galápagos tortoises. *Proceedings. Biological Sciences / the Royal Society*, 271(1537):341–345, 2004. doi:10.1098/rspb.2003.2607.

[311] J. M. Miller, M. C. Quinzin, N. Poulakakis, et al. Identification of Genetically Important Individuals of the Rediscovered Floreana Galápagos Giant Tortoise (Chelonoidis elephantopus) Provide Founders for Species Restoration Program. *Scientific Reports*, 7(1):11471, 2017. doi:10.1038/s41598-017-11516-2.

[312] C. Attard, L. Beheregaray, J. Sandoval-Castillo, et al. From conservation genetics to conservation genomics: a genome-wide assessment of blue whales (Balaenopteramusculus) in Australian feeding aggregations. *Royal Society Open Science*, 5, 2018.

[313] K. A. Witzenberger and A. Hochkirch. Ex situ conservation genetics: a review of molecular studies on the genetic consequences of captive breeding programmes for endangered animal species. *Biodiversity and Conservation*, 20(9):1843–1861, 2011. ISSN 0960-3115. doi:10.1007/s10531-011-0074-4.

[314] S. Haig and J. Ballou. Pedigree Analyses in Wild Populations. In S. Beissinger and D. McCullough, editors, *Population Viability Analysis*, pages 388–405. University of Chicago Press, Chicago, 2002.

[315] C. E. Grueber and I. G. Jamieson. Quantifying and managing the loss of genetic variation in a free-ranging population of takahe through the use of pedigrees. *Conservation genetics (Print)*, 9(3):645–651, 2008. ISSN 1566-0621. doi:10.1007/s10592-007-9390-3.

[316] I. G. Jamieson. Founder effects, inbreeding, and loss of genetic diversity in four avian reintroduction programs. *Conservation Biology*, 25(1):115–123, 2011. doi:10.1111/j.1523-1739.2010.01574.x.

[317] B. Jiménez-Mena, K. Schad, N. Hanna, and R. C. Lacy. Pedigree analysis for the genetic management of group-living species. *Ecology and evolution*, 6(10):3067–3078, 2016. doi:10.1002/ece3.1831.

[318] K.-P. Koepfli, G. Tamazian, D. Wildt, et al. Whole Genome Sequencing and Resequencing of the Sable Antelope (Hippotragus niger): A Resource for Monitoring Diversity in ex Situ and in Situ Populations. *G3 (Bethesda, Md.)*, 9(6):1785–1793, 2019. doi:10.1534/g3.119.400084.

[319] S. Haig, J. Ballou, and S. Derrickson. Management options for preserving genetic diversity: reintroduction of Guam Rails to the wild. *Conservation Biology*, 4(3):290–300, 1990.

[320] Montgomery, Margaret E., Ballou, Jonathan D., Nurthen, Roderick K., et al. Minimizing kinship in captive breeding programs. *Zoo Biology*, 16(5):377–389, 1997. doi:10.1002/(SICI)1098-2361(1997)16:5<377::AID-ZOO1>3.0.CO;2-7.

[321] S. Christie, C. Lees, and J. Wilcken. *MANAGING ZOO POPULATIONS: compiling and analysing studbook data*. 3rd edition, 1998.

[322] R. Lacy. Clarification of genetic terms and their use in the management of captive populations. *Zoo Biology*, 14:565–578, 1995.

[323] J. Ballou, C. Lees, L. Faust, et al. Demographic and Genetic Management of Captive Populations. In D. Kleiman, K. Thompson, and C. Kirk-Baer, editors, *Wild Mammals in Captivity*. University of Chicago Press, 2nd edition, 2012.

[324] R. Gooley, C. J. Hogg, K. Belov, and C. E. Grueber. No evidence of inbreeding depression in a Tasmanian devil insurance population despite significant variation in inbreeding. *Scientific Reports*, 7(1):1830, 2017. doi:10.1038/s41598-017-02000-y.

[325] S. Haig, J. Ballou, and S. Derrickson. Management options for preserving genetic diversity: reintroduction of Guam Rails to the wild. *Conservation Biology*, 4(3):290–300, 1990.

[326] F. C. Ceballos, S. Hazelhurst, and M. Ramsay. Assessing Runs of Homozygosity: A comparison of SNP Array and Whole Genome Sequence low coverage data. *BioRxiv*, 2017. doi:10.1101/160705.

[327] P. J. Crutzen. Geology of mankind. *Nature*, 415(6867):23, 2002. doi:10.1038/415023a.

[328] W. C. Funk, B. R. Forester, S. J. Converse, et al. Improving conservation policy with genomics: a guide to integrating adaptive potential into U.S. Endangered Species Act decisions for conservation practitioners and geneticists. *Conservation genetics (Print)*, 20(1):1–20, 2018. ISSN 1566-0621. doi:10.1007/s10592-018-1096-1.

[329] J. M. Miller, J. Poissant, J. T. Hogg, and D. W. Coltman. Genomic consequences of genetic rescue in an insular population of bighorn sheep (Ovis canadensis). *Molecular Ecology*, 21(7):1583–1596, 2012. doi:10.1111/j.1365-294X.2011.05427.x.

[330] A. V. Stronen, E. Salmela, B. K. Baldursdóttir, et al. Genetic rescue of an endangered domestic animal through outcrossing with closely related breeds: A case study of the Norwegian Lundehund. *Plos One*, 12(6):e0177429, 2017. doi:10.1371/journal.pone.0177429.

[331] B. J. Novak, T. Maloney, and R. Phelan. Advancing a new toolkit for conservation: from science to policy. *The CRISPR Journal*, 1:11–15, 2018. ISSN 2573-1599. doi:10.1089/crispr.2017.0019.

[332] Koumbarakis, Hirschi, Meier, et al. Nature is too big to fail. Technical report, World Wildlife Fund, 2020.

Appendix A

A.1 Chapter 2: Whole genome variation in the pink pigeon *Nesoenas mayeri*

A.1.1 Checking synteny

Dotplots were produced using genome ribbon [200] (http://genomeribbon.com/) to show how the pink pigeon genome aligned to other bird species. These plots confirmed the high levels of synteny between the pink pigeon other bird species (see Figure A.1) providing evidence that it is a good candidate for a reference assisted chromosome assembly.

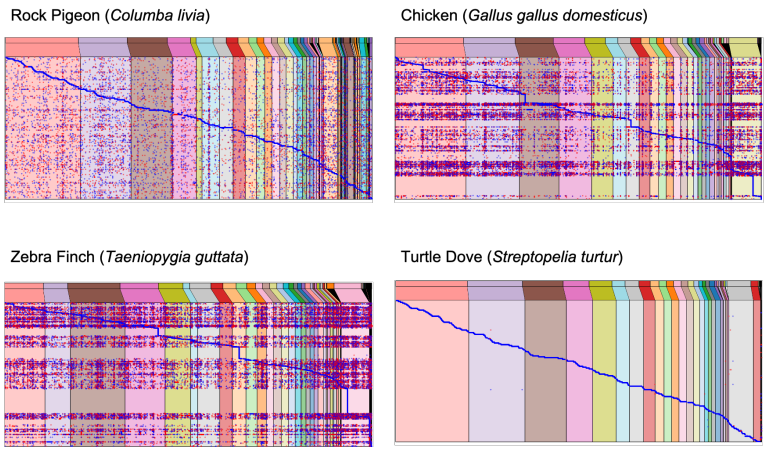
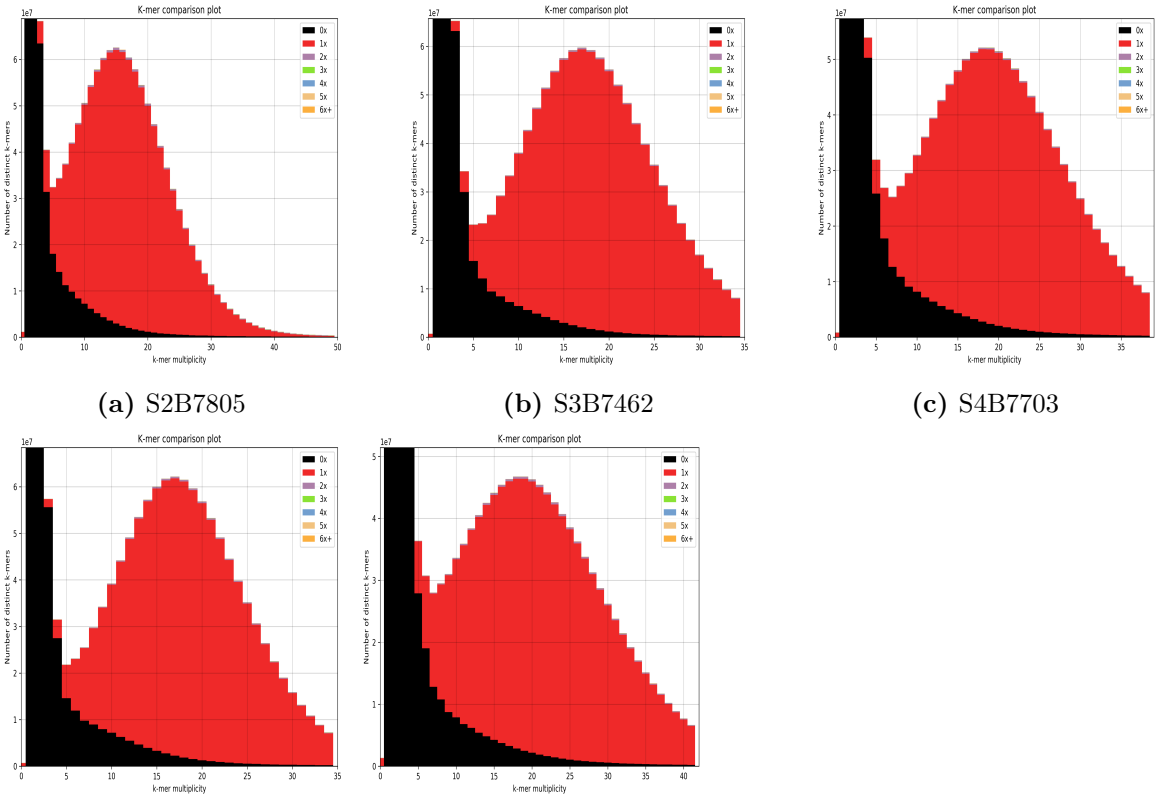


Figure A.1 Dotplots showing the results of an alignment between the pink pigeon and four other bird species. In all different coloured block represents different chromosomes (the numbers have been removed for clarity) of the reference genome. In each plot the reference genome is the species indicated by the title and the query is the pink pigeon genome. Blue shows reads aligned in forward directions, red shows reads aligned in the reverse direction.

A.1.2 Supernova genome assembly quality assessment

The genome assemblies assembled using Supernova v2.1.1 for use in the PSMC analysis were assessed to ensure that there were of high enough quality. The completeness, correctness and contiguity of the genome assemblies were assessed using KAT (see Figure A.2), BUSCO v3.1.0 (see Table A.1) and Abyss v1.9 (see Table A.2) respectively.



(a) S2B7805 **(b)** S3B7462 **(c)** S4B7703

(d) S6W1688 **(e)** S7W1687

Figure A.2 K-mer spectra comparing comparing the k-mers in the processed 10x reads to those in the final assembly. The heterozygous peak is present however due to the stacked bar charts and overall lower k-mer coverage it is difficult to see. This was confirmed by re-plotting the spectra with the missing k-mers (black) removed.

ID	Complete and single-copy BUSCOs
S2B7805	90.0%
S3B7462	91.0%
S4B7703	91.7%
S6W1688	92.1%
S7W1687	88.6%

Table A.1 BUSCO results for five 10x assemblies. Showing only the proportion of BUSCOs that were found complete and as a single-copy.

Table A.2 Assembly statistics for five genome assembled using Supernova v2.1.1 produced using Abyss v1.9.0

Name	n	Min	N50	E-size	Max	Sum
S2B7805	30076	558	4083223	5254941	20.97e6	1.117e9
S3B7462	17834	800	1902505	2587881	11.01e6	1.106e9
S4B7703	15788	802	7362214	11.14e6	41.84e6	1.107e9
S6W1688	16397	823	13.17e6	15.9e6	45.9e6	1.11e9
S7W1687	34114	801	14.05e6	15.73e6	49.61e6	1.111e9

A.1.3 Comparing genome wide heterozygosity in 41 different bird species

Table A.3 Information about the 41 bird species used to compare genome wide levels of heterozygosity. All data (apart from pink pigeon) taken from [6]

Latin names	Common Names	Status*	#SNPs	Heterozygosity (10
Nipponia nippon	Crested Ibis	EN	478,836	0.43
Balearica regulorum	Grey crowned-crane	EN	1,996,436	1.88
Chlamydotis macqueenii	MacQueen's Bustard	VU	2,755,985	2.67
Haliaeetus albicilla	White-Tailed Eagle	LC	429,933	0.4
Haliaeetus leucocephalus	Bald Eagle	LC	476,460	0.43
Mesitornis unicolor	Brown Mesite	VU	1,386,012	1.34
Nestor notabilis	Kea	EN	900,391	0.91
Pelecanus crispus	Dalmation Pelican	NT	652,406	0.6
Acanthisitta chloris	Rifleman	LC	1,631,045	1.67
Anas platyrhynchos	Mallard	LC	2,917,640	2.91
Apaloderma vittatum	Bar-tailed Trogon	LC	2,196,131	2.18
Calypte anna	Anna's Hummingbird	LC	2,305,206	2.55
Caprimulgus carolinensis	Chuck-will's-widow	NT	4,134,396	3.87
Cariama cristata	Red-legged Seriema	LC	2,705,802	2.55
Cathartes aura	Turkey Vulture	LC	1,087,091	1.18
Charadrius vociferus	Killdeer	LC	3,392,281	3.05
Colius striatus	Speckled Mousebird	LC	2,898,272	2.84
Columba livia	Rock Dove	LC	1,742,895	1.72
Corvus brachyrhynchos	American Croq	LC	1,695,667	1.72
Cuculus canorus	Common Cuckoo	LC	5,626,221	5.42
Egretta garzetta	Little Egret	LC	2,683,899	2.51
Eurypyga helias	Sunbittern	LC	3,492,244	3.46
Falco peregrinus	Peregrine Falcon	LC	1,094,615	1.01
Fulmarus glacialis	Northern Fulmar	LC	2,244,766	2.06
Gavia stellata	Red-throated Loon	LC	2,371,415	2.24
Geospiza fortis	Medium Ground-finch	LC	1,053,816	1.1
Leptosomus discolor	Cuckoo Roller	LC	2,648,923	2.47
Manacus vitellinus	Golden-collared Manakin	LC	3,826,080	4.02
Melopsittacus undulatus	Budgerigar	LC	4,316,449	4.31
Merops nubicus	Northern Carmine Bee-eater	LC	3,641,719	3.67
Opisthocomus hoazin	Hoatzin	LC	2,495,191	2.29
Phaethon lepturus	White-tailed Tropicbird	LC	1,581,419	1.5
Phalacrocorax carbo	Great Cormorant	LC	1,416,666	1.39
Phoenicopterus ruber	American Flamingo	LC	3,542,785	3.34
Picoides pubescens	Downy Woodpecker	LC	4,008,877	3.9
Podiceps cristatus	Great Crested Grebe	LC	2,174,676	2.08
Pterocles guturalis	Yellow-throated Sandgrouse	LC	1,449,156	1.46
Struthio camelus	Common Ostrich	LC	1,915,423	1.71
Tauraco erythrolophus	Red-crested Turaco	LC	2,676,542	2.48
Tyto alba	Common Barn-owl	LC	1,252,107	1.25

A.1.4 Scripts and Data

All files listed in this appendix are available at:

- https://ueanorwich-my.sharepoint.com/:f:/g/personal/vmw16vqu_uea_ac_uk/EuWg_gbZRM1FrPU-GyFQV3EB_KqZxCNjOv1nc9qXG-Rdog?e=sw3xg6
- Password: **pinkpigeon**
- It was not possible to upload the raw reads because of their large file sizes however they are available on request (camilla.ryan@earlham.ac.uk).

- **Data**
  1. Genome assemblies
  2. VCF
  3. Heterozygosity
  4. RoH
  5. ZooRoH
  6. PSMC
  7. Files needed for processing the data

- **Scripts**
  1. Code for generating the assembly
  2. Code for generating the human k-mer spectra
  3. Workflow for processing the whole genome resequencing data
  4. R script for calculating observed heterozygosity
  5. R script for generating genome-wide variation plot
  6. R script for generating plot to compare levels of observed heterozygosity in the pink pigeon and 40 other avian species
  7. Code for generating data for runs of homozygosity
  8. R script for calculating F_{FRoH} and generating plots
  9. Code for generating data for ZooRoH
  10. R script for running ZooRoH
  11. Code for generating data for PSMC and running PSMC analysis

A.2 Chapter 3: RADiKal: a novel tool for generating an overview of genome-wide variation in a population at-a-glance

A.2.1 Scripts and Data

All files listed in this appendix are available at:

– https://ueanorwich-my.sharepoint.com/:f:/g/personal/vmw16vqu_uea_ac_uk/EuWg_gbZRM1FrPU-GyFQV3EB_KqZxCNjOv1nc9qXG-Rdog?e=sw3xg6

– Password: **pinkpigeon**

- RADiKal_workflow
- **Data**
  1. Data used to plot number of published papers using RAD-seq data
  2. An example of the output generated by RADiKal for pink pigeon chromosome 1, including the pink pigeon chromosome 1 .fasta file. and the scores produced from creating the classifier. This output includes a directory of classifier images which are putative RAD, Non-RAD that scored below 0 or above 0 respectively. It also includes a directory of scores for each individual.
  3. File containing the length of each chromosome
  4. Files produced by some of the scripts and then used to calculate the density of sites across the genome
- **Scripts**
  1. Code for plotting graph of published RAD studies
  2. Directory containing scripts to create RADiKal's classifier and run RADiKal.
     (a) Code to run RADiKal and initiate classifier.py.
     (b) Classifier function. This is the workhorse of RADiKal and contains all the classes and functions necessary to create classifiers and score sites.
  3. Directory containing all the scripts needed to perform the analyses in RADiKal and created painted chromosomes

(a) Code to create plot of spread of scores from the the RAD and non-RAD sites used to create the classifier.

(b) Code to determine how many sites in classifier and to create plot to find the number of sites in each of the three scored peaks.

(c) Code to create the plot of density of RAD sites across the genome.

(d) Code to generates bins used in subsequent steps.

(e) Code to generate scoring matrix used to pain chromosomes.

(f) Code to create painted chromosomes and density plots (including those for Chromosome 5).

(g) Code to produce variation plots for the whole genome resequencing data (including those for Chromosome 5).

A.3 Chapter 4: Genetic rescue recommended: An updated PVA for the pink pigeon *Nesoenas mayeri*

The following pages contain the parameters used in producing the Vortex models.

Input Variable	Variable	Description	Reference
	Scenario settings		
Number of iterations	1000	-	-
Number of years	100	-	-
Duration year in days	365	-	-
Extinction definition	one sex remains	-	-
Number of populations	6*	-	-
Order of events	EV, Breed, Mortality, Disperse, Harvest, Supplement, Breed, Age	Pink pigeons can and do breed before a year old. To model this, another Breed event is included. This also provides the newly supplemented birds the opportunity to breed in the same year they are introduced.	(Jones 1987)
	Species Description		
Lethal Equivalents	15.13	These were calculated using logistic regression method with the small sample size correction.	Calculated from studbook data
Percentage inbreeding due to LE	50%	Default	(Lacy and Traylor-Holzer 2017) (Simmons and Crow 1977)
Environmental Correlation between reproduction and survival	0.75	Pink pigeons breed during most of the year and do not disperse far, and therefore, a good year for reproduction is also likely to be a good year for survival (barring catastrophes)	(Jones 1987)
Environmental correlation among populations	0.75	All the free-living subpopulations are relatively geographically close (except Ile Aux Aigrette) and will be subject to similar environmental fluctuations	(Jones 1987)
	Dispersal		

(Concannon 2014)						
	PW	**BF**	**BO**	**IAA**	**CO**	**Captive**
PW	75.43	4.17	2.22	0.18	18.18	0.00
BF	8.70	87.66	0.00	0.00	3.64	0.00
BO	4.35	0.00	90.20	0.00	5.45	0.00
IAA	0.00	0.00	2.22	97.78	0.00	0.00
CO	26.09	3.33	11.11	0.00	59.47	0.00
Captive	0.00	0.00	0.00	0.00	0.00	100

Reproductive system			
System	Long term monogamy	-	(Jones 1987)
Age of first offspring female	1	-	(Jones 1987)
Age of first offspring male	1	-	(Jones 1987)
Max lifespan	15	-	(Jones 1987)
Max age female reproduction	5	-	(Jones 1987)
Max age male reproduction	15	-	(Jones 1987)
Max number broods per year	12	Pink pigeons have the ability to produce multiple broods per year particularly if a clutch fails. However, the number of successful broods is far lower than what they are capable of and there is a negative relationship between the number of broods and the number of fertile eggs an individual lays	(Lind 1989)
Max progeny per brood	2	-	(Jones 1987)
Sex ratio at birth	54	-	
Reproductive rates			

% adult female breeding	80							
SD in % breeding due to EV	10							
Distribution of broods per year			Wild		Captive			
	0		9.52		36			
	1		30		20			
	2		50.96		10			
	3		7.14		10			
	4		2.38		5			
	5				5			
	6				14			
	7							
Distribution of number of offspring per brood	1		88		78		The distribution of number of offspring per wild brood was taken from the original PHVA conducted in 1991. Whereas the figures for the captive population were calculated from studbook data.	(US. Seal and Bruford 1991); pink pigeon studbook
	2		12		22			
Mortality								
	PW	BF	BO	IAA	CO	CA		
Mortality 0-1	54.4	54.4	54.4	54.4	54.4	34	The total mortality of individuals includes mortality due to inbreeding depression† This figure represents the mortality of moderately inbred birds from hatching to fledging.	(Swinnerton et al. 2004)
SD in mortality 0-1 due to EV	5	5	5	5	5	5	-	(US. Seal and Bruford 1991)
Mortality 1+	14.45	11.65	19.4	12.8	15.6	15	-	(Bunbury 2006)
SD in mortality 1+ due to EV	7.25	6.27	5.55	9.17	4.80	5	-	(Bunbury 2006)
Initial population size								
PW	70							
PL	115						2010 estimates, which are the most recent comprehensive estimates available	(Concannon 2014)
BO	50							
IAA	60							

CO	50			
Captive	79			
Age distribution	Stable			
Carrying capacity (K) (Jones 1987; US. Seal and Bruford 1991)				
PW	200		The carrying capacity of the free-living metapopulation is higher than would naturally occur because MWF provide supplementary food to the pink pigeons which increases the carrying capacity and 200 is thought to be ideal. Except for the Ile Aux Aigrette which is limited by it being a closed population with a limited number of territories	(US. Seal and Bruford 1991; Jones 1987)
PL	200			
BO	200			
IAA	80			
CO	200			
Captive	500			
Catastrophes				
Cyclones	6.7%		Cited as a major cyclone once every 15 years	(Jones 1987)
Harvest				
Harvest and supplementation are closely linked in vortex. To supplement from the captive population, it is necessary to first harvest form it. There has been some mortality observed from supplementations which is accounted for in supplementation. The supplementation/ harvest routine decided upon was one considered achievable (given past reintroductions).				
Percent survival during translocation	100		-	-
Population harvested	CA		-	-
First year of harvest	1		-	-
Last year of harvest	99		-	-
Interval between harvests	5		-	-
Number of each sex to be harvested from age 0-1	20		-	-
Number of each sex to be harvested after age 1	5		-	-
Supplementation				
Percent survival during translocation	89.87		-	(Concannon 2014)
Population supplemented	PW BF BO IAA CO		-	

First year of supplement	5	6	7	8	9		-	
Last year of supplement	95	96	97	98	99		-	
Number of each sex to be supplemented from age 0 - 1	4	4	4	4	4		Juvenile birds (< 6 months) are preferred for translocations which is why there are a greater number of younger birds being supplemented.	(Concannon 2014);Zuel per comms. 2017
Number of each sex to be supplemented after age 1	1	1	1	1	1		The total number of birds chosen to be supplemented was an arbitrary choice because supplementation regimes are not the primary focus of this paper. However, the proposed numbers for the model are based on historical reintroduction attempts, and this number of birds can be taken from the captive population without significantly depleting it.	
Genetics								
Number of neutral loci to be modelled	22						-	
Loci to be included in summary statistics	Additional loci only						Only include the statistics from our data otherwise Vortex will include analyses of the alleles it has modelled by default.	
Number of loci to be subject to mutation	22						-	
Mutation rate	0.0001						-	(Harrisson et al. 2016)
Start populations with all inbreeding and kinships set to	PW 0.15	BF 0.15	BO 0.15	IAA 0.15	CO 0.15	CA 0.10/0.15	Calculated based on inbreeding coefficients generated from the studbook data by the software PMX§	(Swinnerton et al. 2004), Studbook

The following pages contain the results from literature search of the publications on Vortex's website (http://www.vortex10.org/VortexReferences.aspx [accessed June 2018]) listing studies which included inbreeding depression (heterosis model) in their simulations.

Common name	Scientific name	Status at time of PVA[A]	Current status[B]	Purpose of PVA	Default LE used? (#)[C]	Sensitivity Testing (min,max)[D]	Reference
Western Lowland Gorilla	*Gorilla gorilla gorilla*	CE	CE	Examine long term viability of two reintroduced populations	Y (3.14)	Y (1,5)	(King et al. 2014)
Mountain Brushtail possum	*Trichosurus cunninghami*	NA	LC	Examine the impact of fragmenting habitat and the subdivision of populations on genetic variability of the species	Y (3.14)	N	(Lacy and Lindenmayer 1995)
Bison	*Bison bison*	NT	NT	Looking for evidence of inbreeding depression and whether there is a need to put in place management plans to prevent it.	No (range values)	N	(Licht 2017)
Bison	*Bison bison*	NT	NT	Looking at the feasibility of reintroducing Bison to national park	Y (6.29)	N	(Licht 2014)
Lynx	*Lynx canadensis*	LC	LC	Feasibility and viability of reintroducing Lynx to Isle Royale	Y (6.29)	Y (0, 12.58)	(Licht et al. 2017)
Lynx	*Lynx canadensis*	LC	LC	Modelled the historic Isle Royale population to learn about how they maintained their population on the island and the eventual causes for lynxs decline	Y (3.14)	N but model no inbreeding	(Licht et al. 2015)
Mountain Brushtail possum & Greater glider	*Trichosurus cunninghami* & *Petauroides volans*	NL & LC	LC & V	Look at how metapopulation dynamics effects two species in same guild (they also compared to third species previously published)	Y (3[E])	N but model no inbreeding	(Lindenmayer and Lacy 1995a)
Leadbeater's Possum	*Gymnobelideus leadbeateri*	E	CE	Viability analysis of the possum considering the size and number of availlable habitat patches within wood production forests	Y (3[E])	N	(Lindenmayer and Lacy 1995b)

Common name	Scientific name	Status at time of PVA[A]	Current status[B]	Purpose of PVA	Default LE used? (#)[C]	Sensitivity Testing (min,max)[D]	Reference
Greater glider & Mountain brushtail possum & Ring tailed possum	*Petauroides volans* & *Trichosurus cunninghami* & *Pseudocheirus peregrinus*	LC & NA & LC	LC & V & LC	Ran scenarios to predict the abundance of three arboreal mammals and determine which models best predicted the collected field data	Y (3.14)	N but model no inbreeding	(Lindenmayer et al. 2000)
Bush rat & Agile antechinus	*Rattus fuscipes* & *Antechinus agilis*	LC & LC	LC & LC	Comparing patch occupancy as modelled by Vortex and collected field data	N (2[F])	N	(Lindenmayer and Lacy 2002)
Bottlenose dolphins	*Tursiops aduncus*	DD	DD	Evaluate the relative importance of reproduction and survival on two populations of bottlenose dolphins	NA	Y (0,6.28)	(Manlik et al. 2016)
Sumatran & Bornean Orangutan	*Pongo abelii* & *Pongo pygmaeus*	NA & E	CE & CE	Use PVA to look at how Orangutan ecology has implications for its conservation under different types of threats	N (4.06)	Y (0, 6)	(Marshall et al. 2008)
Thylacine	*Thylacinus cynocephalus*	Ex	Ex	Used PVA as part of a new meta model approach which aims at modelling species interactions to design more accurately models, this meta model was used to reassess the reasons for the Thylacine's extinction	Y (3.14)	Y (model no inbreeding or inbreeding)	(Prowse et al. 2013)
Anacapa Island Deer mice	*Peromyscus maniculatus anacapae*	LC[G]	LC[G]	Used to develop captive breeding and reintroduction plan	N (1.28)	N	(Pergams et al. 2000)
Koala	*Phascolarctos cinereus*	LC	V	Demographic forecasting of Koala popualtions with and without management. They then retro fitted models to see how well the models performed	Na	Y (0, 3.14)	(Penn et al. 2000)
Brown bear, Eurasian Lynx, Wolverine	*Ursus arctos*, *Lynx lynx*, *Gulo gulo*	LC, LC, LC	LC, LC, LC	Formed scientific basis of report to EU commission needed in accordance with the habitats directive and added to current management plans	N (6.50/10.50)	NA they did but not altering LE	(Nilsson 2013)

Common name	Scientific name	Status at time of PVA[A]	Current status[B]	Purpose of PVA	Default LE used? (#)[C]	Sensitivity Testing (min,max)[D]	Reference
White lipped peccary	*Tayassu pecari*	V	V	Simulate population trajectories and extinction risk under the current hunting pressure and different management scenarios.	Y (3.14)	N	(Rivera 2014)
Feral horse	*Equus caballus*	NL	NL	Model the impact of using contraception as way to control the population growth as population too large. Also to see the impact of changing age structure and inbreeding that would result from such management	Y (3.14)	N	(Ballou et al. 2008)
Jaguar	*Panthera onca*	NT	NT	Produce baseline model during workshop that can be adapted to be used for studying any specific Jaguar population in any Biome.	Y (6[E])	N	(Desbiez et al. 2012)
Helmeted honey eater	Lichenostomus melanops cassidix	LC[G]	LC[G]	Wanted to see whether hybridisation with common neighbouring subspecies was viable conservation strategy incorporating microsatellite data. Consider genetic rescue plan alongside a demographic rescue.	Y (6.29)	N	(Harrisson et al. 2016)
Woylie	Bettongia penicillata	CE[G]	CE[G]	Developed PVA to identify demographic and genetic responses to different variables (predation, predator control, population size, heterozygosity and inbreeding depression) and use that information to develop a framework that could be used with other bettong species.	N (3.14 – 10)	Y	(Pacioni et al. 2017)

A – Status according to IUCN Redlist data, this information is for the species as a whole and therefore does not include population level statuses or national or federal listings. NL - Not Listed; DD - Data Deficient; LC - Least Concern; NT - Near Threatened; V - Vulnerable; E - Endangered; CE – Critically Endangered; Ex – Extinct.
B – From the most recent assessment
C - The default number of LE varies depending on the version of Vortex that was used. Vortex version 9 used LE = 3.14 and Vortex version 10 uses LE = 6.29 see Vortex manual for further details.
D – Specifically sensitivity for LE or inbreeding depression
E – Based on default no reason given for why rounded
F – Used for both species
G – The subspecies specifically was not listed individually in IUCN so the statuses refer to the species

Bibliography for literature search

Ballou, J.D., Traylor-Holzer, K., Turner, A., Malo, A.F., Powell, D., Maldonado, J. and Eggert, L. 2008. Simulation model for contraceptive management of the Assateague Island feral horse population using individual-based data. Wildlife Research 35(6), p. 502.

Desbiez, A., Taylor-Holzer, K., Lacy, R., Beisiegel, B., Breitenmoser-Wursten, C., Sana, D., Moraes Jr, E., Carvalho Jr, E., Boulhosa, F., Paula, R., Morato, R., Cavalcanti, S. and De Oliveira, T. 2012. Population Viability Analysis of jaguar populations in Brazil. CATnews (Special Issue) 7, pp. 35–37.

Harrisson, K.A., Pavlova, A., Gonçalves da Silva, A., Rose, R., Bull, J.K., Lancaster, M.L., Murray, N., Quin, B., Menkhorst, P., Magrath, M.J.L. and Sunnucks, P. 2016. Scope for genetic rescue of an endangered subspecies though re-establishing natural gene flow with another subspecies. Molecular Ecology 25(6), pp. 1242–1258.

King, T., Chamberlan, C. and Courage, A. 2014. Assessing reintroduction success in long-lived primates through population viability analysis: western lowland gorillas Gorilla gorilla gorilla in Central Africa. Oryx 48(02), pp. 294–303.

Lacy, R.C. and Lindenmayer, D.B. 1995. A simulation study of the impacts of population subdivision on the mountain brushtail possum Trichosurus caninus ogilby (phalangeridae: marsupialia), in south-eastern Australia. ii. loss of genetic variation within and between subpopulations . Biological conservation 73, pp. 131–142.

Licht, D. 2014. Restoration of Bison (Bison bison) to Agate Fossil Beds National Monument. Fort Collins, CO: National Park Service.

Licht, D.S., Moen, R.A., Brown, D.P., Romanski, M.C. and Gitzen, R.A. 2015. The Canada Lynx (Lynx canadensis) of Isle Royale: over-harvest, climate change, and the extirpation of an island population. The Canadian Field-Naturalist 129(2), p. 139.

Licht, D.S. 2017. Bison conservation in northern great plains national parks: no need to panic. Great Plains Research 27(2), pp. 83–92.

Licht, D.S., Moen, R.A. and Romanski, M. 2017. Modeling viability of a potential canada lynx reintroduction to isle royale national park. Natural Areas Journal 37(2), pp. 170–177.

Lindenmayer, D. and Lacy, R. 1995a. Metapopulation Viability of Arboreal Marsupials in Fragmented Old-Growth Forests: Comparison Among Species. Ecological Applications 5(1), pp. 183–199.

Lindenmayer, D. and Lacy, R. 1995b. Metapopulation Viability of Leadbeater's Possum, Gymnobelideus Leadbeateri, inFragmented Old-Growth Forests. Ecological Applications 5(1), pp. 164–182.

Lindenmayer, D.B., Lacy, R.C. and Pope, M.L. 2000. TESTING A SIMULATION MODEL FOR POPULATION VIABILITY ANALYSIS. Ecological Applications 10(2), pp. 580–597.

Lindenmayer, D.B. and Lacy, R.C. 2002. Small mammals, habitat patches and PVA models: a field test of model predictive ability. Biological Conservation 103(3), pp. 247–265.

Manlik, O., McDonald, J.A., Mann, J., Raudino, H.C., Bejder, L., Krützen, M., Connor, R.C., Heithaus, M.R., Lacy, R.C. and Sherwin, W.B. 2016. The relative importance of reproduction and survival for the conservation of two dolphin populations. Ecology and evolution 6(11), pp. 3496–3512.

Marshall, A.J., Lacy, R., Ancrenaz, M., Byers, O., Husson, S.J., Leighton, M., Meijaard, E., Rosen, N., Singleton, I., Stephens, S., Traylor-Holzer, K., Utami Atmoko, S.S., van Schaik, C.P. and Wich, S.A. 2008. Orangutan population biology, life history, and conservation. In: Orangutans. Oxford University Press, pp. 311–326.

Nilsson, T. 2013. Population viability analyses of the Scandinavian populations of bear (Ursus arctos), lynx (Lynx lynx) and wolverine (Gulo gulo). Stockholm: Swedish Environmental Protection Agency.

Pacioni, C., Williams, M.R., Lacy, R.C., Spencer, P.B.S. and Wayne, A.F. 2017. Predators and genetic fitness: key threatening factors for the conservation of a bettong species. Pacific Conservation Biology 23(2), p. 200.

Penn, A.M., Sherwin, W.B., Gordon, G., Lunney, D., Melzer, A. and Lacy, R.C. 2000. Demographic forecasting in koala conservation. Conservation Biology 14(3), pp. 629–638.

Pergams, O.R.W., Lacy, R.C. and Ashley, M.V. 2000. Conservation and management of anacapa island deer mice. Conservation Biology 14(3), pp. 819–832.

Prowse, T.A.A., Johnson, C.N., Lacy, R.C., Bradshaw, C.J.A., Pollak, J.P., Watts, M.J. and Brook, B.W. 2013. No need for disease: testing extinction hypotheses for the thylacine using multi-species metamodels. The Journal of Animal Ecology 82(2), pp. 355–364.

Rivera, C.J. 2014. Facing the 2013 Gold Rush: A Population Viability Analysis for the Endangered White-Lipped Peccary (Tayassu pecari) in Corcovado National Park, Costa Rica. Natural Resources 05(16), pp. 1007–1019.

A.3.1 Scripts and Data

All files listed in this appendix are available at:

- https://ueanorwich-my.sharepoint.com/:f:/g/personal/vmw16vqu_uea_ac_uk/EuWg_gbZRM1FrPU-GyFQV3EB_KqZxCNjOv1nc9qXG-Rdog?e=sw3xg6
- Password: **pinkpigeon**

- Data
  1. Data for generating mean inbreeding
  2. Studbook data taken from PMx for captive birds from 1977-1998 used to calculate inbreeding plots
  3. Data used for the calculation of inbreeding plots
  4. Data from all living birds used for the calculation of inbreeding plots
  5. Inbreeding coefficients for GDEWS and wild population
  6. Data used in regression analysis (longevity and inbreeding coefficients) to calculate lethal equivalents
  7. Directory containing allele frequency files use in Vortex models
     (a) Allele frequencies of the wild and captive population used for Scenario 1 and Scenario 3.
     (b) Allele frequencies of wild birds nad hypothetical wild sub population used in Scenario 2
- Scripts
  1. Code to plot mean inbreeding over time in captive population
  2. Code to plot distribution of inbreeding coefficients
  3. Code to plot inbreeding (GDEWS and wild) and number birds fledged from GDEWS
  4. Code to perform regression analysis used to calculate lethal equivalents and plot the results
  5. Code to calculate and plot the results from the 3 main vortex scenarios

6. Code to calculate and plot the results from the single factor sensitivity testing.

- Vortex_files
  - Contains vortex project file and files produced by Vortex models. Also contains some of the results of VortexR analyses.

A.4 Chapter 5: I Choose You: A simple tool to help preserve genetic diversity

A.4.1 Scripts and Data

All files listed in this appendix are available at:

- https://ueanorwich-my.sharepoint.com/:f:/g/personal/vmw16vqu_uea_ac_uk/EuWg_gbZRM1FrPU-GyFQV3EB_KqZxCNjOv1nc9qXG-Rdog?e=sw3xg6
- Password: **pinkpigeon**

- Data
  1. Pink pigeon international studbook
  2. Data for use with ICY contains data from the genetics module of PMx
  3. Data for use with ICY pairwise kinship matrix generated in PMx
  4. Data for use with ICY founder representation from PMx
  5. Results of the proof of concept ICY vs. ten random groups
  6. Results of the proof of concept comparing founder equivalents in the group suggested by ICY vs. 1000 random groups
- ICY_report
  - downloadable report from ICY for the pink pigeon using the same filters as the proof of concept (3 males, 3 females etc.). The same report can be found in this appendix at the end of list of scripts.
- ICY_software

  - All scripts necessary to run ICY, as would be found on GitHub (https://github.com/Mills33/ICY). This includes all the documentation about how to use ICY.

- Scripts

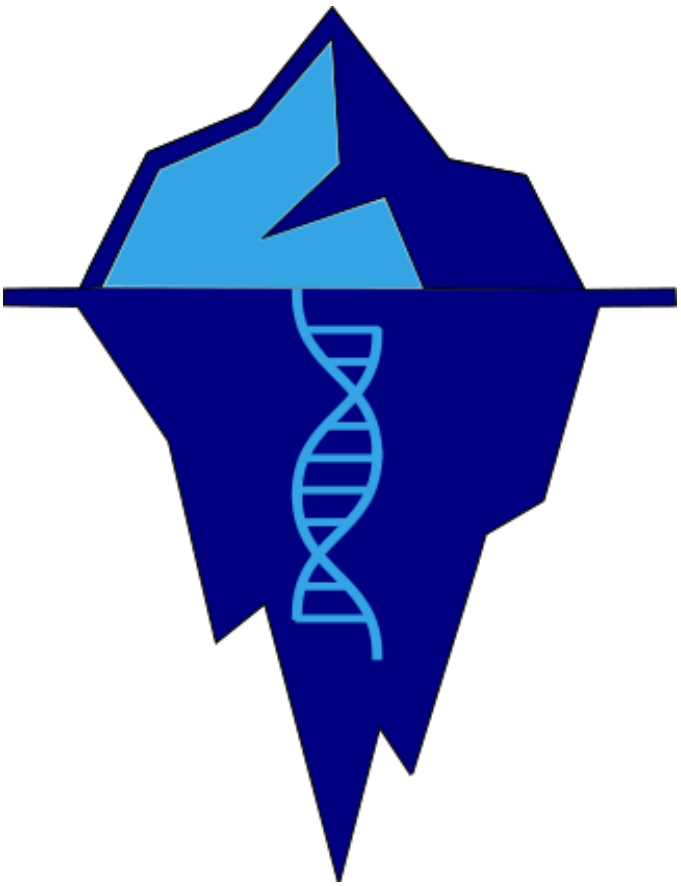
  1. Code to generate 10 random groups of 3 males and 3 female birds
  2. Code to plot the results of the proof of concept
  3. Code to generate 1000 random groups for the proof of concept comparing founder equivalents
  4. Code to plot the results of proof of concept comparing founder equivalents
  5. Code to plot the pink pigeon captive population and founder representation over time

A.4.2 Example ICY report

from ICY for the pink pigeon using the same filters as the proof of concept (3 males, 3 females etc.) Ane example of a report produced by ICY which includes information on the pink pigeon captive popualtio genetic health based on studbook infroamtion.

I.C.Y Report

19 August, 2020



How to use this report

The following report uses studbook data to generates a group of individuals recommended for reintroduction (where group size is equal to the number of females needed + number males needed). This report also provides tables and plots designed to give practitioners information about all the individuals that were in the selection they uploaded to ICY. The suitability of a individual for translocation is based on (1) the number of founder equivalents (Fe) in an individual, its (2) mean kinship coefficient (MK), which is a measure of a individuals average relatedness to other individuals in the population and (3) it relatedness to other individuals that would be released with it.

Founder equivalents are a way of measuring genetic diversity from pedigree data. They can be thought as how many wild caught founders would be needed to recreate the current level of genetic diversity and can be calculated for individuals or for the population as a whole. Founder equivalents is a metric that takes into account both the number of founder genomes that an individual is descended from and also each founder's relative contributions (or **founder representation**). It is perhaps intuitive that the more founders an individual is descended from the greater its genetic diversity. However the relative contribution of each of the founders is incredibly important as uneven contributions lead to a loss in genetic diversity and an increase in inbreeding. The maximum number of founder equivalents that an individual could have would be equal to the number of individuals that founded the population (according to the studbook) and would only occur if all of the founders were represented evenly within that individual.

Figure 1 gives a simple example of how founder representation is calculated, it also demonstrates how this can be graphically represented, **figure 2** gives an example demonstrating how graphs of even and uneven founder representations may look.

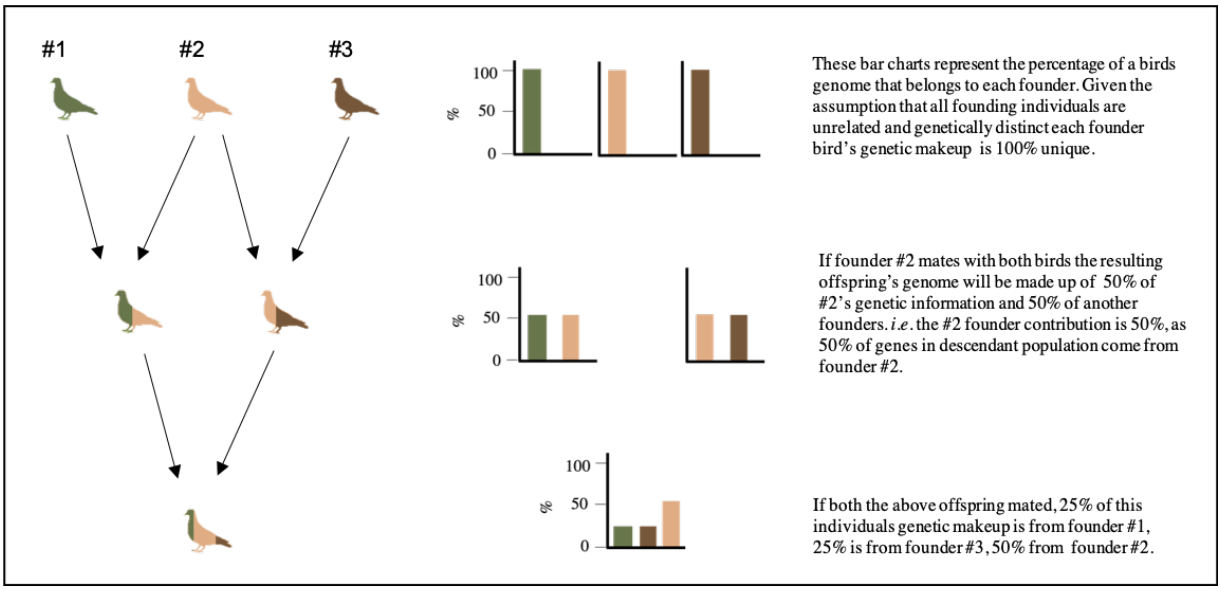


Figure 1: An example of how founder representation is calculated and how it can be visualised with bar charts

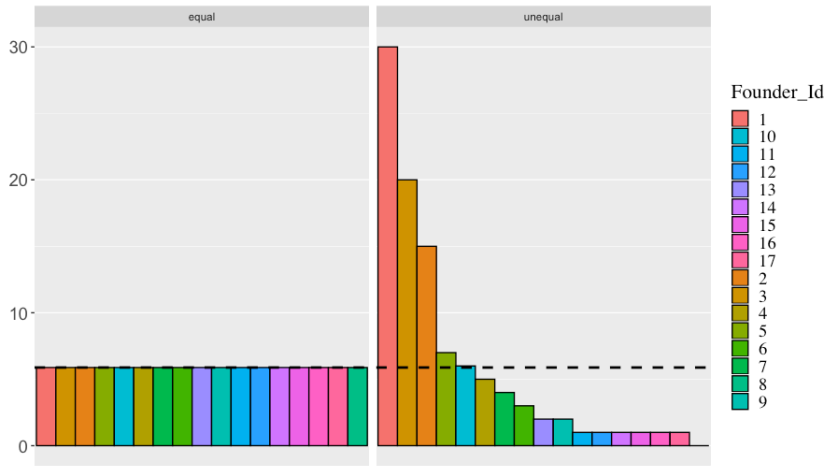


Figure 2: An example of how equal and unequal founder representation graphs may look. The dashed line represents the expected height of the bars if founder representation was equal which in this example, with 17 founders, is 5.88%

How to choose individuals for reintroduction

This report should make it easier to choose a group of individuals based on their possible genetic diversity and relatedness however the rankings do not include any other information that may, or may not, make an individual suitable for reintroduction. Ultimately it will be up to the person reading this report to decide what other factors (other than genetic ones) make a individual suitable for translocation. The points below provide further guidance for selecting individuals to maximise genetic diversity in the released individuals and ensuring the overall genetic health of the captive and released populations.

- **Group recommended by ICY** - These are not simply the highest ranked individuals (rankings based on MK and Fe). But the individuals with the highest rankings that are less related the the specified relatedness threshold. This is an important point related individuals may be highly ranked but if you released a group of highly related individuals the resulting population would not be genetically healthy and would likely suffer from inbreeding depression.

- **Pragmatic considerations** - Age, location, possible disease or health issues, are the individuals available for translocations?

- **How valuable are the individuals to the captive population?**- When you remove a individual from captivity you remove its genes and so a balance must be struck between translocating individuals and ensuring a healthy captive population that can continue to produce healthy, genetically diverse individuals for future reintroductions. Therefore it may be better to initially choose slightly lower ranking

individuals because they will still have a positive impact on the wild population but will ensure a healthy, useful captive population.

- **How inbred are the individuals?** - The more inbred a individual the lower its genetic diversity and the greater the likelihood it will suffer from inbreeding depression. Inbreeding is measured by the inbreeding coefficient.

- **How related are the individuals to any that may have been previously translocated into the wild?**- This may not be relevant for the very first translocation but should be considered in any future translocations.

- **Make sure to select an even number of males and females.** - An uneven sex ratio during reintroductions has been shown to contribute to decreased genetic diversity and increased inbreeding.

Report Contents

All section titles below (in blue) are linked directly to the sections they refer to (just click on them!), any words in blue link to the glossary at the end of the report, any references in blue are hyper-linked to the relevant manuscripts URLs.

Main reports

Founder representation report - Summary of the number of founder equivalents in the current population and a bar chart which shows what proportion of the genes in the current population come from each founder (founder representation). This gives an indication of the overall genetic health of the captive population and which founders may be under-represented and therefore whose genetic contribution is in danger of being lost from the population.

Individuals recommended for reintroduction - Table with group of individuals most suitable for reintroduction based on ICYs algorithm (group of birds least related with most Fe, lowest MK). Graphs showing founder contribution of the chosen animals and how it compares with the overall founder representation with the population as a whole.

Supplementary reports

Female suitability report - Bar charts of founder representation for each living female individual in the population, this is followed by a spreadsheet showing the ranks of females based on the number of founder equivalents and mean kinship coefficient.

Male suitability report -Bar charts of founder representation for each living male individual in the population, this is followed by a spreadsheet showing the ranks of males based on the number of founder equivalents and mean kinship coefficient.

Overall suitability report- Spreadsheet of most to least suitable individuals to supplement (regardless of sex), including more details about founder representations.

Glossary & references

Glossary- Provides definitions of a few key terms (found in blue throughout the report)

Recommended reading & references - Suggestions for further reading and references that contributed to this report.

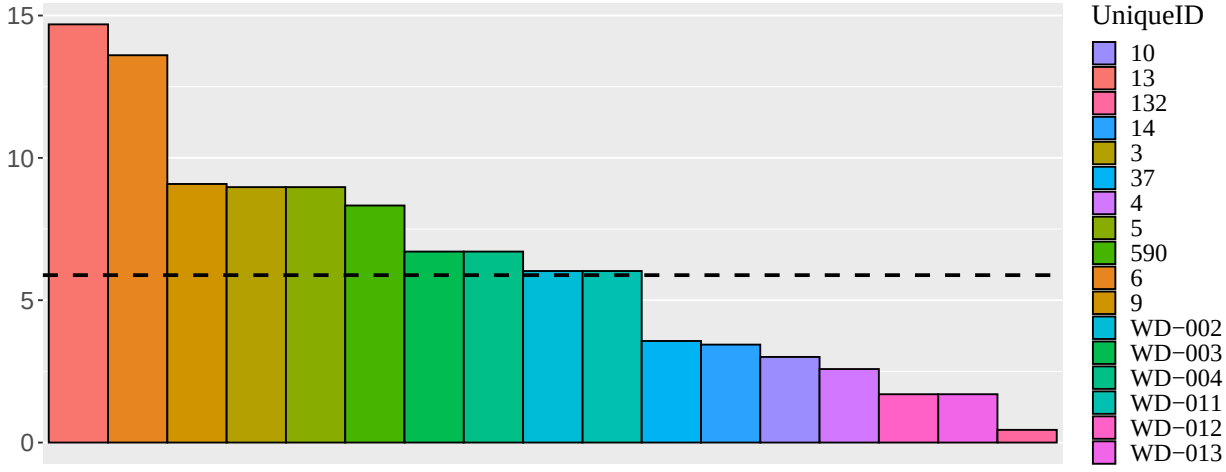
Founder representation report

Summary

The number of founders represented in the population is 17, if the founders were all represented evenly they would be present at 5.88% within the population.The number of founder equivalents found on average in the population could be increased by tactical breeding to even out representations.

Founder representation in the current population

The graph below shows how each founder is represented in the current population the dashed line represents the expected height of the bars if founder representation is equal 5.88. For example it can be seen that the highest proportion of genes in the current population belong to individual 13 . In contrast only a small proportion of the genes present in today's population come from founder 132. This may mean that the population could be in danger of losing all genes from 132 unless careful attention is paid in the following breeding seasons to maximise 132 contribution to the population. If a founders' contribution is lost from the population, that population loses genetic diversity and will likely be less healthy and robust.



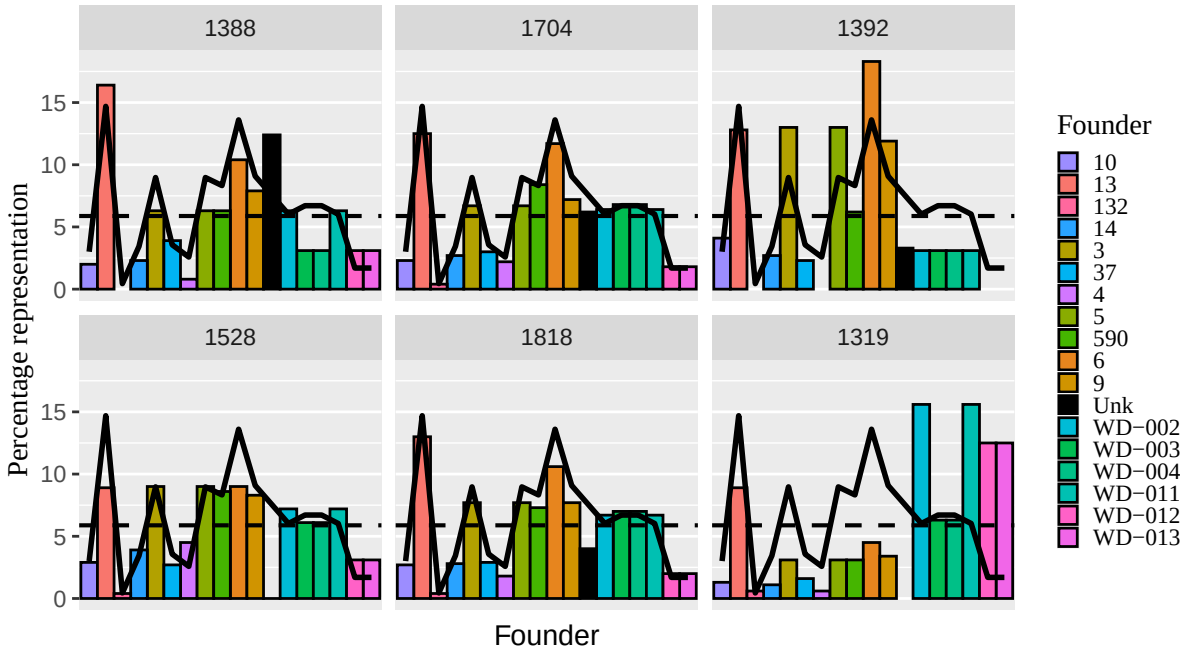
Individuals recommended for reintroduction

The table below contains the individuals that are recommended for reintroduction. These individuals are not just the highest ranked but are the highest ranked individuals that as a group of individuals are the least related. The rankings are based on the individuals mean kinship and the number of founders in their genome however it is important that the group of individuals being released are as unrelated as possible to avoid inbreeding depression and increase genetic diversity. The graphs show the founder representation of each bird (bar chart) and the line shows how the different founders are represented in the population as a whole (this the same data as represented in the founder representation bar chart on the previous page).

Table 1: Individuals recommended for reintroduction/translocation.

Rank	UniqueID	Location	Sex	F	MK	AgeYears	Number of founders	Fe
1	1388	SD-WAP	Female	0.039	0.111	13	16	14.254
3	1704	PRAHA	Female	0.068	0.144	1	17	14.030
9	1528	PRAHA	Male	0.074	0.115	7	17	14.011
17	1818	MULHOUSE	Male	0.049	0.123	0	17	13.601
96	1319	PRAHA	Male	0.016	0.077	16	17	9.760
99	1392	NY BRONX	Female	0.065	0.095	13	13	9.220

Founder representation of the individuals recommended for reintroduction (coloured bars) compared to the founder representation in current population (black line)



Female suitability report

The bar charts below shows the founder representation for each female individual in the current population. The Unique ID for a individual is at the top of each plot and the dashed line represents the expected height of the bars if founder representation is equal (5.88%). The graphs are arranged from highest ranking individual (top left) to lowest ranking individual (bottom right). The page after the graphs contains a table detailing the highest to lowest ranked female individuals (all rankings based on the number of founder equivalents (Fe) and mean kinship coefficients (MK)).The table also contains information that may be useful when selecting individuals for translocation such as age, location of zoo & inbreeding coefficient (F) .

Table 2: Information of females most to least suitable for translocation, rankings based on the number of founder equivalents (Fe) in an individual and its mean kinship (MK)

Rank	UniqueID	Location	Sex	F	MK	AgeYears	Number of founders	Fe
1	1388	SD-WAP	Female	0.039	0.111	13	16	14.254
3	1704	PRAHA	Female	0.068	0.144	1	17	14.030
6	1801	JERSEY	Female	0.068	0.144	0	17	14.030
10	1554	KOLN	Female	0.074	0.115	6	17	14.011
14	1610	FARNHAM	Female	0.074	0.136	2	17	14.011
18	1612	CHESTER	Female	0.075	0.156	2	17	13.357
20	1614	FARNHAM	Female	0.075	0.156	2	17	13.357
21	1618	KOLN	Female	0.075	0.156	2	17	13.357
22	1619	MULHOUSE	Female	0.075	0.156	2	17	13.357
25	1703	PRAHA	Female	0.075	0.156	1	17	13.357
26	1728	JERSEY	Female	0.075	0.156	1	17	13.357
27	1729	JERSEY	Female	0.075	0.156	1	17	13.357
28	1739	JERSEY	Female	0.075	0.156	1	17	13.357
30	1807	JERSEY	Female	0.075	0.156	0	17	13.357
32	1815	JERSEY	Female	0.075	0.156	0	17	13.357
40	1827	FARNHAM	Female	0.184	0.140	0	17	13.169
41	1833	FARNHAM	Female	0.184	0.140	0	17	13.169
43	1707	FARNHAM	Female	0.184	0.141	1	17	13.169
44	1839	JERSEY	Female	0.181	0.162	0	17	13.048
45	1847	JERSEY	Female	0.181	0.162	0	17	13.048
48	1803	WILDPLACE	Female	0.072	0.128	0	17	12.837
51	1853	FARNHAM	Female	0.071	0.138	0	17	12.388
55	1731	JERSEY	Female	0.204	0.169	1	17	12.016
56	1736	JERSEY	Female	0.204	0.172	1	17	12.016
58	1738	JERSEY	Female	0.204	0.174	1	17	12.016
60	1814	JERSEY	Female	0.372	0.176	0	17	12.016
65	1604	JERSEY	Female	0.093	0.158	3	17	11.945
67	1481	BURFORD	Female	0.045	0.136	10	15	11.939
70	1589	JERSEY	Female	0.065	0.170	4	17	11.921
71	1596	JERSEY	Female	0.065	0.174	4	17	11.921
72	1706	MULHOUSE	Female	0.080	0.112	1	17	11.872
74	1575	JERSEY	Female	0.349	0.117	5	15	11.058
77	1622	NY BRONX	Female	0.063	0.104	2	13	10.512
79	1441	JERSEY	Female	0.158	0.115	11	15	10.304
82	1720	WILDPLACE	Female	0.082	0.129	1	17	10.127

Table 2: Information of females most to least suitable for translocation, rankings based on the number of founder equivalents (Fe) in an individual and its mean kinship (MK) *(continued)*

Rank	UniqueID	Location	Sex	F	MK	AgeYears	Number of founders	Fe
83	1721	PAIGNTON	Female	0.082	0.129	1	17	10.127
84	1805	WILDPLACE	Female	0.082	0.129	0	17	10.127
87	1719	CHESTER	Female	0.082	0.130	1	17	10.127
88	1620	WILDPLACE	Female	0.082	0.131	2	17	10.127
89	1630	BURFORD	Female	0.082	0.131	2	17	10.127
90	1536	NY BRONX	Female	0.123	0.110	7	16	9.909
91	1588	NY BRONX	Female	0.123	0.110	5	16	9.909
92	1615	SD-WAP	Female	0.123	0.110	2	16	9.909
93	1624	SD-WAP	Female	0.123	0.110	2	16	9.909
94	1734	SD-WAP	Female	0.123	0.110	1	16	9.909
95	1746	SD-WAP	Female	0.123	0.110	1	16	9.909
98	1810	CHESTER	Female	0.107	0.116	0	17	9.295
99	1392	NY BRONX	Female	0.065	0.095	13	13	9.220
101	1496	NY BRONX	Female	0.094	0.097	9	14	8.380
102	1464	WILDPLACE	Female	0.144	0.118	10	8	7.940
103	1351	NY BRONX	Female	0.102	0.112	14	8	7.912
105	1379	NY BRONX	Female	0.092	0.091	14	8	7.077
107	1416	NY BRONX	Female	0.078	0.097	11	9	6.832
110	1203	TRACY AV	Female	0.179	0.102	20	8	5.884

Male suitability report

The bar charts below shows the founder representation for each male individual in the current population. The Unique ID for a individual is at the top of each plot and the dashed line represents the expected height of the bars if founder representation is equal (5.88%). The graphs are arranged from highest ranking individual (top left) to lowest ranking individual (bottom right). The page after the graphs contains a table detailing the highest to lowest ranked male individuals (all rankings based on the number of founder equivalents (Fe) in an individual and its mean kinship coefficients (MK)).The table also contains information that may be useful when selecting individuals for translocation such as age, location of zoo & inbreeding coefficient (F) value.

Table 3: Information of females most to least suitable for translocation, rankings based on the individual founder equivalent (Fe) and mean kinship (MK)

Rank	UniqueID	Location	Sex	F	MK	AgeYears	Number of founders	Fe
2	1456	JERSEY	Male	0.039	0.124	10	16	14.254
4	1730	PLZEN	Male	0.068	0.144	1	17	14.030
5	1800	JERSEY	Male	0.068	0.144	0	17	14.030
7	1802	JERSEY	Male	0.068	0.144	0	17	14.030
8	1735	JERSEY	Male	0.068	0.147	1	17	14.030
9	1528	PRAHA	Male	0.074	0.115	7	17	14.011
11	1534	PRAHA	Male	0.074	0.120	7	17	14.011
12	1525	WILDPLACE	Male	0.074	0.122	7	17	14.011
13	1611	BURFORD	Male	0.074	0.127	2	17	14.011
15	1552	JERSEY	Male	0.172	0.132	6	16	13.885
16	1431	BURFORD	Male	0.048	0.108	11	17	13.607
17	1818	MULHOUSE	Male	0.049	0.123	0	17	13.601
19	1613	JERSEY	Male	0.075	0.156	2	17	13.357
23	1701	JERSEY	Male	0.075	0.156	1	17	13.357
24	1702	JERSEY	Male	0.075	0.156	1	17	13.357
29	1806	JERSEY	Male	0.075	0.156	0	17	13.357
31	1811	JERSEY	Male	0.075	0.156	0	17	13.357
33	1842	JERSEY	Male	0.075	0.156	0	17	13.357
34	1708	PAIGNTON	Male	0.184	0.140	1	17	13.169
35	1709	FARNHAM	Male	0.184	0.140	1	17	13.169
36	1823	FARNHAM	Male	0.184	0.140	0	17	13.169
37	1824	FARNHAM	Male	0.184	0.140	0	17	13.169
38	1825	FARNHAM	Male	0.184	0.140	0	17	13.169
39	1826	FARNHAM	Male	0.184	0.140	0	17	13.169
42	1851	FARNHAM	Male	0.184	0.140	0	17	13.169
47	1741	PAIGNTON	Male	0.072	0.128	1	17	12.837
49	1723	BURFORD	Male	0.072	0.131	1	17	12.837
50	1724	BURFORD	Male	0.072	0.131	1	17	12.837
52	1505	CHESTER	Male	0.109	0.140	8	15	12.057
53	1705	JERSEY	Male	0.204	0.169	1	17	12.016
54	1727	JERSEY	Male	0.204	0.169	1	17	12.016
57	1737	JERSEY	Male	0.204	0.174	1	17	12.016
61	1820	JERSEY	Male	0.372	0.176	0	17	12.016
62	1821	JERSEY	Male	0.372	0.176	0	17	12.016
64	1606	JERSEY	Male	0.093	0.150	3	17	11.945

Table 3: Information of females most to least suitable for translocation, rankings based on the individual founder equivalent (Fe) and mean kinship (MK) *(continued)*

Rank	UniqueID	Location	Sex	F	MK	AgeYears	Number of founders	Fe
66	1605	JERSEY	Male	0.093	0.162	3	17	11.945
68	1454	FARNHAM	Male	0.045	0.139	10	15	11.939
69	1592	JERSEY	Male	0.065	0.158	4	17	11.921
73	1398	JERSEY	Male	0.079	0.114	13	15	11.242
75	1462	WILDPLACE	Male	0.064	0.134	10	17	10.586
76	1461	JERSEY	Male	0.064	0.145	10	17	10.586
78	1440	KOLN	Male	0.158	0.106	11	15	10.304
80	1625	MULHOUSE	Male	0.082	0.129	2	17	10.127
81	1627	WILDPLACE	Male	0.082	0.129	2	17	10.127
85	1626	MULHOUSE	Male	0.082	0.130	2	17	10.127
86	1717	FARNHAM	Male	0.082	0.130	1	17	10.127
96	1319	PRAHA	Male	0.016	0.077	16	17	9.760
97	1396	FARNHAM	Male	0.107	0.088	13	17	9.463
100	1836	NY BRONX	Male	0.065	0.096	0	13	9.220
104	1326	SD-WAP	Male	0.092	0.090	16	8	7.077
106	1312	NY BRONX	Male	0.040	0.095	17	11	6.909
108	1513	CHESTER	Male	0.078	0.097	8	9	6.832
109	1446	SD-WAP	Male	0.088	0.105	10	8	6.130

Overall suitability report

The table overleaf shows all current individuals ranked from most suitable to least suitable for translocation. The individuals are ranked (as in the previous tables) by the number of founder equivalents (Fe) and mean kinship coefficient (MK). The table below details more precise information (than the previous reports) about the founder representation in each individual in the current population.

Table 4: Information about individuals (both sexes) most to least suitable for translocation

Rank	UniqueID	Location	Sex	F	MK	Age(years)	Number of founders	Founders	Founder contribution(%)
1	1388	SD-WAP	Female	0.039	0.111	13	16	3, 4, 5, 6, 9, 10, 13, 14, 37, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.3, 0.8, 6.3, 10.4, 7.9, 2, 16.4, 2.3, 3.9, 3.1, 3.1, 6.3, 6.3, 6.3, 3.1, 3.1, 12.4
2	1456	JERSEY	Male	0.039	0.124	10	16	3, 4, 5, 6, 9, 10, 13, 14, 37, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.3, 0.8, 6.3, 10.4, 7.9, 2, 16.4, 2.3, 3.9, 3.1, 3.1, 6.3, 6.3, 6.3, 3.1, 3.1, 12.4
3	1704	PRAHA	Female	0.068	0.144	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.7, 2.2, 6.7, 11.7, 7.2, 2.3, 12.5, 2.7, 3, 0.4, 6.8, 6.8, 8.4, 6.4, 6.4, 1.8, 1.8, 6.2
4	1730	PLZEN	Male	0.068	0.144	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.7, 2.2, 6.7, 11.7, 7.2, 2.3, 12.5, 2.7, 3, 0.4, 6.8, 6.8, 8.4, 6.4, 6.4, 1.8, 1.8, 6.2
5	1800	JERSEY	Male	0.068	0.144	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.7, 2.2, 6.7, 11.7, 7.2, 2.3, 12.5, 2.7, 3, 0.4, 6.8, 6.8, 8.4, 6.4, 6.4, 1.8, 1.8, 6.2
6	1801	JERSEY	Female	0.068	0.144	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.7, 2.2, 6.7, 11.7, 7.2, 2.3, 12.5, 2.7, 3, 0.4, 6.8, 6.8, 8.4, 6.4, 6.4, 1.8, 1.8, 6.2
7	1802	JERSEY	Male	0.068	0.144	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.7, 2.2, 6.7, 11.7, 7.2, 2.3, 12.5, 2.7, 3, 0.4, 6.8, 6.8, 8.4, 6.4, 6.4, 1.8, 1.8, 6.2
8	1735	JERSEY	Male	0.068	0.147	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.7, 2.2, 6.7, 11.7, 7.2, 2.3, 12.5, 2.7, 3, 0.4, 6.8, 6.8, 8.4, 6.4, 6.4, 1.8, 1.8, 6.2
9	1528	PRAHA	Male	0.074	0.115	7	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9, 4.5, 9, 9, 8.3, 2.9, 8.9, 3.9, 2.7, 0.4, 6.1, 6.1, 8.6, 7.2, 7.2, 3.1, 3.1, 0
10	1554	KOLN	Female	0.074	0.115	6	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9, 4.5, 9, 9, 8.3, 2.9, 8.9, 3.9, 2.7, 0.4, 6.1, 6.1, 8.6, 7.2, 7.2, 3.1, 3.1, 0
11	1534	PRAHA	Male	0.074	0.120	7	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9, 4.5, 9, 9, 8.3, 2.9, 8.9, 3.9, 2.7, 0.4, 6.1, 6.1, 8.6, 7.2, 7.2, 3.1, 3.1, 0
12	1525	WILDPLACE	Male	0.074	0.122	7	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9, 4.5, 9, 9, 8.3, 2.9, 8.9, 3.9, 2.7, 0.4, 6.1, 6.1, 8.6, 7.2, 7.2, 3.1, 3.1, 0
13	1611	BURFORD	Male	0.074	0.127	2	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9, 4.5, 9, 9, 8.3, 2.9, 8.9, 3.9, 2.7, 0.4, 6.1, 6.1, 8.6, 7.2, 7.2, 3.1, 3.1, 0
14	1610	FARNHAM	Female	0.074	0.136	2	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9, 4.5, 9, 9, 8.3, 2.9, 8.9, 3.9, 2.7, 0.4, 6.1, 6.1, 8.6, 7.2, 7.2, 3.1, 3.1, 0
15	1552	JERSEY	Male	0.172	0.132	6	16	3, 4, 5, 6, 9, 10, 13, 14, 37, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	5.4, 1.2, 5.4, 12.8, 6.4, 1.6, 9.4, 2, 2.7, 4.7, 4.7, 9.4, 9.4, 9.4, 4.7, 4.7, 6.1

Table 4: Information about individuals (both sexes) most to least suitable for translocation *(continued)*

Rank	UniqueID	Location	Sex	F	MK	Age(years)	Number of founders	Founders	Founder contribution(%)
16	1431	BURFORD	Male	0.048	0.108	11	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	7.1, 3, 7.1, 6, 7, 2.5, 8.2, 2.6, 1.6, 0.3, 6.3, 6.3, 7.8, 10.9, 10.9, 6.3, 6.3
17	1818	MULHOUSE	Male	0.049	0.123	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.7, 1.8, 7.7, 10.6, 7.7, 2.7, 13, 2.8, 2.9, 0.4, 7, 7, 7.3, 6.7, 6.7, 2, 2, 4
18	1612	CHESTER	Female	0.075	0.156	2	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
19	1613	JERSEY	Male	0.075	0.156	2	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
20	1614	FARNHAM	Female	0.075	0.156	2	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
21	1618	KOLN	Female	0.075	0.156	2	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
22	1619	MULHOUSE	Female	0.075	0.156	2	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
23	1701	JERSEY	Male	0.075	0.156	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
24	1702	JERSEY	Male	0.075	0.156	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
25	1703	PRAHA	Female	0.075	0.156	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
26	1728	JERSEY	Female	0.075	0.156	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
27	1729	JERSEY	Female	0.075	0.156	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
28	1739	JERSEY	Female	0.075	0.156	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
29	1806	JERSEY	Male	0.075	0.156	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
30	1807	JERSEY	Female	0.075	0.156	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3

Table 4: Information about individuals (both sexes) most to least suitable for translocation *(continued)*

Rank	UniqueID	Location	Sex	F	MK	Age(years)	Number of founders	Founders	Founder contribution(%)
31	1811	JERSEY	Male	0.075	0.156	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
32	1815	JERSEY	Female	0.075	0.156	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
33	1842	JERSEY	Male	0.075	0.156	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.4, 1.8, 6.4, 12.1, 6.9, 2.1, 11.2, 2.4, 2.8, 0.3, 6.7, 6.7, 10.3, 7.6, 7.6, 2.7, 2.7, 3.3
34	1708	PAIGNTON	Male	0.184	0.140	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	9.9, 5.2, 9.9, 10.6, 8.9, 3.1, 9.3, 4.6, 3.3, 0.4, 6, 6, 9, 5.4, 5.4, 1.6, 1.6
35	1709	FARNHAM	Male	0.184	0.140	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	9.9, 5.2, 9.9, 10.6, 8.9, 3.1, 9.3, 4.6, 3.3, 0.4, 6, 6, 9, 5.4, 5.4, 1.6, 1.6
36	1823	FARNHAM	Male	0.184	0.140	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	9.9, 5.2, 9.9, 10.6, 8.9, 3.1, 9.3, 4.6, 3.3, 0.4, 6, 6, 9, 5.4, 5.4, 1.6, 1.6
37	1824	FARNHAM	Male	0.184	0.140	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	9.9, 5.2, 9.9, 10.6, 8.9, 3.1, 9.3, 4.6, 3.3, 0.4, 6, 6, 9, 5.4, 5.4, 1.6, 1.6
38	1825	FARNHAM	Male	0.184	0.140	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	9.9, 5.2, 9.9, 10.6, 8.9, 3.1, 9.3, 4.6, 3.3, 0.4, 6, 6, 9, 5.4, 5.4, 1.6, 1.6
39	1826	FARNHAM	Male	0.184	0.140	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	9.9, 5.2, 9.9, 10.6, 8.9, 3.1, 9.3, 4.6, 3.3, 0.4, 6, 6, 9, 5.4, 5.4, 1.6, 1.6
40	1827	FARNHAM	Female	0.184	0.140	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	9.9, 5.2, 9.9, 10.6, 8.9, 3.1, 9.3, 4.6, 3.3, 0.4, 6, 6, 9, 5.4, 5.4, 1.6, 1.6
41	1833	FARNHAM	Female	0.184	0.140	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	9.9, 5.2, 9.9, 10.6, 8.9, 3.1, 9.3, 4.6, 3.3, 0.4, 6, 6, 9, 5.4, 5.4, 1.6, 1.6
42	1851	FARNHAM	Male	0.184	0.140	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	9.9, 5.2, 9.9, 10.6, 8.9, 3.1, 9.3, 4.6, 3.3, 0.4, 6, 6, 9, 5.4, 5.4, 1.6, 1.6
43	1707	FARNHAM	Female	0.184	0.141	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	9.9, 5.2, 9.9, 10.6, 8.9, 3.1, 9.3, 4.6, 3.3, 0.4, 6, 6, 9, 5.4, 5.4, 1.6, 1.6
44	1839	JERSEY	Female	0.181	0.162	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.9, 2.6, 6.9, 12, 7.1, 2.5, 11.7, 2.8, 2.8, 0.6, 8.2, 8.2, 9.7, 6.3, 6.3, 1.2, 1.2, 3
45	1847	JERSEY	Female	0.181	0.162	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.9, 2.6, 6.9, 12, 7.1, 2.5, 11.7, 2.8, 2.8, 0.6, 8.2, 8.2, 9.7, 6.3, 6.3, 1.2, 1.2, 3

Table 4: Information about individuals (both sexes) most to least suitable for translocation *(continued)*

Rank	UniqueID	Location	Sex	F	MK	Age(years)	Number of founders	Founders	Founder contribution(%)
46	1848	JERSEY	Unknown	0.181	0.162	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	6.9, 2.6, 6.9, 12, 7.1, 2.5, 11.7, 2.8, 2.8, 0.6, 8.2, 8.2, 9.7, 6.3, 6.3, 1.2, 1.2, 3
47	1741	PAIGNTON	Male	0.072	0.128	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.3, 2.4, 9.3, 8.7, 9.1, 3.1, 15.9, 3.6, 3.7, 0.4, 5.2, 5.2, 6.2, 5.1, 5.1, 2, 2, 3.7
48	1803	WILDPLACE	Female	0.072	0.128	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.3, 2.4, 9.3, 8.7, 9.1, 3.1, 15.9, 3.6, 3.7, 0.4, 5.2, 5.2, 6.2, 5.1, 5.1, 2, 2, 3.7
49	1723	BURFORD	Male	0.072	0.131	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.3, 2.4, 9.3, 8.7, 9.1, 3.1, 15.9, 3.6, 3.7, 0.4, 5.2, 5.2, 6.2, 5.1, 5.1, 2, 2, 3.7
50	1724	BURFORD	Male	0.072	0.131	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.3, 2.4, 9.3, 8.7, 9.1, 3.1, 15.9, 3.6, 3.7, 0.4, 5.2, 5.2, 6.2, 5.1, 5.1, 2, 2, 3.7
51	1853	FARNHAM	Female	0.071	0.138	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.7, 2.8, 9.7, 9.5, 9.5, 3.2, 16.1, 3.9, 4, 0.4, 5.1, 5.1, 6.3, 4.2, 4.2, 1.2, 1.2, 3.9
52	1505	CHESTER	Male	0.109	0.140	8	15	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002	11, 5.7, 11, 9.8, 10.1, 3.5, 8.6, 4.7, 2.7, 0.2, 6.1, 6.1, 10.9, 4.9, 4.9
53	1705	JERSEY	Male	0.204	0.169	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.2, 3, 7.2, 12.3, 7, 2.6, 10.8, 2.9, 2.5, 0.7, 9.6, 9.6, 10.9, 6.2, 6.2, 0.6, 0.6, 0.1
54	1727	JERSEY	Male	0.204	0.169	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.2, 3, 7.2, 12.3, 7, 2.6, 10.8, 2.9, 2.5, 0.7, 9.6, 9.6, 10.9, 6.2, 6.2, 0.6, 0.6, 0.1
55	1731	JERSEY	Female	0.204	0.169	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.2, 3, 7.2, 12.3, 7, 2.6, 10.8, 2.9, 2.5, 0.7, 9.6, 9.6, 10.9, 6.2, 6.2, 0.6, 0.6, 0.1
56	1736	JERSEY	Female	0.204	0.172	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.2, 3, 7.2, 12.3, 7, 2.6, 10.8, 2.9, 2.5, 0.7, 9.6, 9.6, 10.9, 6.2, 6.2, 0.6, 0.6, 0.1
57	1737	JERSEY	Male	0.204	0.174	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.2, 3, 7.2, 12.3, 7, 2.6, 10.8, 2.9, 2.5, 0.7, 9.6, 9.6, 10.9, 6.2, 6.2, 0.6, 0.6, 0.1
58	1738	JERSEY	Female	0.204	0.174	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.2, 3, 7.2, 12.3, 7, 2.6, 10.8, 2.9, 2.5, 0.7, 9.6, 9.6, 10.9, 6.2, 6.2, 0.6, 0.6, 0.1
59	1813	GOVT MAUR	Unknown	0.372	0.176	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.2, 3, 7.2, 12.3, 7, 2.6, 10.8, 2.9, 2.5, 0.7, 9.6, 9.6, 10.9, 6.2, 6.2, 0.6, 0.6, 0.1
60	1814	JERSEY	Female	0.372	0.176	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.2, 3, 7.2, 12.3, 7, 2.6, 10.8, 2.9, 2.5, 0.7, 9.6, 9.6, 10.9, 6.2, 6.2, 0.6, 0.6, 0.1

Table 4: Information about individuals (both sexes) most to least suitable for translocation *(continued)*

Rank	UniqueID	Location	Sex	F	MK	Age(years)	Number of founders	Founders	Founder contribution(%)
61	1820	JERSEY	Male	0.372	0.176	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.2, 3, 7.2, 12.3, 7, 2.6, 10.8, 2.9, 2.5, 0.7, 9.6, 9.6, 10.9, 6.2, 6.2, 0.6, 0.6, 0.1
62	1821	JERSEY	Male	0.372	0.176	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.2, 3, 7.2, 12.3, 7, 2.6, 10.8, 2.9, 2.5, 0.7, 9.6, 9.6, 10.9, 6.2, 6.2, 0.6, 0.6, 0.1
63	1840	JERSEY	Unknown	0.372	0.176	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.2, 3, 7.2, 12.3, 7, 2.6, 10.8, 2.9, 2.5, 0.7, 9.6, 9.6, 10.9, 6.2, 6.2, 0.6, 0.6, 0.1
64	1606	JERSEY	Male	0.093	0.150	3	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7, 3.5, 7, 13, 6.5, 2.6, 8.6, 3.1, 2.1, 0.8, 10.5, 10.5, 10.6, 6.6, 6.6, 0.4, 0.4, 0.2
65	1604	JERSEY	Female	0.093	0.158	3	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7, 3.5, 7, 13, 6.5, 2.6, 8.6, 3.1, 2.1, 0.8, 10.5, 10.5, 10.6, 6.6, 6.6, 0.4, 0.4, 0.2
66	1605	JERSEY	Male	0.093	0.162	3	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7, 3.5, 7, 13, 6.5, 2.6, 8.6, 3.1, 2.1, 0.8, 10.5, 10.5, 10.6, 6.6, 6.6, 0.4, 0.4, 0.2
67	1481	BURFORD	Female	0.045	0.136	10	15	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002	10.8, 5.9, 10.8, 12.1, 9.6, 3.3, 9.7, 5.3, 3.9, 0.4, 5.9, 5.9, 9.4, 3.5, 3.5
68	1454	FARNHAM	Male	0.045	0.139	10	15	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002	10.8, 5.9, 10.8, 12.1, 9.6, 3.3, 9.7, 5.3, 3.9, 0.4, 5.9, 5.9, 9.4, 3.5, 3.5
69	1592	JERSEY	Male	0.065	0.158	4	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.3, 2.4, 7.3, 11.5, 7.5, 2.6, 13, 2.8, 2.9, 0.7, 8.7, 8.7, 11.1, 5.9, 5.9, 0.8, 0.8, 0.1
70	1589	JERSEY	Female	0.065	0.170	4	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.3, 2.4, 7.3, 11.5, 7.5, 2.6, 13, 2.8, 2.9, 0.7, 8.7, 8.7, 11.1, 5.9, 5.9, 0.8, 0.8, 0.1
71	1596	JERSEY	Female	0.065	0.174	4	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.3, 2.4, 7.3, 11.5, 7.5, 2.6, 13, 2.8, 2.9, 0.7, 8.7, 8.7, 11.1, 5.9, 5.9, 0.8, 0.8, 0.1
72	1706	MULHOUSE	Female	0.080	0.112	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	5.8, 3.2, 5.8, 12.8, 5.3, 2, 3.3, 2.4, 1.2, 0.4, 9.8, 9.8, 10.9, 10.5, 10.5, 3.1, 3.1, 0.1
73	1398	JERSEY	Male	0.079	0.114	13	15	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002	7, 4.8, 7, 10.4, 5.7, 2.9, 4.2, 3.2, 0.8, 0.9, 13.3, 13.3, 9.4, 8.6, 8.6
74	1575	JERSEY	Female	0.349	0.117	5	15	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002	6.7, 4.7, 6.7, 14.6, 5.5, 2.7, 4.2, 3.3, 1.4, 1, 12.3, 12.3, 10.2, 7.2, 7.2
75	1462	WILDPLACE	Male	0.064	0.134	10	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	6.3, 0.8, 6.3, 10.7, 7.4, 2.5, 19.8, 2.3, 3.7, 0.8, 8.6, 8.6, 7.4, 5.9, 5.9, 1.6, 1.6

Table 4: Information about individuals (both sexes) most to least suitable for translocation *(continued)*

Rank	UniqueID	Location	Sex	F	MK	Age(years)	Number of founders	Founders	Founder contribution(%)
76	1461	JERSEY	Male	0.064	0.145	10	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	6.3, 0.8, 6.3, 10.7, 7.4, 2.5, 19.8, 2.3, 3.7, 0.8, 8.6, 8.6, 7.4, 5.9, 5.9, 1.6, 1.6
77	1622	NY BRONX	Female	0.063	0.104	2	13	3, 5, 6, 9, 10, 13, 14, 37, WD-003, WD-004, 590, WD-011, WD-002, Unk	8.1, 8.1, 17.6, 10.2, 2.5, 17.6, 1.6, 3.1, 3.1, 3.1, 6.3, 3.1, 3.1, 12.5
78	1440	KOLN	Male	0.158	0.106	11	15	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002	6.5, 4.5, 6.5, 18.8, 5.3, 2.5, 4.2, 3.4, 2, 1, 11.3, 11.3, 10.9, 5.9, 5.9
79	1441	JERSEY	Female	0.158	0.115	11	15	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002	6.5, 4.5, 6.5, 18.8, 5.3, 2.5, 4.2, 3.4, 2, 1, 11.3, 11.3, 10.9, 5.9, 5.9
80	1625	MULHOUSE	Male	0.082	0.129	2	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.6, 0.4, 9.6, 8.4, 10, 3.4, 22.8, 3.3, 4.6, 0.4, 4.3, 4.3, 3.7, 2.9, 2.9, 0.8, 0.8, 7.8
81	1627	WILDPLACE	Male	0.082	0.129	2	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.6, 0.4, 9.6, 8.4, 10, 3.4, 22.8, 3.3, 4.6, 0.4, 4.3, 4.3, 3.7, 2.9, 2.9, 0.8, 0.8, 7.8
82	1720	WILDPLACE	Female	0.082	0.129	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.6, 0.4, 9.6, 8.4, 10, 3.4, 22.8, 3.3, 4.6, 0.4, 4.3, 4.3, 3.7, 2.9, 2.9, 0.8, 0.8, 7.8
83	1721	PAIGNTON	Female	0.082	0.129	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.6, 0.4, 9.6, 8.4, 10, 3.4, 22.8, 3.3, 4.6, 0.4, 4.3, 4.3, 3.7, 2.9, 2.9, 0.8, 0.8, 7.8
84	1805	WILDPLACE	Female	0.082	0.129	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.6, 0.4, 9.6, 8.4, 10, 3.4, 22.8, 3.3, 4.6, 0.4, 4.3, 4.3, 3.7, 2.9, 2.9, 0.8, 0.8, 7.8
85	1626	MULHOUSE	Male	0.082	0.130	2	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.6, 0.4, 9.6, 8.4, 10, 3.4, 22.8, 3.3, 4.6, 0.4, 4.3, 4.3, 3.7, 2.9, 2.9, 0.8, 0.8, 7.8
86	1717	FARNHAM	Male	0.082	0.130	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.6, 0.4, 9.6, 8.4, 10, 3.4, 22.8, 3.3, 4.6, 0.4, 4.3, 4.3, 3.7, 2.9, 2.9, 0.8, 0.8, 7.8
87	1719	CHESTER	Female	0.082	0.130	1	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.6, 0.4, 9.6, 8.4, 10, 3.4, 22.8, 3.3, 4.6, 0.4, 4.3, 4.3, 3.7, 2.9, 2.9, 0.8, 0.8, 7.8
88	1620	WILDPLACE	Female	0.082	0.131	2	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.6, 0.4, 9.6, 8.4, 10, 3.4, 22.8, 3.3, 4.6, 0.4, 4.3, 4.3, 3.7, 2.9, 2.9, 0.8, 0.8, 7.8
89	1630	BURFORD	Female	0.082	0.131	2	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	9.6, 0.4, 9.6, 8.4, 10, 3.4, 22.8, 3.3, 4.6, 0.4, 4.3, 4.3, 3.7, 2.9, 2.9, 0.8, 0.8, 7.8
90	1536	NY BRONX	Female	0.123	0.110	7	16	3, 4, 5, 6, 9, 10, 13, 14, 37, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.9, 0.4, 7.9, 20.6, 9.8, 2.4, 17.4, 2, 3.5, 1.6, 1.6, 3.1, 3.1, 3.1, 1.6, 1.6, 12.4

Table 4: Information about individuals (both sexes) most to least suitable for translocation *(continued)*

Rank	UniqueID	Location	Sex	F	MK	Age(years)	Number of founders	Founders	Founder contribution(%)
91	1588	NY BRONX	Female	0.123	0.110	5	16	3, 4, 5, 6, 9, 10, 13, 14, 37, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.9, 0.4, 7.9, 20.6, 9.8, 2.4, 17.4, 2, 3.5, 1.6, 1.6, 3.1, 3.1, 3.1, 1.6, 1.6, 12.4
92	1615	SD-WAP	Female	0.123	0.110	2	16	3, 4, 5, 6, 9, 10, 13, 14, 37, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.9, 0.4, 7.9, 20.6, 9.8, 2.4, 17.4, 2, 3.5, 1.6, 1.6, 3.1, 3.1, 3.1, 1.6, 1.6, 12.4
93	1624	SD-WAP	Female	0.123	0.110	2	16	3, 4, 5, 6, 9, 10, 13, 14, 37, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.9, 0.4, 7.9, 20.6, 9.8, 2.4, 17.4, 2, 3.5, 1.6, 1.6, 3.1, 3.1, 3.1, 1.6, 1.6, 12.4
94	1734	SD-WAP	Female	0.123	0.110	1	16	3, 4, 5, 6, 9, 10, 13, 14, 37, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.9, 0.4, 7.9, 20.6, 9.8, 2.4, 17.4, 2, 3.5, 1.6, 1.6, 3.1, 3.1, 3.1, 1.6, 1.6, 12.4
95	1746	SD-WAP	Female	0.123	0.110	1	16	3, 4, 5, 6, 9, 10, 13, 14, 37, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	7.9, 0.4, 7.9, 20.6, 9.8, 2.4, 17.4, 2, 3.5, 1.6, 1.6, 3.1, 3.1, 3.1, 1.6, 1.6, 12.4
96	1319	PRAHA	Male	0.016	0.077	16	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	3.1, 0.6, 3.1, 4.5, 3.4, 1.3, 8.9, 1.1, 1.6, 0.6, 6.3, 6.3, 3.1, 15.6, 15.6, 12.5, 12.5
97	1396	FARNHAM	Male	0.107	0.088	13	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012	4.6, 1.6, 4.6, 3.4, 3.8, 2.3, 9.2, 1.4, 1.2, 1.6, 17.2, 17.2, 2.3, 11.7, 11.7, 3.1, 3.1
98	1810	CHESTER	Female	0.107	0.116	0	17	3, 4, 5, 6, 9, 10, 13, 14, 37, 132, WD-003, WD-004, 590, WD-011, WD-002, WD-013, WD-012, Unk	8.7, 1.8, 8.7, 19.2, 9.5, 2.8, 19.8, 4, 5.4, 0.2, 2.1, 2.1, 1.9, 1.5, 1.5, 0.4, 0.4, 10
99	1392	NY BRONX	Female	0.065	0.095	13	13	3, 5, 6, 9, 10, 13, 14, 37, WD-003, WD-004, 590, WD-011, WD-002, Unk	13, 13, 18.3, 11.9, 4.1, 12.8, 2.7, 2.3, 3.1, 3.1, 6.2, 3.1, 3.1, 3.3
100	1836	NY BRONX	Male	0.065	0.096	0	13	3, 5, 6, 9, 10, 13, 14, 37, WD-003, WD-004, 590, WD-011, WD-002, Unk	13, 13, 18.3, 11.9, 4.1, 12.8, 2.7, 2.3, 3.1, 3.1, 6.2, 3.1, 3.1, 3.3
101	1496	NY BRONX	Female	0.094	0.097	9	14	3, 4, 5, 6, 9, 10, 13, 14, 37, WD-003, WD-004, 590, WD-011, WD-002, Unk	10.4, 1.6, 10.4, 24.2, 10.4, 3.1, 14.8, 3.7, 4.3, 1.6, 1.6, 3.1, 1.6, 1.6, 7.6
102	1464	WILDPLACE	Female	0.144	0.118	10	8	3, 5, 6, 9, 10, 13, 14, 37, Unk	12.9, 12.9, 6.2, 12.7, 4.2, 25.7, 4.3, 5.5, 15.6
103	1351	NY BRONX	Female	0.102	0.112	14	8	3, 5, 6, 9, 10, 13, 14, 37, Unk	8, 8, 5.5, 10.9, 2.7, 30.5, 3.1, 6.3, 25
104	1326	SD-WAP	Male	0.092	0.090	16	8	3, 5, 6, 9, 10, 13, 14, 37, Unk	17.8, 17.8, 6.8, 14.5, 5.8, 20.9, 5.5, 4.7, 6.2
105	1379	NY BRONX	Female	0.092	0.091	14	8	3, 5, 6, 9, 10, 13, 14, 37, Unk	17.8, 17.8, 6.8, 14.5, 5.8, 20.9, 5.5, 4.7, 6.2
106	1312	NY BRONX	Male	0.040	0.095	17	11	3, 5, 6, 9, 10, 13, WD-003, WD-004, 590, WD-011, WD-002	8.2, 8.2, 29.7, 9.4, 2.3, 4.7, 6.3, 6.3, 12.5, 6.3, 6.3
107	1416	NY BRONX	Female	0.078	0.097	11	9	3, 4, 5, 6, 9, 10, 13, 14, 37, Unk	7.7, 3.1, 7.7, 30.1, 9, 2.1, 16.8, 4.7, 6.3, 12.5

Table 4: Information about individuals (both sexes) most to least suitable for translocation *(continued)*

Rank	UniqueID	Location	Sex	F	MK	Age(years)	Number of founders	Founders	Founder contribution(%)
108	1513	CHESTER	Male	0.078	0.097	8	9	3, 4, 5, 6, 9, 10, 13, 14, 37, Unk	7.7, 3.1, 7.7, 30.1, 9, 2.1, 16.8, 4.7, 6.3, 12.5
109	1446	SD-WAP	Male	0.088	0.105	10	8	3, 5, 6, 9, 10, 13, 14, 37, Unk	9.5, 9.5, 30.9, 11.7, 2.9, 18.4, 1.6, 3.1, 12.4
110	1203	TRACY AV	Female	0.179	0.102	20	8	3, 5, 6, 9, 10, 13, 14, 37, Unk	9.1, 9.1, 7.4, 10.2, 2.1, 35.5, 4.7, 9.4, 12.5

Glossary

Founder equivalents

- A measure of genetic diversity and health. It calculates the number of founder genomes present in the current population or an individual and their relative representations. If all founders are represented equally the number of founder equivalents equal the number of founders.

Founder representation

- The proportion of each founders' genes found in an individual in the population

Genome

- All the genes in an individual

Inbreeding coefficient

- Measure of inbreeding denoted by 'F', the higher the inbreeding coefficient the more inbred the individuals are. This can be thought of as how much of an individuals genome is is likely to share with other individuals in the population. When individuals are highly inbred they can experience inbreeding depression which can manifest as reduced survival, fecundity or an increase in health problems.

Mean kinship

- Measure of relatedness, denoted in tables by 'MK'. The mean kinship is on average how related a individual is to the other individuals in the population, the higher the MK the higher the average relatedness. The mean kinship coefficient = mean inbreeding coefficient of the hypothetical offspring between the individual individual and any other in the population.

Recommended reading & references

Genetics in management

Frankham, R., Jonathan, B. and David, B. 2012. Introduction To Conservation Genetics. 2nd ed. New York: Cambridge University Press.

Grueber, C.E. and Jamieson, I.G. 2008. Quantifying and managing the loss of genetic variation in a free-ranging population of takahe through the use of pedigrees. Conservation genetics (Print) 9(3), pp. 645–651.

Jamieson, I.G. 2011. Founder effects, inbreeding, and loss of genetic diversity in four avian reintroduction programs. Conservation Biology 25(1), pp. 115–123.

Loercher, F., Keller, L. and Hegglin, D. 2013. Low genetic diversity of the reintroduced bearded vulture (*Gypaetus barbatus*) population in the Alps calls for further releases. In: pp. 473–478.

Founder Equivalents

The equation that was used to calculate the number of founder equivalents comes originally from Lacy, 1989 and is as follows:

$Fe = 1/\sum(pi^2)$

Where *Fe* is the number of founder equivalents and *pi* is a founders representation.

This translates to:

An individuals founder equivalents = 1 divided by the sum of all the founder representations squared

Founder contributions were generated from studbook data automatically in Pmx (version 1.4.2).

Hedrick, P.W., Hoeck, P.E.A., Fleischer, R.C., Farabaugh, S. and Masuda, B.M. 2016. The influence of captive breeding management on founder representation and inbreeding in the ʻAlalā, the Hawaiian crow. Conservation genetics (Print) 17(2), pp. 369–378.

Lacy, R. 1989. Analysis of Founder Representation in Pedigrees: Founder Equivalents and founder equivalents. Zoo Biology 8, pp. 111–123.

Software used to analyse studbook data (free)

Ballou, J., Lacy, R. and Pollak, J. 2010. PMx: software for demographic and genetic analysis and management of pedigreed populations. Brookfield, Illinois, USA.: Chicago Zoological Society.