This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Cavill, Emily L

Title:
Conservation genomics of the endangered Seychelles magpie robin (Copsychus sechellarum)

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

• Your contact details
• Bibliographic details for the item, including a URL
• An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Conservation Genomics of the Endangered Seychelles Magpie Robin (*Copsychus sechellarum*)

Emily Louisa Cavill

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of MSc by Research in the Faculty of Life Sciences.

School of Biological Sciences *May 2019*

Word Count: 16,007

// Abstract

‘Threatened’ species often exist as small, isolated populations, such as those on islands. Characteristic consequences of this are inbreeding – the mating of closely related individuals – which can result in reduced heterozygosity and increased risk of inbreeding depression. These species sometimes undergo ‘rescue’, whereby well–intentioned parties translocate individuals from a healthy (or sometimes the last remaining) population to a new destination to help the species recover. Repeated translocations resulting in repeated founder effects can potentially lead to a decrease in genetic diversity in the new populations. Translocations have been an integral factor in the recovery of the Seychelles magpie robin (*Copsychus sechellarum*), an endangered species endemic to Seychelles, but the potential genetic consequences of their translocation history have not yet been explored.

For this project, 110 Seychelles magpie robin genomes, representing the five islands within the Seychelles Archipelago on which the species currently exists, were re–sequenced and mapped against a reference genome which was constructed *de novo* for the species. Mapping the re–sequenced genomes against the species–specific reference allowed for identification of the genetic variation in the form of single nucleotide polymorphisms (SNPs). These SNPs were used to identify patterns of heterozygosity, homozygosity and autozygosity in each individual to investigate inbreeding and genetic diversity across the islands.

With minimal genetic research previously undertaken on the Seychelles magpie robin, this project offered the first analysis of the genetic profile of this endangered species. A very low level of heterozygosity was observed, coupled with long homozygous segments that suggest recent inbreeding, probably a consequence of the most recent population bottleneck experienced. Three of the four translocated populations displayed less genetic diversity than the founder population from which they were taken – the familiar pattern observed as a result of the evolutionary force of genetic drift following founder events. Furthermore, and perhaps surprising given the recent time since the new populations were established, population structure was observed among translocated populations.

New awareness of this inbreeding and continued monitoring of the population will allow for genetically informed management decisions, particularly concerning the future translocations planned for this species.

Dedication & Acknowledgements

First and foremost, I would like to thank SMART, with my whole heart, for giving me the incredible opportunity to help save a bird I fell so deeply in love with during my time working on Cousin Island. I want to thank Cheryl Sanchez for the encouragement you have always given me to pursue this research – you are, and will always be, an inspiration to me. I also would like to thank Bronwyn and Stuart Dunlop on Cousine Island for training me to withdraw blood from a magpie robin – quite possibly my fondest memory of Seychelles! I look forward to sequencing it! Eric Blais, none of this would be possible without your facilitation of getting the samples from the islands in Seychelles to the labs in Copenhagen, and the endless permits we needed to get them here! A big part of this work was also made possible by the long–term care and generous handover of Aride samples by Kate Lessells and Christine Mateman at NIOO – you helped us answer important questions for this species.

Of course, this project would not have been possible without my two supervisors at Bristol University who first had faith in my research ideas and ability to undertake this work – so for that, Gareth Jones and Martin Genner, I am grateful.

Tom Gilbert – I feel so privileged to have worked as part of your team over the past two years. Thank you for having confidence in me, for letting me explore the potential of this research, and for investing so many resources in my work and the magpies... we are *all* truly grateful.

Every single member of Tom's research team deserves a special thank you for making me feel so comfortable during my visits to Copenhagen University. I turned up in a new country, to a new institution, feeling a bit lost and out of my depth trying to understand a new subject (and then the subsequent two years where I just turned up overwhelmed for no good reason) and you made me feel so welcome every single visit. Equally I'd like to thank everyone in the bat lab at Bristol University, especially Parvathy Venugopal, Lia Gilmour and Penelope Fialas – those last few weeks were gruelling, but your support and kind words helped me get to the end!

I need to thank all lab staff, especially Tina Brand, Pernille Olsen and Charlotte Hansen for orders and printing and generally making all things 'lab' possible. I am so grateful to George Pacheco and Lara Puetz for taking the time around your own PhDs to train me so thoroughly in all lab protocols, Ángela

Ribeiro for helping me work through the numerous problems that came with the task of sexing the magpie robins, and Sarah Mak for being the lab god to us all.

Goodness. Bioinformatics. Shyam Gopalakrishnan, I actually cannot thank you enough for being my teacher, for guiding me, and for enduring my endless questions. I can confidently say that without you I would not have got this far. There are so many people who helped me with ideas, research, support and feedback for this dissertation. I value everybody's contributions, no matter how big or small, as without you all this project would not be what it is.

Having a magpie robin annotated reference genome is an amazing achievement, especially for a Masters project. So, I extend a huge thanks to Love Dalen, Bent Petersen, Guojie Zhang, Josefin Stiller and the team at the Genome B10K Consortium for creating the reference genome for this species that was invaluable to my research.

I am especially grateful to ERC for the funding that made this whole project possible. And I would like to extend this thanks to everyone who donated to my 'Save A Species' campaign run through the fundraising platform 'Experiment'. Although the project surpassed what I could have ever expected to achieve, your donations helped get this whole project off the ground.

Rebeca Orts Jover, your artistic contribution really made me feel like this thesis was finally complete. Thank you for creating such a beautiful and unique representation of the Seychelles magpie robin, and for allowing me the privilege to display it so proudly on this thesis.

Lastly, but by no means least, I would like to thank Rute Fonseca and Ida Moltke for not only offering their expertise and guidance, but for becoming my role models as strong and empowering female figures in this field.

University of
BRISTOL

UNIVERSITY OF
COPENHAGEN

Dedicated to my best pal Amy – who has no background in science but who, after her daily support throughout my Masters, could probably write an exceptional genetics paper on the magpie robins. I cannot believe after all these years of friendship the only photos we have together are of our trumpets and our feet.

Author’s declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... DATE: ....

Table of Contents

Abstract 2

Dedication & Acknowledgements 3

Author's declaration........ 6

Chapter 1 12

General Introduction 12

1.1 A Brief History of the Seychelles Magpie Robin........ 13

1.2 Why Save the Magpie Robins? 16

1.3 Genetics as a Tool for Conservation........ 17

1.3.1 The Significance of SNPs 18

1.3.2 Gene Flow & Mutation 18

1.3.3 Genetic Drift & Heterozygosity 19

1.3.4 Linkage Disequilibrium & Haplotype Blocks........ 21

1.4 Conservation in the Era of Genomics........ 21

1.5 Project Aims & Contributions........ 22

1.6 Thesis Overview........ 23

Chapter 2 24

Genome Sequencing – A Unique Insight into The Recent History of a Precious Endemic Bird 24

2.1 Introduction........ 25

2.1.1 The Seychelles Magpie Robin........ 25

2.1.2 Next Generation Sequencing........ 25

2.1.3 Why Sequence Whole Genomes?........ 27

2.1.4 The Avian Genome 28

2.2 Methods 29

2.2.1 Sampling 29

2.2.2 DNA Extraction........ 29

2.2.3 Molecular Sexing.. 30

2.2.4 Nucleic Acid Fragmentation.. 31

2.2.5 Library Construction.. 31

2.2.6 Sequencing.. 31

2.2.7 Reference Genome Assembly.. 32

2.2.8 Read Alignment & Quality Control.. 32

2.2.9 Analysis.. 33

2.3 Results.. 42

2.3.1 DNA extraction & Library Construction.. 42

2.3.2 Molecular Sexing.. 42

2.3.3 Reference Genome Assembly.. 43

2.3.4 Sequencing Success, Failures, & Quality Control.. 43

2.3.5 Principal Component Analysis.. 45

2.3.6 Admixture.. 47

2.3.7 Heterozygosity.. 48

2.3.8 The Site Frequency Spectrum.. 51

2.3.9 Weir & Cockerham's F_{ST}.. 52

2.3.10 Inbreeding Coefficients (F).. 53

2.3.11 Runs of Homozygosity.. 53

2.4 Discussion of Results.. 56

2.4.1 Question 1: Are the five geographically defined populations of Seychelles magpie robin genetically distinct?.. 56

2.4.2 Question Two: Have translocations had an impact on the genetic diversity of this species?.. 57

2.4.3 Question Three: What is the state of inbreeding in each of the five populations?. 60

2.5 Conclusion.. 64

Chapter 3 66
General Discussion 66
3.1 Overview of Project 67
3.2 Applications to Conservation 68
3.3 Future Research 70
References 72
List of Abbreviations 84
Appendices 85
Appendix 1 85
Appendix 2 86
Appendix 3 87
Appendix 4 88
Appendix 5 89
Appendix 6 90
Appendix 7 91
Appendix 8 92
Appendix 9 95
Appendix 10 96
Appendix 11 97
Appendix 12 98
Appendix 13 103

List of Tables

Table 1 Table of all known inter–island migrations of the Seychelles magpie robin 19

Table 2 A table outlining all basic filters applied to the dataset across all software......................36

Table 3 A table displaying the maximum number of iterations used to ensure convergence for each 'k' for admixture analysis 37

Table 4 Average sequencing depth of coverage by island 44

Table 5 Weighted pairwise F_{ST} values...52

List of Figures

Figure 1 A) A cropped map of Seychelles showing *relative* location of all 5 'magpie robin' islands. **B)** Table summarising translocation details **C)** population trends of the magpie robin census population from 1988–2014 15

Figure 2 Visualisation of how a genetic bottleneck can affect the genetic diversity of a population 20

Figure 3 An example of molecular sexing results 43

Figure 4 A) and **B)** Principal Component Analyses of 110 samples from all islands........ 46

Figure 5 Admixture analysis of 110 samples from all islands........ 47

Figure 6 A) Global autosomal heterozygosity estimates for the magpie robins. **B)** Genome–wide average heterozygosity estimates for 11 IUCN Red List species including magpie robins........ 49

Figure 7 Local heterozygosity estimates for five high coverage individuals (one from each island population) 50

Figure 8 Folded population level Site Frequency Spectrum 51

Figure 9 A example of a neutral expected Site Frequency Spectrum........ 52

Figure 10 A) Median total length of Runs of Homozogysoity categories by population, **B)** A boxplot showing F_{ROH} for each population **C)** and **D)** show the population distribution of ROH by length categories 4–6Mb and 6–8Mb respectively 55

Chapter 1

General Introduction

'Early English colonists bestowed the name magpie robin, because the bird's bold plumage resembles that of the European magpie, and its close relationship with humans is reminiscent of European robins.'

Shah, N. (2002)

1.1 A Brief History of the Seychelles Magpie Robin

The Seychelles magpie robin (Scientific name: *Copsychus sechellarum;* Seychellois Creole name: Pi Santez; herein also referred to as 'magpie robin') is a bird species in the family Muscicapidae in the order Passeriformes. Once classified as a thrush – (Copsychus being the Greek for blackbird) – and initially described in Newton's 'On an apparently undescribed bird from the Seychelles Islands' (1865) – the Seychelles magpie robin, and indeed other *Copsychus* species, are now treated as part of the Old World Flycatcher family due to advances in phylogenetic analyses (Sangster *et al.*, 2010).

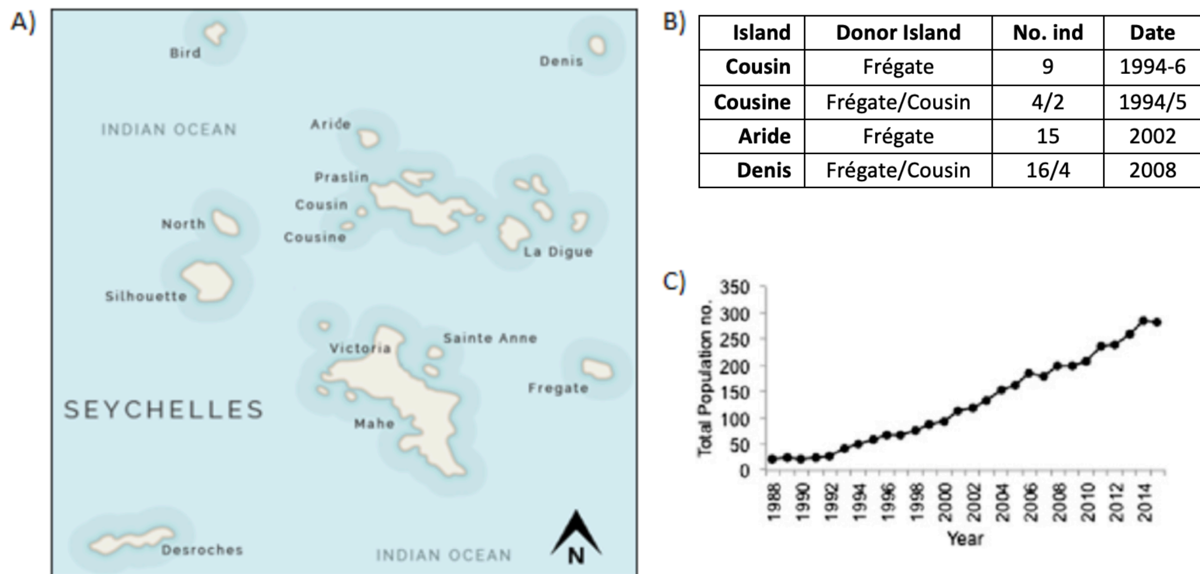
This charismatic bird has had a tumultuous history. One of 13 endemic bird species and the most endangered (Gaymer *et al*., 1969; Nature Seychelles, 2018), the magpie robin once inhabited at least eight of the 115 islands that make up the Seychelles Archipelago, and is native to Mahé (Newton, 1867), Praslin (Blackburn, 1878), Aride (Hartlaub, 1877), South East Island, St Anne (Skerrett, 2000), La Digue, Marianne and Alphonse (Ridgway, 1895; Vesey–Fitzgerald, 1940). The Seychelles magpie robin is just one of many species to have been affected by human actions over past centuries. Several factors are responsible for driving the species to near–extinction in the mid–1900s, including agricultural intensification, the introduction of exotic predators such as rats and cats and, as suggested from the records of a visit by Landz and his research team in 1878, and specimen over–collection (Gaymer *et al.*, 1969; Wilson and Wilson, 1978; Watson *et al.*, 1992). The outcome for this species looked bleak in the middle of the last century. However, recent conservation efforts from organisations including Birdlife International, the Royal Society for the Protection of

Birds (RSPB) and the Seychelles Magpie Robin Recovery Team (SMART) have led to a recovery, by means of population expansion, from near extinction.

Records suggest the magpie robin populations were already experiencing a decline in the mid 1800's (Newton, 1867) and by 1965 there was a total of eight sighted and an estimated total of 15 individuals remaining solely on the island of Frégate – the lowest ever recorded number for the species (Dawson *et al.*, 1965). Following this, multiple research expeditions were undertaken to study the land birds of the Seychelles, and to monitor the magpie robin population. The population size fluctuated over the following decades, with a maximum count of 38 birds recorded in 1973 (High, 1974, cited in Watson *et al.*, 1992) plummeting once more in the 1980s (Edwards, undated). This final fall in numbers was the driving force behind the recovery program implemented in 1990, which involved the eradication of birds and mammals that prey upon the magpie robins, and the removal of invasive plant species and subsequent habitat restoration: it resulted in an immediate population increase (Burt *et al.*, 2016).

Arguably the biggest contributing factor to the recent conservation success has been the translocation of birds to environmentally suitable islands to help population recovery (Burt *et al.*, 2016). Burt *et al.* (2016) published a comprehensive overview of all translocations that have occurred, by compiling both published and unpublished reports and island data, which I outline here: In 1994, Cousin was the first island to receive birds by translocation from Frégate. This occurred over three dates between 1994–1996 totalling 9 birds, and as the population flourished no future translocations were required. Cousine Island received six birds from Frégate in 1995–6 and one bird from Cousin. Again, no further translocations were deemed necessary. Numerous early attempts were made to establish populations on Aride (even prior to Cousin) but none was successful (Watson, 1978; Lucking & Ayrton, 1994; Parr, 1998; Millet & Shah, 2000). In 2002, another larger–scale attempt was made whereby 15 individuals were moved to Aride Island from Frégate (Shah & Parr, 2002). The

future for this population seemed promising until 2014 when more than three–quarters of the population perished with no confirmed reason. The most recent translocation was to Denis Island, and took place in 2008, with 20 birds translocated (16 from Frégate, 4 from Cousin). The Denis island population has seen continuous expansion and therefore not received any more translocations. The census population now stands at over 300 individuals across five islands (see Figure 1) Cousin: 50, Cousine: 50, Aride: 13, Frégate: 145, and Denis has a minimum of 78 birds (SMART, 2019). As each population was established from only a few founders, and mate choice was and remains restricted due to isolation and small population sizes, inbreeding is likely to have occurred (Keller & Waller, 2002; Frankham *et al.*, 2014).



Island	Donor Island	No. ind	Date
Cousin	Frégate	9	1994-6
Cousine	Frégate/Cousin	4/2	1994/5
Aride	Frégate	15	2002
Denis	Frégate/Cousin	16/4	2008

Figure 1 A) A cropped map of Seychelles (adapted from africaodyssey.com) showing *relative* location of all 5 'magpie robin' islands. The distance between Frégate and Denis is 65 km, Aride and Denis is 45km, Aride to Cousin is 10km and Cousin to Cousine is 2km. The Seychelles archipelago has a total land mass of 459km^2. The total size inhabited by the magpie robins is only 4.86km^2 **B)** Table summarising the island receiving donors, the donor island, the number of individuals translocated ('No. ind') and the dates on which the translocations occurred **C)** Population trends of magpie robin whole population 1988–2014, adapted from Burt *et al.* (2016).

If an individual survives to adulthood, the average lifespan is 4.2 years (Nature Seychelles, *unpublished*) but magpie robins have been documented to live for up to 15 years and have an estimated generation time of 3.6 years (Birdlife International, 2017). A breeding pair may occupy a territory for substantially longer than this, living with up to 8 non–breeding

subordinates within the territory. Related subordinate magpie robins have been observed to act as ‘helpers’ in the group – feeding the chicks. However, the presence of subordinates is largely recognised as causing conflict within the territory, therefore given the small islands and limited territories and resources, this can ultimately slow the process of recovery (López–Sepulcre *et al.*, 2009).

Part of the active management plan developed and implemented by SMART 29 years ago (Appendix 1) is the identification banding and close monitoring of the magpie robins. The resulting documented history of the populations makes them a good study species from a conservation perspective. Staff have collected blood samples and taken biometric measurements from most individuals, and observed and documented sexual and social behaviours, nesting patterns and breeding success, along with any other important observations. We can now look at population trends and behaviours over time and combine this with genetic data gained from this research for a more complete picture of the demographic history of the species.

1.2 Why Save the Magpie Robins?

First, we must ask ‘why is biodiversity important?’ Ecologically, birds are known to influence beneficial environmental processes across ecosystems. Their impact is evident through ‘nutrient cycling’, seed dispersal and pollination of many plants, and ‘pest’ control and maintenance of insect numbers (Whelan *et al.*, 2015; Zacheis *et al.*, 2002).

Economically, diverse and unique flora and fauna compositions are seen to boost tourism. Visits to private islands of the Seychelles are often driven by endeavours to witness endemic birds and as such it can be argued that there are also local financial incentives to preserve endemic species.

The International Union for the Conservation of Endangered Species (IUCN) Red List classifies 'threatened' species as those that are 'critically endangered', 'endangered', or 'vulnerable'. In 2017, the IUCN identified that of the 10,966 extant species of bird assessed, 13% are considered threatened, and therefore at some degree of risk of extinction in the wild (IUCN, 2017). The Seychelles magpie robin is currently classified as 'endangered', having been downgraded from 'critically endangered' in 2005 following a consistent increase in population numbers, one of only 32 bird species to have the conservation status downgraded between 1988–2008 (Birdlife International, 2008). The current recovery program and management plan in place for the species, implemented in 1990 and adapted with population expansion, has been successful thus far. However, 2014 saw a sudden and catastrophic crash of the magpie robin population on the island of Aride, from 30 birds to just 6 (SMART, *pers. comm*). A wildlife vet visited the island soon after, but could not determine the cause for this crash, although a number of possible factors were highlighted such as hygiene risks – associated with living closely to humans – and inbreeding depression (SMART*, pers. comm.*). It is the uncertainty surrounding the collapse of the Aride population that raises cause for concern about the remaining individuals and prompted the need for further research to investigate the extent of genetic variability and potential inbreeding.

1.3 Genetics as a Tool for Conservation

Population genetics is the study of genetic differences between biological populations. Traditional conservation genetic analysis techniques often use neutral markers such as microsatellites to make inferences about population structure (the differences in allele frequencies and patterns of genetic diversity of a population), and demographic history (changes in a population over time) (Holderegger *et al.*, 2006; Frankham *et al.*, 2010). There are a number of factors that shape these processes including mutation rate, genetic drift, bottlenecks and gene flow. These factors are represented by genetic features such as heterozygosity and can be measured by individual variation.

1.3.1 The Significance of SNPs

Single nucleotide polymorphisms (SNPs) are variations that occur at a *single* base (*nucleotide* – A, C, T, and G), whereby the commonly observed base has been replaced by another. For example, in a diploid species an AA genotype may be found at a specific locus in the genome in some individuals in a population, but in one individual a C nucleotide is observed i.e. a *polymorphism* occurs at this site, and hence the site is a SNP. Depending on whether this variation results in a different amino acid and subsequently a different protein being produced, it can potentially result in phenotypic variation within the population. SNP sites are used to identify and measure variation across whole genomes at both the individual and population level and can be used to assess demographic and evolutionary history of both populations and a whole species.

1.3.2 Gene Flow & Mutation

As the magpie robins exist on five islands with very little migration between them, there is limited potential for gene flow to influence genetic diversity. When considering conservation 'ideals', the suggested level of gene flow to minimise loss of heterozygosity is 1–10 introduced migrants per generation (Mills & Allendorf, 1996). There have only been 14 reported cases of magpie robins migrating, mostly between close islands (Table 1). There has been a 50% success rate of migrating individuals establishing a presence on a new island (Burt *et al.*, 2016). It is assumed the migration is in response to territory conflict on the origin island, and survival of the migrating individual is seen to be dependent on acceptance by the 'new' island magpie robin population. When gene flow by migration is limited, new variation in a population will primarily arise by *de novo* mutation.

Table 1 Table of all known inter–island migrations of the Seychelles magpie robin (adapted from Burt *et al.*, 2016). Returned* = without assistance / Transferred* = with assistance

Date	From	To	Outcome
1997	Cousin	Praslin	Disappeared
2000	Cousine	Cousin	Success
2000	Cousin	Cousine	Success
2002	Cousine	Cousin	Success
2003	Cousin	Praslin	Returned* to Cousin
2003	Cousine	Cousin	Success
2003	Cousine	Praslin	Unknown
2003	Cousin	Praslin	Returned* to Cousin
2004	Aride	Denis	Disappeared
2005	Cousin	Cousine	Success
2006	Cousin	Cousine	Unknown
2007	Aride	Curieuse	Transferred* to Cousin
2008	Cousin	Aride	Success
2010	Cousine	Cousin	Success

1.3.3 Genetic Drift & Heterozygosity

Genetic drift is one of the driving forces of evolution, and happens independently of natural selection or mutation (Wright, 1931). Genetic drift describes fluctuations in allele frequencies that happen by chance such as the phenomenon of genetic bottlenecks and the founder effect. The effects of genetic drift and the occurrence of certain alleles becoming lost or fixed are more profound in small populations (Kimura, 1983). A genetic bottleneck occurs, often after a natural disaster, when the number of individuals in a population is reduced (see Figure 2). The founder effect takes place when a small number of individuals (and the alleles they harbour) establish a new population after being segregated from a main population. In the new population there is often reduced representation of the original range of alleles, potentially losing some 'advantageous alleles', such as those desired in fighting off disease, e.g. alleles within the MHC region of the genome. As a result, might be more susceptible to certain diseases (Matzaraki *et al.*, 2017). Alternatively, the new population may harbour a specific detrimental allele in high frequency, culminating in a persistent 'disadvantageous' trait. In the case of the magpie robins on Frégate island that were reduced to just a few breeding pairs in the 1960's, there will have undoubtedly been a genetic bottleneck resulting in a high likelihood of the founder effect with any subsequent founder translocations from this population.

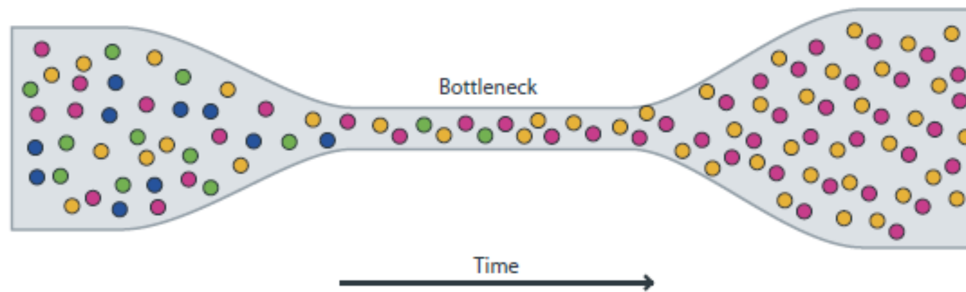


Figure 2 Visualisation of how a genetic bottleneck can affect the genetic diversity of a population (Jobling *et al.*, 2012). The coloured circles on the left represent the alleles present in a population before the bottleneck and those on the right are representative of the alleles that remain after a bottleneck, demonstrating how allelic diversity is lost due to this event.

Small population numbers that often follow a bottleneck also lead to increased chances of inbreeding – as there are fewer potential mates for the remaining individuals, and over generations the offspring are likely to become closely related (Andersen *et al.*, 2004; Herfindal *et al.*, 2014). All scenarios of genetic drift suggest a reduction in *heterozygosity* in a population. Heterozygosity is the state (genotype) where an individual has two *different* alleles of the same gene at a given location (locus) in the genome. This is contrary to the homozygous state, where an individual has inherited two *identical* copies of an allele. The level of heterozygosity that exists within a population is often viewed as a good predictor of the capacity of the population to be able to adapt in response to environmental change (Lande, 1988; Snustad & Simmons, 2011). Inbreeding over many generations leads to an increase in the proportion of offspring that exhibit long stretches of homozygous DNA known as 'Runs of Homozygosity'. This can often result in 'inbreeding depression' whereby there is an accumulation of recessive alleles with deleterious or harmful effects persist at high frequencies in the population (Charlesworth & Willis, 2009). In populations with reduced genetic variation and in those which suffer from inbreeding depression, detrimental effects on reproductive success, and survival under environmental change have been observed (Keller & Waller, 2002; Haanes *et al.*, 2013).

1.3.4 Linkage Disequilibrium & Haplotype Blocks

A group of alleles that are inherited together from a single parent is known as a haplotype or haplotype block. Linkage disequilibrium (LD) is defined as the non–random association of two or more alleles in a haplotype. When a given haplotype block is observed at a higher (or lower) frequency than that which would be expected under random association, the loci contained within the haplotype block are deemed to be in LD. The amount and extent of LD is expected to be higher in bottlenecked inbred populations, and patterns of LD can offer insight into inbreeding (Gray *et al.*, 2009; Garcia–Gamez *et al*., 2012).

1.4 Conservation in the Era of Genomics

Kardos *et al.* (2016), among others, have demonstrated how advances in genomics have improved the accuracy in measuring inbreeding and inbreeding depression. A key development over recent years is that we are now better able to understand how forces such as mutation and genetic drift effect the whole genome (Supple & Shapiro, 2018). With the advancements in genomic sequencing technologies it is possible to identify variants genome–wide, distinguish between neutral markers (those that are thought to have no impact on fitness) and adaptive markers (those that are under selection), we are also able to identify and investigate the activity of specific genes and study the variation of these within a population, and explore relationships between genes and environment (Allendorf, *et al.*, 2010; Angeloni *et al.*, 2012; Ekblom & Wolf, 2014).

The commercialisation, wider accessibility and decreasing costs of whole genome sequencing has led to genomics becoming the leading scientific approach in setting 'conservation priors' defined as 'specific predetermined objectives that aim to enhance and improve the viability of a population or species' (McMahon *et al.*, 2014). The field of conservation genetics has, thus far, been invaluable in the conservation of endangered species. However, these recent advances in sequencing technology and DNA analysis tools allow effective population size and genetic drift to be estimated more accurately with a much

higher marker density (across the whole genome). Furthermore, genomics allows the development of more complex investigations such as the study of local adaptation – a higher fitness of a population to a certain environment as a result of natural selection acting upon traits that make them better suited to that environment (Williams, 1966). As done here with the magpie robins, it is now possible to study divergence patterns between populations of a species, and to further explore reasons behind population struggle and survival rates. Genome–wide scans can also be used to identify a gene or set of genes exhibiting variability between populations, thus allowing investigation of any significance to the population.

1.5 Project Aims & Contributions

The research undertaken for this Masters project aims to introduce new methods to aid in the recovery of the Seychelles magpie robin, and to focus on promoting the survival of current populations. Although minimal genetics work has been carried out on Seychelles magpie robins previously, the need for such research has been recognised (Burt *et al.*, 2016). There has only been one published piece of genetic research for the Seychelles magpie robins, which used two genes from historical mitochondrial DNA as part of a phylogeographic study of all *Copsychus* species, examining the evolutionary history of the genus in an attempt to determine species divergence dates (Lim *et al.*, 2010). Conservation efforts up to this point have been invaluable in the recovery of the magpie robins, however the species is still not considered 'safe'. Introducing recommended conservation methods arising from this project will reduce the need for reactionary 'emergency' conservation (Redford *et al.*, 2011) and will contribute to the long–term management of the species. The most recent, and rapid, decline of the population of Aride from 30 to 6 birds in 2014 (Nature Seychelles, *pers. comm.*) not only highlights how fragile the future of this species is, but also reinforces the need for implementation of new conservation methods – particularly guided by population genetics data. The Seychelles warbler (*Acrocephalus sechellensis*) has a similar history to the magpie robins, and through recent microsatellite genotyping, important information has been gained with regards to population structure and suitability for

individuals involved in translocations (Wright *et al*, 2014). Due to extensive research, favourable management, and translocations, the Seychelles warbler has experienced a truly spectacular population increase from less than 30 individuals on Cousin Island in 1968 to an estimated 3000 individuals across five islands in Seychelles (Birdlife International, 2016). While habitat availability, territory size and small social grouping does not allow for such high population density of the magpie robins, proposed conservation actions for the species include translocating birds to new islands to allow population expansion.

This research is being undertaken in collaboration with the Seychelles Magpie Robin Recovery Team (SMART) in line with proposed conservation actions: to investigate the potential genetic influence on the downfall of the population on Aride, before considering future translocation to this island (SMART, 2018), and to provide genetic insight for the future translocations intended for this species (Birdlife International, 2017). As such, I addressed three main questions in this research project: are the five populations genetically distinct?; have translocations had an impact on genetic diversity?; and, what is the level of inbreeding within each subpopulation?

1.6 Thesis Overview

Chapter One is the introduction of key topics of focus for this thesis. Chapter Two introduces the Next Generation Sequencing technology used for this project. This chapter also outlines the methodological journey including blood collection and sample preparation, whole genome resequencing, the first ever glance of the genetic makeup of the magpie robins and examining the genetic diversity within and between the five small, geographically defined populations that exist for the Seychelles magpie robin. I will analyse whole population and sub–population relationships to investigate zygosity and determine the extent of inbreeding. Chapter Three offers a broader project discussion, focusing on genomics–based conservation of the endangered Seychelles magpie robin.

Chapter 2

Genome Sequencing – A Unique Insight into The Recent History of a Precious Endemic Bird

2.1 Introduction

2.1.1 The Seychelles Magpie Robin

In the last century, the Seychelles magpie robin species suffered a bottleneck followed by a period of decades where the population size remained small. The population did not see a recovery until conservation organisations intervened with a management plan that involved translocating the birds from the last remaining magpie robin island population in order to establish new populations on different islands. Translocated populations from previously bottlenecked populations, especially those that are geographically isolated, are particularly susceptible to inbreeding because of the small founder groups with already reduced diversity (Frankham, 1997). The increased homozygosity that often occurs as a result of inbreeding can lead to a build–up and expression of deleterious recessive alleles – this is what is referred to as inbreeding depression because the heterozygous advantageous sites are depressed due to inbreeding (Charlesworth & Charlesworth, 1987; Keller & Waller 2002). Furthermore, when fragmented populations exist and migration is limited, there is very little chance of the introduction of new genetic material into the gene pool of an inbred population. All aforementioned factors influenced the decision to study, the questions addressed, and methods used to study the Seychelles magpie robin at the genomic level. This chapter will discuss how methods relating to whole genome sequencing and resequencing were used to answer fundamental questions regarding the population structure, genetic diversity and levels of inbreeding of the current populations of Seychelles magpie robins.

2.1.2 Next Generation Sequencing

DNA sequencing is considered a pivotal advancement in the scientific approach to modern biological studies (Motahari *et al.* 2012). The capabilities of genetic sequencing have improved exponentially over recent years. The advent of DNA sequencing was in the 1970's with chemical cleavage methods used to determine nucleotides by fragment size, and subsequently nucleotide order (Maxam & Gilbert, 1977). From this evolved the Sanger Sequencing base–by–base method (Sanger *et al.*, 1977) known as 'first generation'

sequencing which is considered to have paved the way for all future sequencing platforms. This revolutionary technology allowed the first large genome to be sequenced – the draft sequence of the human genome – in 2001 (International Human Genome Sequencing Consortium, 2001). This approach was used almost exclusively until the 1990's and the development of pyrosequencing 'sequencing–by–synthesis' and Roche454. These were the first automated sequencing platforms, and the introduction to 'second generation' or 'Next Generation' Sequencing (NGS) technology. NGS allows the parallel sequencing of *millions* of short DNA fragments (Liu *et al.*, 2012). NGS currently dominates the sequencing world and it is this technology that was used to sequence the birds in this project.

Many NGS platforms use a 'shotgun' approach to sequencing, where short fragments of DNA are sequenced to produce short reads. These short reads are aligned to a reference genome, from which variants within the genomes of individuals in a population can be discovered. The use of reference genomes and individual alignments has dramatically improved our ability to study *non–model* organisms, paving the way for large scale phylogenetic studies across taxa. Whole genome resequencing provides insight into population history and demography by enabling researchers to produce, among many other analyses, summary statistics for structural variation, population allele frequencies and genome–wide heterozygosity from huge datasets.

Quality issues in sequencing data arise due to sample preparation techniques and sequencing errors made from NGS techniques and the short reads used for shotgun sequencing (Beltman *et al.*, 2016, Pfeiffer *et al.*, 2018). This affects patterns at the ends reads, as well as other base–calling and alignment errors (Bentley *et al.*, 2008; Nielsen *et al.*, 2011) although this can be reduced by using paired-end sequencing technology and other considerations (Pfeiffer *et al.*, 2018). Of particular concern are sequencing issues unaccounted for, causing bias in downstream analyses (Pool *et al.,* 2010). However, crucial

filtering steps are taken to ensure and quantify quality when mapping, creating a trade–off between the amount and quality of data used.

Depth of coverage ('coverage' hereafter) refers to the number of times each single nucleotide on each DNA fragment is sequenced. The average coverage for an individual roughly equates to ‘number of bases sequenced/size of genome’, with some margin of error. DNA fragments generated prior to library building are sheared randomly and sequencing is not uniform, so this can lead to a very variable number of reads for each site, and indeed each individual in the pool, even though samples are pooled equimolarly. Coverage correlates with the credibility of correctly assigning genotypes. If a single diploid site in an individual is sequenced twice, and one read matches the reference nucleotide and one read differs, that variability is present 50% of the time. This results in uncertainty about whether this is a true genotype, or a sequencing error. Untangling this uncertainty in apparent variability is possible with more reads at each base to isolate anomalies as sequencing error. Chances of sequencing errors occurring increase when working with diploid genomes rather than haploid genomes as there is greater potential for error in genotype calling e.g. if only one nucleotide of the pair is sequenced ‘correctly’, or one nucleotide of a pair is not sequenced at all.

Because the cost and demand of deeply sequencing multiple individuals is still relatively high, the trend in research is to sequence a large number of individuals at low coverage, which is driving the development of better tools to analyse such low coverage population sequencing data.

2.1.3 Why Sequence Whole Genomes?

Studies using microsatellite markers and Restriction Site Associated DNA (RAD) markers can inform aspects of the population genetic structure. However, in species suspected with high levels of inbreeding, these methods would not be appropriate as they rely on too few

markers to effectively represent the whole genome. Therefore, such methods would not be able to be used effectively to address the pertinent questions identified for my study species. Based on this, considerations were moved to whether to sequence the exomes (the protein coding regions of the genome/genes) only or to sequence whole genomes. Lelieveld *et al.* (2015) identified the benefits of using whole genomes as most SNPs (sites of variation), useful for this study on the expectedly inbred populations, occur in intergenic regions (the regions between genes) which would be missed if only exomes were targeted.

2.1.4 The Avian Genome

The ability to study birds at the genomic level is relatively new. In 2010, only 3 whole avian genomes had been sequenced, with only one of these being sequenced using NGS technology (Dalloul *et al.*, 2010). However, since then, NGS technologies have enabled researchers to sequence and publish over 50 whole avian genomes (Zhang *et al.*, 2014a) and the 'B10k project' is now working to sequence the genome of all extant bird species (Zhang *et al.*, 2014b) hence this number is growing rapidly. The Seychelles magpie robin nuclear genome, sequenced for this first time as part of the broader work for this thesis, has now contributed to this endeavour.

The Animal Genome Size Database reports that bird genomes range from 0.9Gb (0.9 billion bases) to 2.16Gb (2.16 billion bases) with an average of 1.35Gb across 898 species (Gregory, 2018). The development of special bioinformatic tools allows those studying genomics to filter and utilise this wealth of information in the most appropriate and advantageous way to answer the questions at hand. The Seychelles magpie robin nuclear genome was assembled to a total length of 1.05Gb, which sits comfortably within the published bird genome size range.

2.2 Methods

2.2.1 Sampling

Blood samples from 135 birds were collected from five islands in Seychelles, by trained local conservation staff. Blood sampling was carried out by the withdrawal of blood (maximum 70µL) using a syringe or capillary tube via the wing vein. Unlike the red blood cells of mammals, the red blood cells of birds are nucleated. Thus, nuclear DNA extraction from blood was chosen as the preferred method to yield high amounts of DNA, and the collection method is less invasive than using tissue samples from live birds.

Blood samples used for this project were mostly collected between 2015–2017. Some sample dates were unknown, but it is suspected that no sample was older than 10 years. However, the 24 samples from the crashed population of Aride collected pre–2012 were sent by Dr Kate Lessells and Christa Mateman (Netherlands Institute of Ecology (NIOO)), and do not have specific dates of sampling. In total, 135 blood samples were received from Seychelles and 24 from NIOO, totalling 159 samples, 110 of which were able to be used for this project. Due to the small population sizes, this number covers a largely representative sample of the populations on each of the five islands inhabited by the magpie robins: Frégate (15/145), Denis (32/78), Cousin (22/50), Cousine (16/50), and Aride (25 total: 1 from the surviving population and 24 from the pre-crash population).

2.2.2 DNA Extraction

Blood samples were received in three different storage buffers: Queen's buffer, lysis buffer and 97–100% ethanol. This required two different extraction methods, all focused on genomic DNA extraction. For samples stored in lysis and Queen's buffer, the DNA extraction protocol for the Thermo Scientific Kingfisher Blood DNA kit (ThermoFisher Scientific Inc., 2012) was used, with the adjustment of using 200µL of sample input rather than the recommended 250µL. The Qiagen DNeasy Blood and Tissue kit (Qiagen Inc., 2006) was used for those samples stored in ethanol, following the manufacturers protocol. Finally, the

DNA was eluted in 70µL of Buffer EB, and for the DNeasy kit a second elution was carried out by repeating the elution step to yield more DNA to work with downstream. Samples were quantified using the Qubit Fluorometer 2.0 Broad Range DNA assay kit (ThermoFisher Scientific Inc., 2015a) and DNA quality was assessed using the Agilent TapeStation 2200 and genomic DNA kit (Agilent Technologies Inc., 2015).

2.2.3 Molecular Sexing

Male and female Seychelles magpie robins exhibit minimal morphological differences and as such physical characteristics are not distinct enough to accurately identify sex. As the magpie robins are a non–ratite bird species, it is possible to differentiate between the two sexes molecularly by identifying size differences between the two sex chromosomes (Ellegren & Fridolfsson, 1999). This is considered a 'universal' method and is proven effective at sex identification in species within the Muscicapidae family. Molecular sexing for this project was carried out using Fridolfsson Primers (2550F/2718R) which target the Z– and W– linked CHD (chromodomain–helicase–DNA–binding protein) locus as described by Ellegren & Fridolfsson (1999). A small aliquot of genomic DNA was diluted to a concentration of 3ng/µL and subsequently used to carry out the molecular sexing PCR reactions using an optimised PCR profile for this species (Appendix 2). PCR products were run on a 2% agarose gel electrophoresis to separate the amplified DNA fragments according to size. The results enabled the identification of sex by the presence of a single (ZZ) or double (ZW) band for males and females respectively.

Prior to this project, molecular sexing of the magpie robins had not been routinely carried out since 2012. This method was therefore used to further determine the sex of all 135 individual samples sent directly from the Seychelles. The sexing of the 24 samples received from NIOO had already been completed by Christa Mateman, a laboratory technician at the institution (NIOO, *pers. comm.*).

2.2.4 Nucleic Acid Fragmentation

Current NGS ‘shotgun’ technology requires the genome of an organism to be sheared into small fragments, sequenced, and then ‘rebuilt’ using bioinformatic tools. DNA was sheared using ultrasonic acoustic energy, utilising Covaris M220 Focused–Ultrasonicator equipment. DNA was fragmented following the protocols (with the adaptation of using a treatment time of 175 seconds) for either MicroTUBE 15µL or 50µL (Covaris Inc., 2015), where the volume used was determined by initial DNA concentration, to achieve an average target length of 250bp.

2.2.5 Library Construction

Input DNA for building libraries was normalised in order to reduce the effects of DNA concentration on final library concentration. Using the fragmented genomic DNA, **159*** shotgun libraries were constructed following the ‘BEST 2.0’ protocol (Carøe *et al.*, 2017). The process involved attaching oligonucleotide adapters, required for sequencing, to the fragments. One adjustment was made to the protocol: double volume reactions were used (64µL instead of 32µL) due to the high input DNA amount obtained during DNA extraction. A qPCR was performed on the library product to determine optimum number of cycles for indexing PCR using BGI primers (example of qPCR results in Appendix 5). After a second bead purification, final amplified libraries were quantified using the Qubit Fluorometer 2.0 and the High Sensitivity DNA assay kit (Thermo Fisher Scientific Inc., 2015b) before being equimolarly pooled.

***111 samples were sequenced within the allowed time frame and 110 were sequenced successfully.**

2.2.6 Sequencing

Samples were pooled and sequenced by eight individuals per lane. In order to be able to later identify and separate each individual sample within each pool, known–sequence indices were added to barcode the library for each individual. Each pool is required to be equimolar for input DNA to equalise numbers for reads for each sample. Twenty pools were sent for sequencing. Each of the 20 pools averaged a total concentration of ~100ng/µL.

Genomic shotgun libraries were sent to the BGI Sequencing Centre (formerly known as the Beijing Genomics Institute) to be sequenced using the unique BGIseq 500 platform, which uses an adaptation of the Illumina sequencing platform and utilise DNA NanoBall (DNBseq) Technology (BGI, 2019). Samples were sequenced using 'paired–end' reads of 100bp, meaning that every ~250bp fragment was sequenced twice – one 100bp read starting from the 3' direction on one strand, and one from the 5' direction on the opposite strand. Paired end reads were used as this allows for more accurate alignment to the reference genome (Illumina Inc., 2017), and the slightly larger size of the fragment reducing the likelihood of the reads overlapping in the middle which maximises coverage across the genome.

2.2.7 Reference Genome Assembly

A species–specific reference genome allows for accurate mapping of the data when rebuilding the genomes of each individual to ensure correct sequences of DNA to be used to ascertain genetic variation. The DNA extract for one sample (SAFring no. 4A52064) from Cousin island was used to build a reference genome *de novo*. As females are heterogametic, a female was chosen in order to have both male and female sex chromosomes sequenced and annotated. DNA extraction for the individual used for the reference genome was carried out by myself at the Centre for GeoGenetics at University of Copenhagen. A high DNA yield was required from multiple extractions to ensure there was sufficient DNA for the library build – a total DNA amount of 75667.7ng was used to build libraries (protocol). The library building and subsequent sequencing was carried out by Love Dalen at the SciLifeLab in the Swedish Natural History Museum, using the Illumina HiSeq platform. Sequences were then assembled by Bent Petersen at the Danish Technical University, using SOAPdenovo v2.0 and Allpaths-LG version 52488.

2.2.8 Read Alignment & Quality Control

Using bioinformatic tools, the genome of each individual was reconstructed by mapping the reads against the reference genome using the Paleomix Pipeline (Schubert *et al.*, 2014). First, the raw data reads were demultiplexed: the index primers added during the library

building stage were used to separate reads by individual, and then were removed from the sequence. The Paleomix Pipeline includes several stages carried out by multiple programs. First, the AdapterRemoval software (v2.2.2) ‘cleans’ the resulting fragments of adapters that were added during library construction. BWA–mem (v0.7.15) software maps the ‘cleaned’ reads to the reference genome and removes reads that do not successfully map to the reference. Read quality was defined and filtered at a later stage (see Table 2). PCR duplicates that arose during library construction were removed. Finally, indel–realigned BAM files (binary version of aligned map) were generated sing the Genome Assembly Tool Kit (GATK) (v3.8.0).

2.2.9 Analysis

2.2.9.1 Removal of Sex–Linked & Small Scaffolds

In birds, the male sex chromosome is homozygous (ZZ), while the female sex chromosome is heterozygous (ZW). Furthermore, local mutation rate and nucleotide diversity varies between the avian sex chromosomes (Montell *et al.*, 2001). Therefore, only autosomal scaffolds were included in all analyses undertaken in this research. Scaffolds were identified as autosomal following identification and isolation of Z– and W–linked scaffolds (Rute Fonseca, in–house script, *pers. comm.).*

Scaffolding, used for the assembly of the reference genome for this project, joins contiguous sequences of DNA where the longer (unbroken) sequences increase the reliability of those sequences being correctly orientated.

The total magpie robin genome is 1.05Gb. Using only those scaffolds over 100kb in size and with sex-linked scaffolds removed, a total of 232 scaffolds were used for the analyses of this project, with a total size of the genome used of 0.95Gb.

2.2.9.2 Genotype Likelihoods

Once the reads were cleaned and mapped, ANGSD (Analysis of Next Generation Sequencing) v0.921 was used to generate genotype likelihoods and allele frequency estimates, generated using quality filters described in Section 2.2.9.3 (Korneliussen *et al.*, 2014). Genotype *likelihoods* were preferred over genotype *calling* for this dataset as the data were generated at low to medium coverage (see Section 2.2.4.4) and *likelihood estimates* propagate at least *some* of the uncertainly that is produced with this level of coverage from Next Generation Sequencing, either through error in sequencing or incorrect alignment to the reference genome (Skotte *et al.*, 2013). Genotype likelihoods offer a probability score for each of the three possible genotypes at each site (e.g. AA, Aa, aa) when mapped against the reference genome. The genotype with the highest likelihood is then given as output. If there is variation in that genotype, compared to the reference genome, that site is then considered a polymorphism (SNP) (Li, 2011). Genotype likelihoods are then used to generate allele frequency estimates, being the relative frequency (actual number of observed alleles compared to total number) of an allele at each site.

2.2.9.3 Filtering

When working with NGS data, stringent filtering is one of the most important steps to ensure the quality of the data, allowing only the most informative SNPs to be considered in downstream analysis. This requires consideration of bias in factors such as coverage, sequencing quality and mapping quality. Filters were chosen based on recommended parameters, previous work using low coverage NGS data, and knowledge of the species history. Appropriate filters can help to reduce the number of uncertainties associated with for example low quality or mismatched reads, sequencing errors, effects of poorly assembled scaffolds, paralogs, and repeat regions. Thus, the filtering parameters outlined below were used as a minimum threshold for SNPs to be kept in the dataset when calling genotype likelihoods. The function of some software relies on no prior assumptions being made about

the dataset, and thus filtering options were limited to the recommended basic parameters. These instances are laid out clearly in the text.

SNPs were not filtered for Hardy–Weinberg Equilibrium (HWE) as due to the recent history of the species it is likely many SNPs would not be in HWE and applying this filter would severely deplete the dataset. Instead, inbreeding coefficients (Section 2.3.9.6) were estimated independently and applied as additional filters.

Table 2 A table outlining all basic filters applied to the dataset across all software. Allele frequency estimates and subsequent SNP 'calling' are based on genotype likelihoods, generated with ANGSD.

Allele Frequency Estimates		
FILTER	**THRESHOLD**	**FUNCTION**
–doMajorMinor	1	Estimates allele frequencies using both fixed major and minor alleles, infer both the major and minor allele directly from the genotype likelihood data
–doMaf	1 or 2	Assume known 1) major and minor allele or 2) minor allele only when estimating allele frequencies
–minMaf	0.05	Set a minimum minor allele frequency of 0.05
–doSaf	1 or 2	Estimate site allele frequency where 1: no prior assumptions imposed upon dataset or 2: taking inbreeding coefficients into account)

SNP 'calling'		
FILTER	**THRESHOLD**	**FUNCTION**
–minQ	30	SNP must have a minimum quality (PHRED) score of 30
–minMapQ	30	SNP must have a minimum mapping quality score of 30
–doDepth	1	Dependent for setting a maximum depth threshold
–maxDepth	INT	SNP must have a maximum coverage depth of (3*average coverage*no. individuals)
–setMinDepth	4	SNP must have a minimum coverage of 4 reads
–SNP_pval	1e–6	SNP must have a p–value of less than 0.00001
–BAQ	1	(base alignment quality) Reduces SNPs being called close to potential INDELs (insertions or deletions), which may have caused surrounding bases to be misaligned, therefore reduces the chance of calling a dense region of SNPs that do not have a high probability of being correct
–C50	REF	Will not align a read with excessive mismatches, thus reduces 'overestimated' mapping quality in these areas

2.2.9.4 Population Structure

Some basic clustering analyses were performed to give an overview of genetic differences between the populations and subpopulations and to investigate any potential structure.

Principal Component Analysis (PCA) is a useful preliminary analysis to visualise population structure. The genotype likelihoods and allele frequencies from an input dataset of 110 successfully sequenced individuals were used to identify a total of 717,088 polymorphic sites, which were subsequently used as a prior for generating a pairwise covariance matrix with PCAngsd (Meisner & Albrechsten, 2018). The pairwise covariance matrix was used as input for the prcomp() function in R (R Development Core Team, 2008) to perform the principal component analysis, plotted with 95% confidence ellipses.

Admixture proportions were used to further explore population structure and ancestry proportions were estimated using the allele frequency estimates for each SNP. Using the same genotype likelihood data for all 110 individuals as input, NGSadmix (v32) was used to generate admixture coefficients which represent the degree of admixture between both individuals and populations, and was chosen as the preferred tool to use for this as it has been found to produce reliable results from low to medium coverage data (Skotte, 2013). NGSadmix was run using up to 5 ‘k’ (ancestral) populations, which gives a likelihood of the proportions of the genetic makeup of each of the study individuals from those 2,3,4 and 5 assumed ancestral populations. The number of iterations for each ‘k’ was determined by convergence – when the program reaches 5 likelihood values that are similar enough to be considered a true global maximum (Table 3).

Table 3 A table displaying the maximum number of iterations used to ensure convergence for each 'k' (NGSadmix v32).

k=	No. iterations
2	112
3	212
4	441
5	580

2.2.9.5 Heterozygosity & Genetic Diversity

Several different analyses were used to estimate levels of heterozygosity, homozygosity and genetic diversity that exists in the different magpie robin populations.

Inbreeding Coefficients (F) estimated for each individual are necessary priors for genotype calling in inbred species and have been shown to effect results of the Site Frequency Spectrum (SFS) as when inbreeding is not taken into account for these estimations, heterozygosity can be overestimated. Therefore marker–based F values were determined using ngsF (Vieira *et al.*, 2013) and were incorporated into the SFS and F_{ST} analysis outlined in this section. Measured by homozygous sites across the genome, the inbreeding coefficient represents the proportion of the genome that is more homozygous then expected, given the population allele frequency as a reference panel. Although F values were calculated as a necessity for programs which require an inbreeding coefficient as input for analysing inbred populations, F statistics are not recognised as the 'best' way to measure individual inbreeding (Kardos *et al.*, 2018) therefore alternative measures, explored below (Section 2.2.9.4), were used to answer questions related to extent of inbreeding in the magpie robins.

The ***Site Frequency Spectrum (SFS)*** was used to assess the level of genetic variation within each of the five magpie robin sample populations. SFS estimates were calculated for each population using ANGSD and realSFS. All available sites were used in this analysis to obtain the highest accuracy, however only variable sites contribute to the spectrum presented in Section 2.3.8.

Constructing the SFS for each island required the generation of per–site posterior probability allele frequencies by incorporating inbreeding coefficients (Nielsen *et al.*, 2012). This is traditionally done using an ancestral genome. Currently there is no relevant ancestral

genome available for the Seychelles magpie robin, and therefore we would not be able to reliably differentiate an ancestral allele (original nucleotide) from a derived allele (nucleotide arisen from mutation). Therefore, the SFS was 'folded' (using the –fold 1 option in ANGSD). The –fold option looks for the rarest (minor) allele and considers it the derived state, and the major allele or most common is considered the ancestral state. Thus the 'folded' spectrum relies on minor allele frequency (therefore the *x*–axis on the constructed plot ranges from 0 to n+1). The resulting posterior probabilities were then used to estimate the Site Frequency Spectrum using the ANGSD tool realSFS to generate bootstrapped maximum–likelihood (ML) estimates of the SFS.

Pairwise F_{ST} was then calculated for all populations. The F_{ST} value obtained is a measure of the genetic differences between populations, which can give insight into which populations are more closely related and which have experienced the most differentiation. An unfolded SFS was constructed for each population and was used to generate a pairwise or '2–Dimensional' SFS between each of the populations. ANGSD is not yet able to accurately compute a 2–D SFS using a folded genome (a pairwise site frequency spectrum between populations) therefore this analysis used the unfolded SFS of each population (2n+1). The output was then used as a prior to estimate pairwise weighted F_{ST} for each pair of populations using realSFS. Sites were indexed to ensure the analysis was carried out on the same sites for each population, and then the global estimates were calculated using the 'stats' option (Nielsen *et al.*, 2012).

Global heterozygosity (average heterozygosity across the whole genome) was calculated at the individual level using ANGSD to create a folded Site Frequency Spectrum, and realSFS for the estimation of homozygous and heterozygous sites across the genome (simply defined as heterozygous or homozygous, with no differentiation in the AA or aa homozygous genotype). Heterozygosity was determined by the number of heterozygous genotypes as a proportion of the total number of sites. These values were visualised

together using a bar plot to allow for a comparison of heterozygosity levels between populations. Significance of heterozygosity results between islands was tested using an ANOVA, with heterozygosity as a response variable and islands as the only explanatory variable, validated by visualisation of residual plots following the guidelines of Zuur *et al.*, (2010). On a site–by–site basis, the results from this method of calculating heterozygosity was found to be affected by depth of coverage. In order to control for some of the bias introduced by coverage, a minimum site coverage of 4x was added for analysis carried out at the individual level.

Using the individual global heterozygosity estimates as a prior for ***local heterozygosity***, single sample per site theta estimates were generated to assess the distribution of heterozygosity across the genome. Watterson's Theta estimates (Watterson 1975) were characterised in windows, offering a value corresponding to the density of segregating sites per window region. Windows were defined to be 1Mb long with a step–size of 100Kb. Using the ggplot2 package in R (R Development Core Team, 2018) a genome–wide picture of heterozygosity was created, and, consequently, homozygosity patterns were created. The extended regions with depleted heterozygosity that were found in all individuals encouraged subsequent Runs of Homozygosity analyses outlined in the next section.

2.2.9.6 Individual Inbreeding

Runs of Homozygosity (ROH) were assessed across the genome of 106 individuals (four samples were removed due to low coverage, see Section 2.3.4). ROH analysis is recognised as a reliable way to determine the extent of individual inbreeding (McQuillan *et al.*, 2008; Keller *et al.*, 2011). A ROH is a length of DNA on which alleles at variant sites (SNPs) in a diploid individual are identical (homozygous), which indicates they have been inherited from a common ancestor (Broman & Weber, 1999). How recent this common ancestor is can be inferred by ROH abundance and length (Kirin *et al*, 2010; Pemberton *et al.*, 2012). Genome–wide SNPs were used to assess the proportion of the genome of each individual that is

Identical–by–State (IBS), and subsequently if these homozygous tracts were Identical–By–Descent (IBD).

The panel of 'private' SNPs, variation ascertained from individuals within each island, was generated using ANGSD with the standard filtering options identified in Section 2.2.9.3. These data were then fed into PLINK: The Whole Genome Data Analysis Toolkit, v1.07 (Purcell *et al.*, 2007) to assess genome–wide ROH by using the –homozyg command. The areas of interest in the ROH are not the common sequences across the population, but instead the runs where variants are observed as a homozygous genotype. ROH were called using the following thresholds: the maximum gap allowed between consecutive SNPs for them to still be considered in the same ROH was set at 1000kb bases (—homozyg–gap 1000); sliding windows were set to 300kb (—homozyg–window–kb 300); the minimum number of SNPs for eligibility of a run was 50 (––homozyg-SNP 50), a run being called required at least 1 SNP present every 50kb (––homozyg–density 50); 30 'missing' calls/sites (—homozyg–window–missing 30); 5 heterozygous calls were allowed per window (—homozyg–window–het 5). Although the latter two may be considered high, parameters are decided based on the data available, and these were considered suitable parameters for low coverage data and where data are missing at different sites for different individuals.

For this specific analysis, an additional parameter was imposed upon the dataset. The reference genome for the Seychelles magpie robin is assembled by scaffolds, with inevitable 'breaks' in sequence, mirrored in each individually mapped genome. An ROH can only be called from unbroken sequences of DNA. Therefore, only scaffolds of a minimum length of 10Mb were included in the ROH analysis. Subsequently. The first 30 scaffolds were used for analysis (scaffold_0–29). The longest scaffold was 57,831,083 (57.83Mb) in size, the shortest was 10,483,188 (10.48Mb) with a total used genome size of 730,210,129b (0.73Gb).

2.3 Results

2.3.1 DNA extraction & Library Construction

DNA was successfully extracted from all individuals at average of a 102ng/µL with a range of 4.2–568ng/µL (results in Appendix 3 and Appendix 6). Once extracted, all samples were tested for quality. There was minimal DNA degradation across all samples (example of Tapestation results in Appendix 4). The average fragment length after fragmentation with Covaris was 288bp (Appendix 4), with a standard error of 2.09 across 110 samples.

2.3.2 Molecular Sexing

The fragment that represents the Z chromosome is roughly 450bp. As male birds possess two copies of the Z chromosome, the banding patterns on the gel is single and bright. The fragment size of the W chromosome is roughly 550bp; females have one Z and one W chromosome, thus the gel run shows two separate bands (one at 450bp and the other at 550bp) which, due to the lower level of biological material present at each band, appear comparatively dimmer (Figure 3). The magpie robin samples were run against established male and female 'controls' from a Karoo scrub robin (*Cercotrichas coryphaeus*), another passerine species, which was sexed anatomically as well to confirm molecular sex results (Ángela Ribeiro, *pers. comm.*). These are labelled as 'C' and 'M' for the male control or 'C' and 'F' for the female control in Figure 3 and were used as a baseline for confirmation of the magpie robin results.

Of the 135 samples received from Seychelles sexed as part of this project, 64 were male and 71 were female. The results were sent directly to members of SMART to update their magpie robin database and were not used any further in this project. The results of the molecular sexing confirmed almost every behaviour–based sex assignment given, by trained staff, in the field (134 of 135 samples). The contradicting sex result for this single individual was confirmed in multiple independent tests at request of the island managers, so I am confident this individual was sexed accurately. This has positive implications for the current

method of sex identification through behavioural observations, which is widely used for this species and is particularly beneficial for times when molecular sexing is not available.

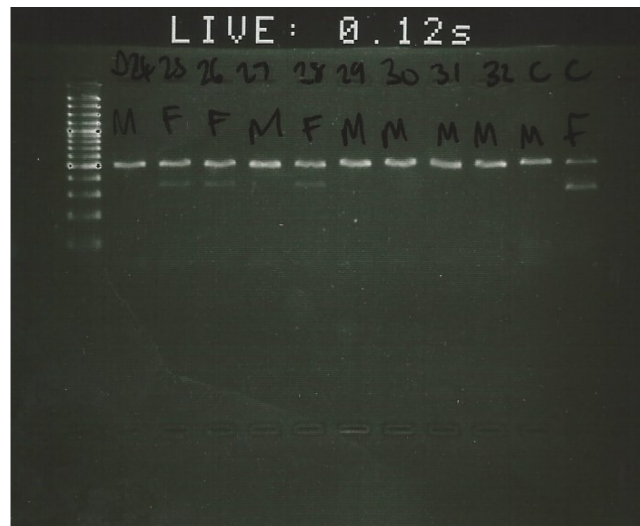


Figure 3 An example of molecular sexing results from Denis Island (sample IDs 24–32), run on a 2% agarose gel electrophoresis: labelled 'M' for male determined by the presence of a single band, and 'F' for female with a double band. The 'C' label is used for identification of the Karoo scrub robin control samples.

2.3.3 Reference Genome Assembly

The reference genome for the Seychelles magpie robin was sequenced at 374x coverage. The total assembled genome size for the Seychelles magpie robin is 1,052,508,292 base pairs (~1.05Gb). The assembled reference genome comprises 3664 scaffolds. The magpie robin reference assembly scaffold n50 size, that is the size of the scaffold where it and longer scaffolds cover 50% of the genome, is ~18.5Mb, with the longest scaffold at a length of ~57.8Mb. The magpie robin genome size is in line with other similar bird species (Gregory, 2018).

2.3.4 Sequencing Success, Failures, & Quality Control

This section provides a broad outline of sequencing summaries. An exhaustive overview of resequencing summaries for each individual is detailed in Appendix 8.

Only one sample, CN_04, was considered to have 'failed' sequencing with only 210,000 retained reads at a coverage of 0.0009x. This sample was removed from the dataset prior to any analysis undertaken.

The median number of mapped reads per sample across the remaining 110 samples was 133 million with a range of 25–404million, and the fraction of these reads were uniquely mapped had a median of 91% with a range of 73–97%. The clonality fractions ranged from 1% to 25% of reads generated, with the higher clonality margins observed within the first samples that were processed.

The depth of coverage from all 110 samples ranges from 2x to 34x. Average coverage per population is detailed in Table 4. Understanding the sequencing technicalities that may have been responsible for the differences in coverage between samples is beyond the scope of this project.

Table 4 Average sequencing depth of coverage by island and number of individuals used per island, standard deviation (SD) shown in brackets.

ISLAND (no.ind)	**ARIDE (25)**	**COUSIN (22)**	**COUSINE (16)**	**DENIS (32)**	**FRÉGATE (15)**
Coverage (SD)	14.17 (8.96)	12.97 (3.33)	11.93 (2.07)	15.54 (5.78)	7.13 (2.89)

It has been established that statistics which are performed at the population level generate fewer false–positive variant calls from low coverage NGS data than when performed at the individual level (Sims *et al.*, 2014). Therefore, for the individual level analysis carried out for ROH, a 5x average coverage threshold was used for data to be considered informative enough to provide meaningful estimates, and consequently four individuals (A_07 (4x coverage), A_09 (2x), F_05 (4x), D_11 (2x)) with an average coverage below this threshold were removed to ensure the integrity of the analysis.

2.3.5 Principal Component Analysis

The Principal Component Analysis was carried out on all 110 individuals collectively; therefore, no prior population relationships were inferred. The results show some population structure (see Figure 4A and 4B). Each principal component is presented as a percentage of the total variance. The founder population of Frégate sits as a fairly central figure, and on no axis do Frégate and Denis segregate. The first three principal components account for 19.2%, 17.0% and 12.6%, respectively, culminating in almost half of the total variation among the populations. The axis of the first principal component shows that Aride differs most from the other populations (highest degree of variation). PC2 and PC3 segregate Cousin and Cousine, respectively. While PC 4 does not appear to offer any discernible variation. Therefore, the first three principal components (PC) account for all island differences (PC5 and PC6 show no island clustering, see Appendix 10), and this segregation is mirrored in the genetic differentiation observed in the F_{ST} results (Section 2.2.3.9, Table 5).

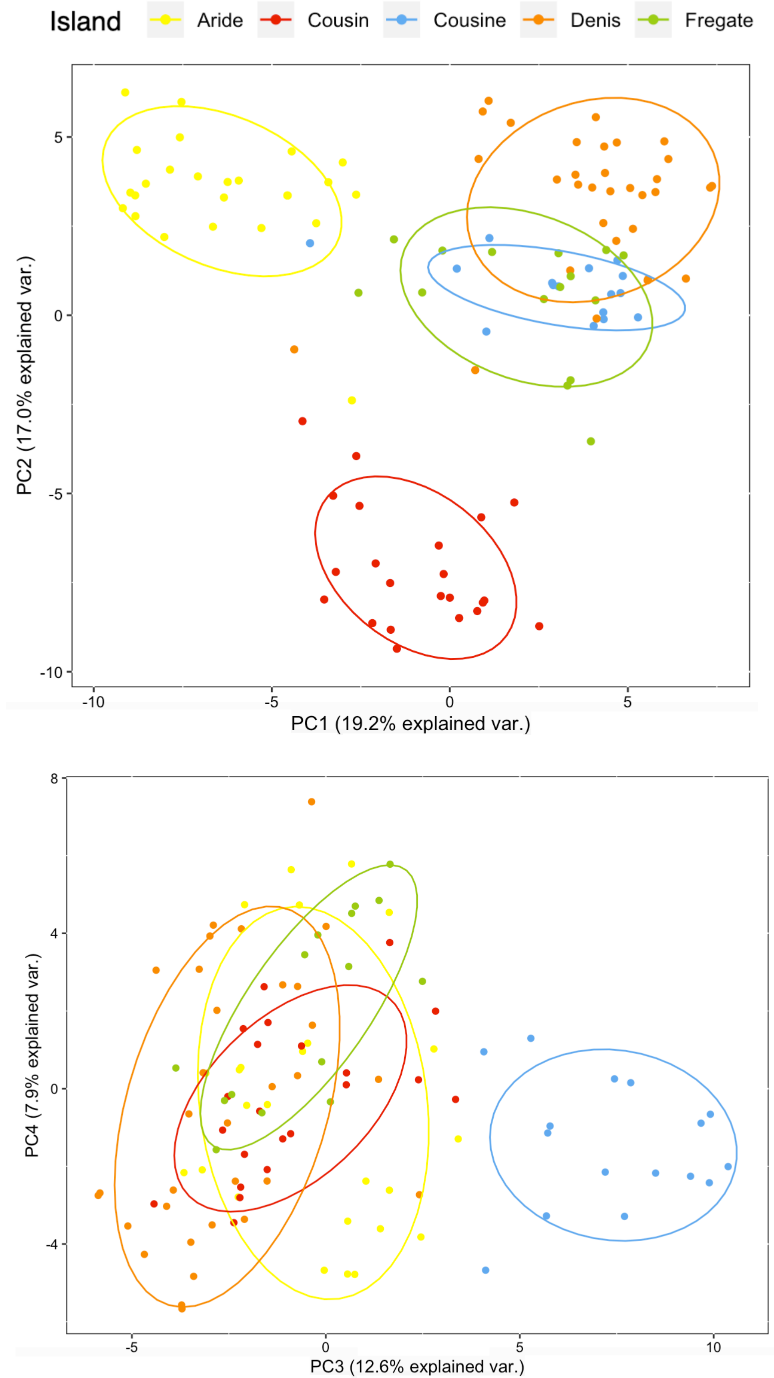


Figure 4 Principal Component analysis for all 110 successfully sequenced individuals representing five island populations A) PCA showing 1st and 2nd principal component, the founder population Frégate is represented by green, and 95% confidence ellipses are shown for each island B) As A) but showing principal components 3 and 4.

2.3.6 Admixture

The same input dataset that was used for the PCA was used to calculate admixture proportions, where no prior assumptions were made on the relationship of the individuals or population. In Figure 5, the populations are ordered from left to right by first to last translocation with the founder population on the right. Clear isolation of population structure would be indicated by a uniform colour observed throughout the population – this would be an 'unadmixed' population. NGSadmix uncovered some apparent population structure in the magpie robins, evident by a preponderance of one colour for most of the populations at k=5. The results show Frégate, the founder population, as the most genetically diverse population. The Denis population was the most recent translocation and can be seen to not have yet been affected by patterns of genetic drift, showing almost equal diversity as Frégate.

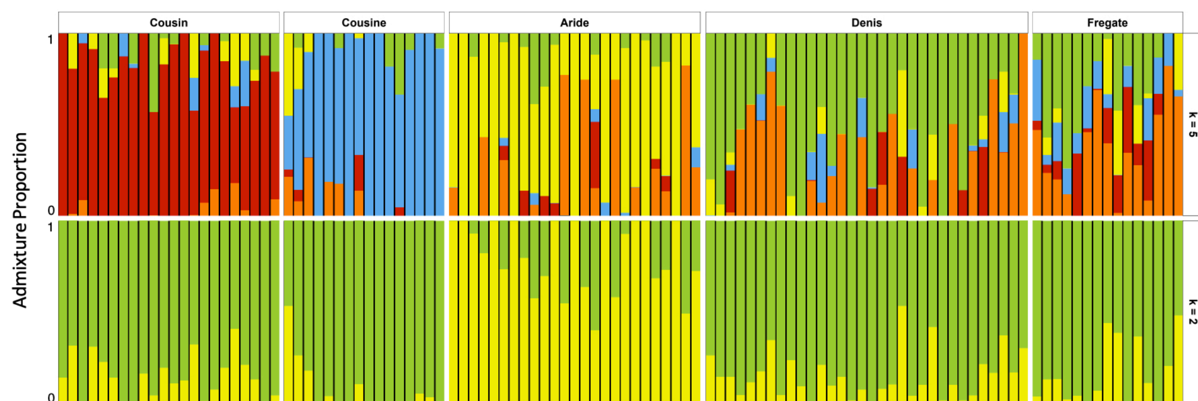
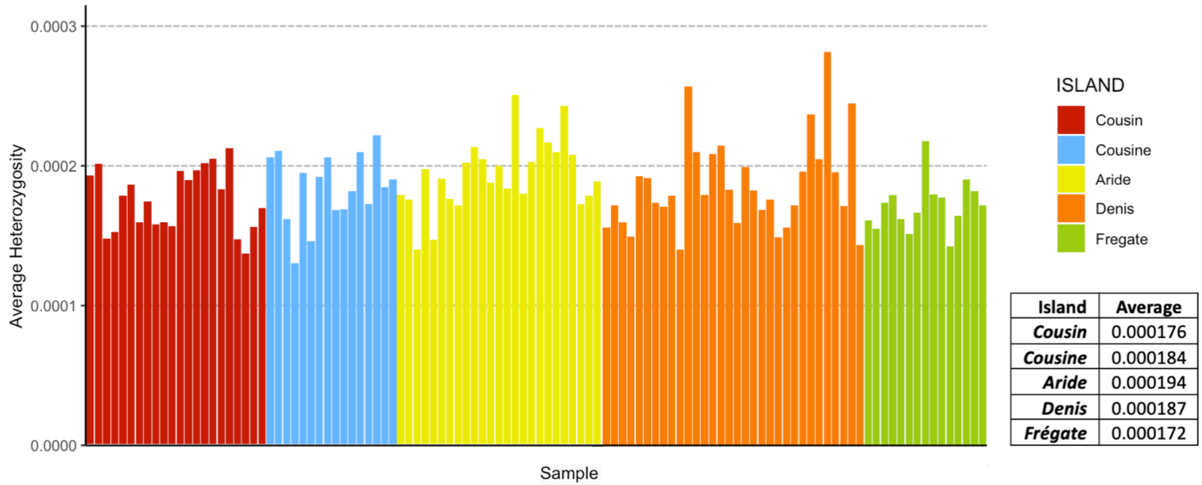


Figure 5 Admixture plot (NGSadmix) showing only k=2 and k=5 to demonstrate population differentiation. Populations are ordered left to right: Cousin–Cousine–Aride–Denis–Frégate, representing the first translocation to the most recent, with the founder population on the far right. Individual samples are represented on the *x axis* (one bar/sample), and admixture proportions are on the *y axis.* Each colour corresponds to an undefined ancestral population – 'k'. The full admixture plot including k=3 and k=4 can be found in Appendix 11.

2.3.7 Heterozygosity

A recent genomic study into the genetic diversity of 42 avian species (Li *et al.*, 2014) uncovered low levels of genome–wide heterozygosity in 'threatened' species, the results from 9 of which are illustrated in figure 6B along with the magpie robin genome–wide heterozygosity estimates obtained from this research. The group of endangered and Vulnerable species used in the avian genomics study found an average heterozygosity estimate of 0.0018 (calculated as the proportion of sites used that were observed heterozygous), with the lowest levels found in the crested ibis (*Nipponia nippon*), bald eagle (*Haliaeetus leucocephalus*) and white–tailed eagle (*Haliaeetus albicilla*) (0.00043–0.00040). The results from a similar analysis on the magpie robins in this research determined the observed heterozygosity to be less than half of this low average value (0.00018) (Figure 6A). While estimates were made at the individual level, the purpose of this was to gauge an understanding of whether the average heterozygosity levels differ between islands. The Aride population has the highest average heterozygosity across the whole sample population (0.000194). However, there was no significant difference among the five populations in average heterozygosity (ANOVA: $F_{4,105} = 2.306$; $p=0.063$).



Island	Average
Cousin	0.000176
Cousine	0.000184
Aride	0.000194
Denis	0.000187
Frégate	0.000172

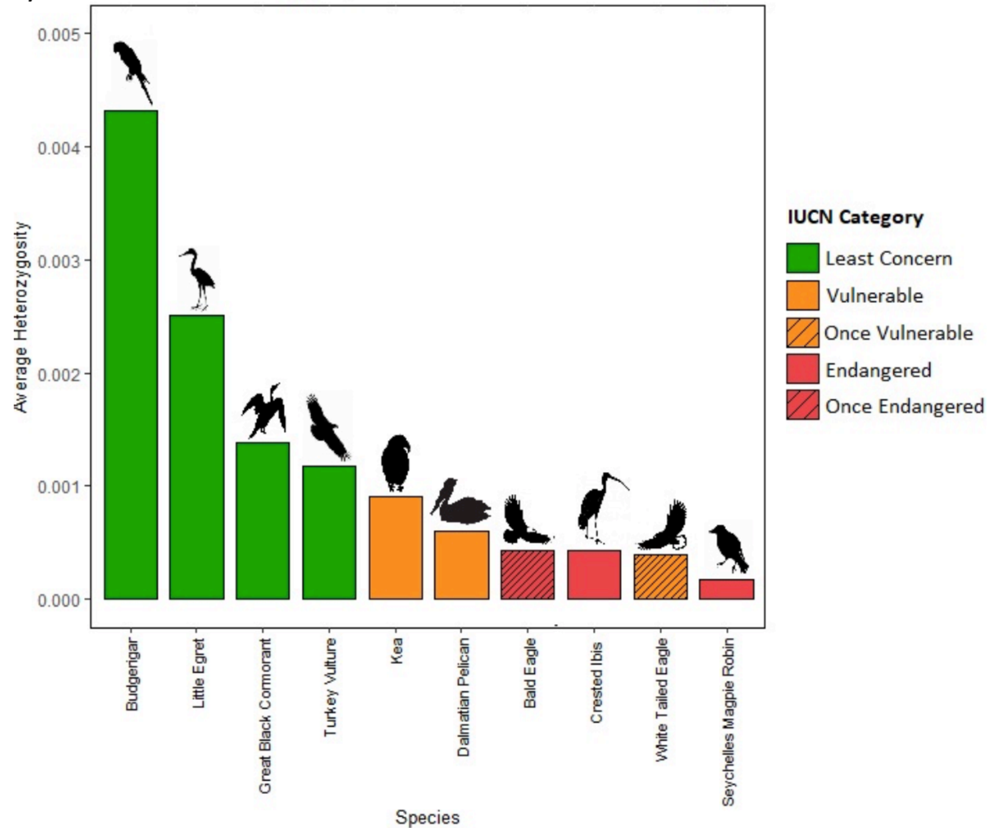


Figure 6 A) Global autosomal average heterozygosity estimates for each individual across all five magpie robin populations. The average calculated from bootstrapped estimates (using ANGSD and realSFS) for each Seychelles magpie robin sample, giving an overview of island level heterozygosity. INSET shows average heterozygosity estimate value for each island population **B)** Genome–wide average heterozygosity estimates for 10 IUCN Red List species, two of which used to be held in the category but have since been downlisted (indicated in legend). Plot constructed with the results from a published avian genomics study (Li *et al.*, 2014) incorporating Seychelles magpie robin results from this study.

Watterson's theta estimates, calculated in windows across the genome, were used to further investigate the low heterozygosity in order to determine any existing patterns in the distribution of heterozygous sites. An initial investigation was carried out for one high coverage individual for each island (Figure 7) and revealed long regions of diminished heterozygosity. When further exploration across the whole population found similar patterns of long homozygous tracts in every individual (Appendix 12). Patterns were noted such as the long stretch of homozygosity observed around 250Mb in four of the samples (CE17, CN20, D30, F08) but this is not observed in A28 (Figure 7). Further analysis will determine if there are shared regions of homozygous tracts within the populations.

Figure 7 Local heterozygosity (Watterson's Theta) estimates for one high coverage individual per sample population/island, estimates across 232 scaffolds above the length of 100kb, culminating in a total length of 900Mb.

2.3.8 The Site Frequency Spectrum

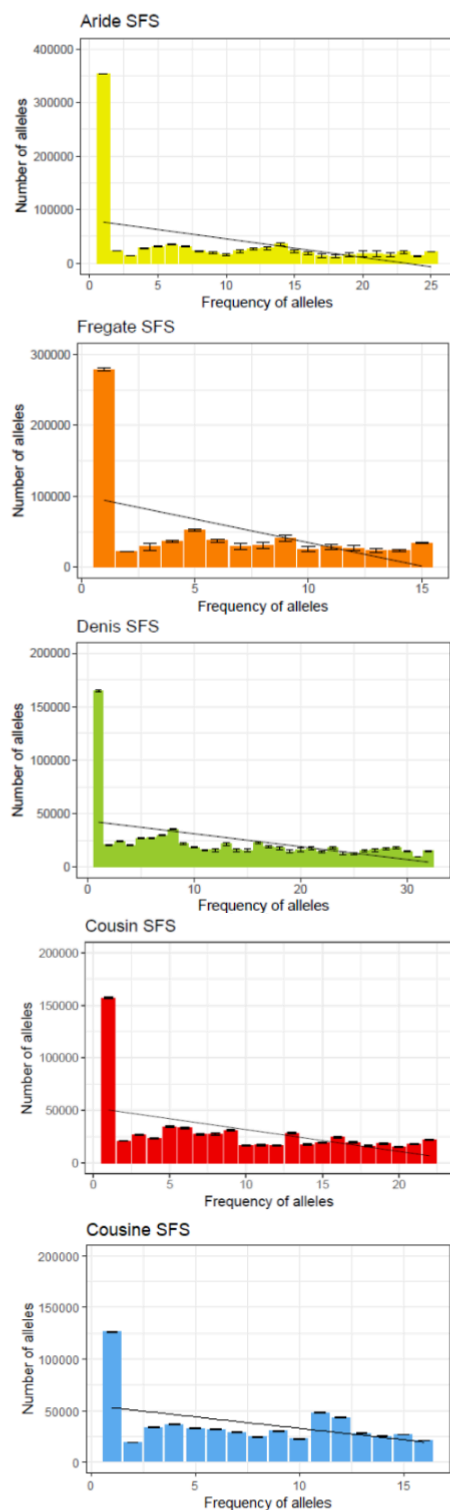


Figure 8 Folded population level SFS. Frequency of alleles observed in the population is plotted against the number of occurrences of alleles of that frequency.

Figure 8 shows a compressed SFS displaying only segregating sites: sites that had 0 variants across the population (bin=0) was removed. The resulting SFS shows the number of alleles with a minor allele frequency of 1 to 'n+1' within each population. The standard error bars represent the variance in the bootstrapped estimates obtained during SFS optimisation.

Over time, mutations and rare alleles can persist in a population, and over generations the frequency of these alleles increases, resulting in the characteristic shape of the *neutral expected SFS* (Figure 9), which assumes random mating and fixed population size (Hudson, 1987; Zhu & Bustamante, 2005). Each of these populations show a depressed distribution, relative to that of the neutral expected SFS.

While all populations hold a high number of singletons (bin=1), the Frégate and Aride populations have a markedly higher number of singletons.

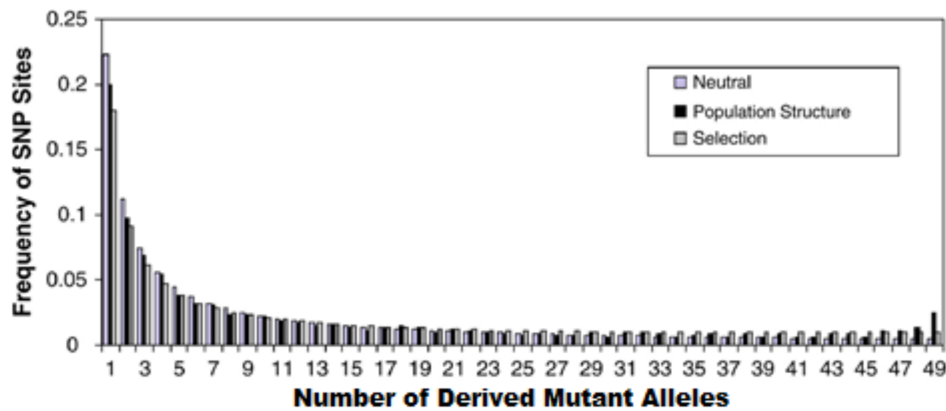


Figure 9 Adapted from Zhu & Bustamante (2005) where 'Neutral' demonstrates the distribution of allele frequencies neutral expected SFS; this represents an unfolded spectrum of 24 diploid individuals. Frequency of 0 not shown

2.3.9 Weir & Cockerham's F_{ST}

F_{ST} was measured between all pairs of populations, culminating in 8 successfully calculated F_{ST} values: an average weighted F_{ST} between each of the compared populations in pairwise analyses, shown in Table 5.

Table 5 Weighted pairwise F_{ST} calculated using ANGSD and realSFS including only indexed SNPs. Two results, indicated by '–', were not successfully computed.

ISLAND	Frégate			
Denis	0.052	Denis		
Aride	0.093	–	Aride	
Cousine	0.101	–	0.145	Cousine
Cousin	0.091	0.104	0.135	0.159

Using Weir–Cockerham's F_{ST}, the highest F_{ST} value (indicating the highest degree of differentiation between the pairs of populations) was observed between Cousin and Cousine with a weighted F_{ST} value of 0.159 (Table 5, highlighted in red). The highest degrees of differentiation were observed in the pairwise comparisons between Cousin, Cousine and Aride, these being the longest established introduced populations. Frégate and Denis were observed to have the least differentiation with a value of only 0.05 (Table 5, highlighted in green) these being around half the value of the other islands when analysed with Frégate. Results from F_{ST} analysis on the crested ibis, a species with a similar history, deemed an F_{ST} value of 0.134 and above relatively high, considering the short time of 20 years since

populations were established (Li *et al.*, 2014), which yields similar implications for the results from the magpie robin populations.

2.3.10 Inbreeding Coefficients (F)

The inbreeding coefficient is a value given as a proportion of 1, with a score of 1.0 indicating that the individual is highly inbred. If the 'observed' number of heterozygotes is equal to that 'expected' (usually under HWE, but in this study based on allele frequencies from the 'private' SNP panel of each island) then F=0. Half of the individuals had F=0, but the range was similar in each population from 0 to 0.22(Aride), 0.23(Cousin), 0.23(Cousine), 0.25(Denis) and 0.26(Frégate). An F score of 0.2 indicates a 20% probability that at any locus across the diploid genome, the two alleles observed will be identical–by–descent i.e. inherited from a common ancestor. The inbreeding coefficients (F) calculated with ngsF were used solely as input for accuracy of SFS and subsequently F_{ST}. A table of all inbreeding coefficients can be viewed in Appendix 13.

2.3.11 Runs of Homozygosity

Short ROH are expected between individuals of any population, therefore ROH less than 0.5Mb were disregarded from consideration as a factor of inbreeding and were not included in any plots or calculations (Cassidy *et al.*, 2016) and only ROH of 2Mb in length or above were included in F_{ROH} analysis as ROH of 2Mb have been considered a factor of distant inbreeding (Pečnerová, 2018).

The sample population of Frégate differentiates from other islands in that the median shows the population has a higher number of short ROHs i.e. categories 0.5–1 and 1–2Mb, and less ROH in the 8–16Mb category (Figure 10A). As the original founder population, this pattern infers more recombination events over time and can therefore be used as a 'baseline' in comparing the ROH patterns of other islands. Cousine most closely resembles the pattern of Frégate, but with a notable increase in the longer ROH, while Aride, Cousin and Denis show a distinct peak in the 4–8Mb category (Figure 10A), with a significant

difference found among the populations for ROH of this length (Kruskal-Wallis: chi-squared=12.15, df=4, p=0.016). This sparked the interest to further investigate the composition of the ROH of this category resulting in it being broken down in to 2Mb length categories (Figure 10C and 10D). Although the mean and median for the 4–6/6–8Mb (Figure 10C and 10D) subcategories is not too different between the populations, the lower peak observed at 4–8Mb for Cousine (Figure 10A) could be a result of a higher proportion of shorter lengths in each category. While the overall degree of relatedness generally appears high in each population, the length of the ROH can be used to infer recentness of inbreeding events (Kirin *et al.*, 2010). All populations show similar trends, the Denis population, however, shows the steepest incline at the ROH 4–8Mb category, and the highest peaks at the 8–16Mb and 16>Mb categories. All translocated populations were observed to have at least one ROH longer than 20Mb in the population.

When considering F_{ROH}, Aride has the highest median (Figure 10B) but the lowest number of 10+Mb lengths and highest 4–8Mb category (Figure 10A). Figure 10C shows Aride to have the highest number of 4–6Mb ROH, and although this population has the widest distribution at 6–8Mb, it has the lowest median of all islands for these categories. Therefore, although more of the genome appears to be IBD for the Aride population (Figure 10B), these tracts appear to be shorter, therefore these results are indicative of *less recent* inbreeding and *more distant* mating of related individuals, than on Denis island, for example, which has a lower F_{ROH} median for percentage of the genome that is F_{ROH} but has the highest number of ROH at both categories above 8Mb.

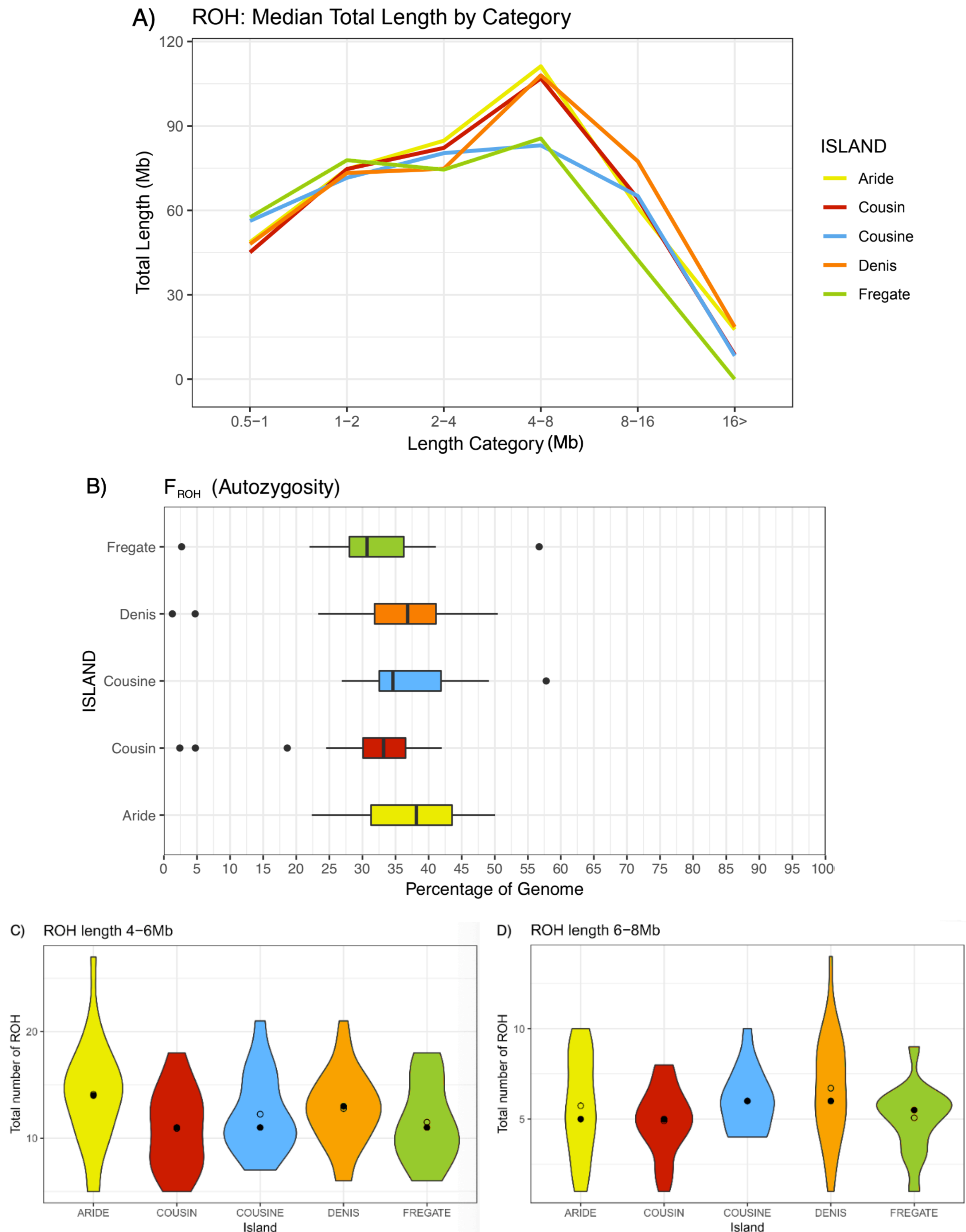


Figure 10 A) Median total length of ROH of length above 0.5Mb across the genome calculated for each length category, by population B) A boxplot depicting the range and median of the F_{ROH} (autozygosity) for each population calculated by 'total length of autosomal ROH in genome/length of autosomal genome covered by homozygous SNPs' (Moyse *et al*., 2016), where ROH included in this calculation were above 2Mb. C) and D) show the population distribution of ROH by length categories A) 4–6Mb and B) 6–8Mb. The median for each population is represented by a filled black circle and the mean is represented by a hollow black circle.

2.4 Discussion of Results

This study is the first genetic assessment of the endemic, IUCN Red–Listed 'endangered' Seychelles magpie robin. This research aimed to answer three basic questions about the current genetic state, using genome–wide SNP markers obtained from a sample of 110 individuals representing the populations on the five islands within the Seychelles archipelago on which this species exists. How each analysis performed in this project contributes to what we can now understand about the current Seychelles magpie robin population is discussed in this section in relation to the original questions posed in the Introduction.

2.4.1 Question 1: Are the five geographically defined populations of Seychelles magpie robin genetically distinct?

The principal component analysis shows some clear clustering by island, most specifically with islands whose populations were translocated earlier (Cousin:1994, Cousine:1995 and Aride:2002), probably a consequence of genetic drift acting upon those populations, although rapid selection has been documented to act upon small populations particularly when faced with adverse conditions (Lamichhaney *et al.*, 2016). Due to the donation of individuals to Cousin and Cousine from Frégate at its smallest population size (23), the contemporaneity of these translocations, and the fact the largest number of migrations have been between these islands due to their close proximity (Figure 1) it was recently suggested that they are more likely to be, genetically, 'one population' (Burt *et al.*, 2016). The differentiation, seen by F_{ST} value of 0.159, between Cousin and Cousine indicate that they are in fact two distinct populations and show a higher degree of differentiation from each other than any of the other islands. The results of this study show the birds on Frégate and Denis as the most similar, which may reflect the size of the donor population when the individuals were taken, the number donated, and the time since introduction.

The dataset for NGSadmix was not filtered to remove closely related individuals, however this does not seem to have caused any clear substructure within any of the populations. If relatedness had an effect on the output from this analysis, some level of clustering between more closely related individuals would be apparent. This having not occurred suggests that either none of the individuals are closely related or that the individuals are largely similar to each other genetically, to the point that the few related individuals did not have any impact on the results. Based on the detailed monitoring data we suspect the former is inaccurate and therefore the latter is more likely. The results from kinship analysis (not shown) were suspected to be unreliable due to this factor.

The F_{ST} values obtained here demonstrate surprisingly profound differentiation between the first three translocated populations, given the relatively short time since translocations. Fluctuation in allele frequencies, time taken for allele fixing, and population differentiation are greater in smaller populations, particularly when there is no migration between populations (Wright, 1931) and this may be the reason for the patterns observed as this is commonplace in introduced populations.

2.4.2 Question Two: Have translocations had an impact on the genetic diversity of this species?

One of the great contributors to the success of the magpie robin population is translocations. With such small populations, removing a few individuals from an island with no prior knowledge of their genetic make–up can have a significant effect not only on the variation that then remains in the island they were removed from, but also determines the genetic variation that can possibly exist within the new population. Whilst the PCA shows genetic distinctiveness, it does not measure genetic diversity. The patterns observed from the NGSadmix results suggest the Frégate and Denis populations have the most diversity by structure, which could suggest that these populations have been less affected by genetic drift than the other islands. This pattern may reflect Frégate harbouring all initial genetic

diversity 30 years ago when translocations began, the Frégate population has had the most generations for mutations to arise, it the largest population and therefore assumed effective population size, and Denis only received translocated individuals from this population 10 years ago. It is exciting to see that the founder effect did not appear to have caused much loss in the genetic diversity of Frégate because of the donations to new islands. Observable patterns, between Frégate and Aride for example, give some insight in to how founder effects are reflected in the genetics of these island populations.

Compared to other published genomic heterozygosity estimates from mammal and avian taxa, the observed heterozygosity in the magpie robins is very low. Familiar examples of low heterozygosity include the cheetah (*Acinonyx jubatus*) and giant panda (*Ailuropoda melanoleuca*). Li *et al*. (2014) ascertained genome–wide heterozygosity from nine bird species from different categories of the Threatened IUCN list. 'Endangered' taxa ranged from 0.00043–0.00060, while 'Least Concern' species ranged from 0.00118–0.00431. Even compared to other endangered birds with a similar history as the Seychelles magpie robin, such as the crested ibis, heterozygosity is low (0.00018). This extremely low level of heterozygosity is considered a 'signature' of endangered species (Li *et al.*, 2014). No large differences were found in the level of heterozygosity between islands, and the general low level observed is more likely a consequence of the earlier genetic bottleneck, rather than of the translocations.

The shape of the SFS for each population is characteristic of a bottlenecked population whose rarer alleles have been lost (Marth *et al.*, 2003). Creating new populations in quick succession over ~two decades has resulted in a rapid population expansion following this bottleneck, most of them private to individuals – which can be a contributing factor to the high number of singletons (alleles that are only found once in the whole population) observed in every population (SFS results, Section 2.3.8, Figure 8). With the dataset and methods used for this project it is important to consider the implications of working with low coverage data. Incorrect genotype assignment is a consequence of missing data or a low

number of reads at a site. In such circumstances, heterozygous genotypes can be mistakenly called homozygous if one of the chromosomes at a diploid sit is missing data or has a low number of reads, it can lead to incorrect genotype calls and an overestimation of diversity. Therefore, this high number of singletons could be partially due to sequencing errors at sites that would otherwise be non–variant homozygotes across the population. While this is a possibility, given the severity of singletons observed, an exaggerated number of heterozygotes called should not have a drastic effect on the general pattern observed in the SFS.

Currently the census population of magpie robins is increasing, and translocating individuals between these already established populations may result in the accompaniment of an increase in genetic diversity with this population growth. When considering translocations to new islands, carefully selecting individuals for genetically informed translocations could allow for factors such as higher levels of heterozygosity and low relatedness to be considered, and could strive towards to the ideal ‘genetic capture’ of 95% of a species’ genetic diversity, to preserve any diversity that exists, using the optimum number of individuals to achieve this (Weeks *et al.*, 2011). Unfortunately, with such a small population the optimum number of donors might not currently be achievable.

Facilitated by the increasing availability of genomic data, it is possible to identify and increasing trend of low genetic diversity patterns in ‘threatened’ species where low diversity does not seem to be compromise survival (Li *et al.*, 2014; Westbury *et al.*, 2016; Robinson *et al.*, 2018; Westbury *et al.*, 2019). However, the focus species of these studies experienced long–term reduced heterozygosity to which they appear to have adapted over time. Contrastingly, the magpie robins suffered a presumably recent and severe reduction in heterozygosity, coupled with subsequent inbreeding (not found to be a factor in aforementioned studies) which would likely make them more vulnerable to the widely recognised reduced fitness characteristics of this genomic signature including susceptibility to disease and maladaptation to changing environments (Frankham, 2005).

Diversity in specific regions of the genome is also found to be more important than genome–wide diversity. For example, the well–known case of the giant panda which exhibits extremely low levels of genome–wide heterozygosity nonetheless harbours relatively high levels of heterozygosity at important immune–function gene regions (Zhu *et al.*, 2013).

2.4.3 Question Three: What is the state of inbreeding in each of the five populations?

While there are several definitions of inbreeding, for the purpose of this question (and corresponding analysis) inbreeding was defined as the mating of closely related individuals.

The picture of local heterozygosity, in all populations, showed regions of the genome with long tracts of diminished heterozygosity, and this influenced the decision to consider ROH analysis. The abundance, length and distribution of ROH can give a glimpse of a 'genealogical timeframe' (Howrigan *et al.*, 2011). The size of tracts coupled with frequency in the sample population allows inferences to be made about population history. Longer tracts are indicative of recent inbreeding. However, frequency must also be considered. High frequency (i.e. observed in the majority of individuals) indicates distant relatives, multiple generations ago, but with little recombination (which could be the case in very inbred individuals), whilst low frequency (i.e. only seen in a few individuals) is highly indicative of true, recent IBD haplotypes. Fewer ancestors also leads to higher proportions of IBD, and more recent inbreeding means less time for the effects of recombination to break up inherited tracts of DNA.

Traditionally, F statistics such as the inbreeding coefficient have been used to estimate what proportion of the genome is IBD. Genomic study now allows the analysis of genome–wide variation and subsequently leads to a more accurate determination of individual inbreeding. However, for this species using F statistics alone for example, would have led to an underestimation in the levels of inbreeding, as half of the individuals had an F score of 0 indicating no alleles are IBD, and this was not apparent from ROH analysis in any individual.

Therefore, using genome–wide markers to assess ROH allowed us to more accurately determine the level of inbreeding that exists in the magpie robins. LD pruning (that is removing sites from analysis that are suspected to be in linkage disequilibrium) is recommended for ROH analysis in PLINK as high LD can lead to high error rates in IBD estimates (Purcell *et al.*, 2007; Hill & Weir, 2011). However, due to the history of this species, many of the SNPs identified in this study were suspected to be in LD and thus pruning for these would have resulted in a depleted dataset that may not have been sufficient enough to attempt to answer this question. However, there may be better developed methods for dealing with such a dataset that need to be explored.

IBS segments are DNA sequences that are identical in two individuals. IBD define those identical DNA sequences in two or more individuals which have been inherited from a common ancestor, or by the same mutation occurring by chance. While PLINK identifies tracts of homozygosity that are IBS, it does not determine whether the homozygous tracts are explicitly due breeding of closely related individuals i.e. IBD. Individuals are usually expected to share short IBD segments as if we look far back enough, we all share a common ancestor (Ceballos *et al*., 2018). Given the sheer volume and length of ROH observed in these birds, it is valid to make assumptions with regards to IBD and inbreeding. Several papers have highlighted the fact that even in outbred populations long ROHs of more than 1Mb and even up to 4Mb can be common (Gibson *et al.*, 2006; Simon–Sanchez, *et al.*, 2007; McQuillan *et al.*, 2008). However, ROHs of 2Mb in length have also been considered an indicator of breeding between distantly related individuals or remnants of past inbreeding (Cassidy *et al.*, 2016) and could be the reason behind the high total ROH at the 1–2Mb category. There are currently no exact figures defined for ROH length determining relatedness of individuals, however in a study that looked at inbreeding in chickens, ROH lengths of 5Mb were found to be due to recent ancestral relatedness (Moyse *et al.*, 2016) and generally this and anything above is considered ‘very’ long. ROH longer than 10Mb

were observed in almost all individuals of the sample populations of the magpie robins, and are strongly indicative of recent inbreeding.

Most studies regarding ROH have been carried out on humans, given the ease and wealth of data available since the International HapMap Project began (International HapMap Consortium (2005), and it has been noted that very little has been done to study ROH from small, endangered population species with populations that have undergone translocations.. It is always important when comparing across species to take this into consideration, however the characteristics of high F_{ROH} and ROH in abundance across the genome of most individuals appears to be a universal indicator of inbreeding.

Very little research has been carried out on patterns of ROH in species with severe inbreeding. The purpose of this analysis was to gain a preliminary understanding of the abundance of ROH across the genome of each individual, which has confidently been carried out. More in–depth analysis will uncover species specific ROH patterns, and island–specific (specific regions shared between individuals of the same population) ROH patterns, with potential to uncover the age of inbreeding events (Grossen *et al.*, 2018). Repeating this method but with the addition of known pedigrees for one or more of the populations (obtained from data that has been collected from long term monitoring) to compare known generations since parental inbreeding with ROH patterns could aid in this understanding.

Although it is clear that each of the populations has experienced inbreeding, at this stage it is uncertain as to whether any of the current populations are suffering from inbreeding depression, or if this genetic pattern is simply characteristic of a species that has been subject to long term inbreeding. In small populations experiencing inbreeding, 'purging selection' can be effective in removing (partially) recessive deleterious alleles from the population, and can thus negate some of the negative effects experienced from inbreeding, as the alleles are not passed down to future offspring (Crnokrak & Barrett, 2002; Garcia–Dorado, 2012). However, this is recognised to be less effective in cases of rapid inbreeding

events (López–Cortegano *et al.*, 2018). The genetic patterns we observe here in the magpie robins, such as diminished heterozygosity and long runs of homozygosity, have been not only been found to be associated with increased deleterious variants (Szpiech, 2013), but were also found as a precursor to extinction in recent genomic studies of extinct species, such as the woolly mammoth (*Mammuthus primigenius*) (Pečnerová, 2018) and may be a cause for concern with regards to the future of the magpie robins species.

While we do not know the 'health' of the magpie robins in relation to the effects of inbreeding, nor whether this characteristic is a longstanding feature in their genetic history due to past events or arose as a result of the recent population bottleneck. Nonetheless, detection and monitoring of inbreeding now may help to prevent future consequences of inbreeding (Grossen *et al.*, 2018).

2.5 Conclusion

The effects of genetic drift can lead to differentiation between populations and loss of variation within populations of the same species and is hypothesised to be the underlying factor responsible for the results obtained from this research. The three islands which were established first are more genetically different when compared to the founding population, likely as there has been more time for the evolutionary forces, such as drift or selection, to act upon the genetic variation in the populations.

It has been argued that serial translocations can greatly reduce the genetic variation of the translocated populations due to repeated founder effects and this previously observed effect is consistent with the findings from my research on the magpie robins as evident by the low overarching genetic diversity. While this research shows translocations have had some impact on the genetic diversity within this species, the consequences of this requires further study.

Inferences with regards to the Aride population can however be made from these findings, as no discernible differences from the other populations were found that could account for the sudden decline in the population size in 2014. Furthermore, although the differences are minimal, the Aride pre-crash population had the highest average level of heterozygosity, and shorter ROH than the other populations, indicating that the birds that perished were relatively ‘healthy’, with health being defined by attributes of the wider magpie robin population.

Having the genome sequenced for every individual would allow for a complete analysis of the genetic make–up of each of the populations and would give a complete picture of exactly what inbreeding is currently happening on each of the islands. This would allow the population to be monitored to ascertain whether inbreeding is increasing, potentially leading to inbreeding depression, and allow such events to be accounted for in management

decisions. Given the severity of the loss of genetic diversity, and how varied it is between individuals, there is a powerful case to develop and implement a long–term genetic monitoring plan, alongside the current management program in place.

This research confirmed well–informed suspicions that the species has experienced inbreeding since the species population bottleneck, and subsequent translocations. The severe degree of inbreeding could not have been guessed and is certainly of more concern than might have been hoped for in terms of progression for the species. It is this which most strongly suggests that the species would benefit from careful genetic monitoring with a specific focus on changing trends and patterns within and between populations.

Chapter 3

General Discussion

3.1 Overview of Project

The Seychelles magpie robin has a small population size, therefore obtaining blood samples for and processing for genome resequencing of half of the population was achievable. In this instance whole genome sequencing and resequencing has allowed us to gain insight into population structure of the magpie robins and to assess inbreeding, providing valuable information that can help influence important management decisions. Overall, the sequencing for this project can be seen as successful. Samples were processed quickly, and most samples (111/159) that were sent for sequencing were received in time to be used in this project and most of these (110/111) could be used for analysis, which allowed for a representative study sample. Most of the remaining samples have since been received (40/48), while one lane is currently being sequenced and all will be incorporated into these analyses prior to publication of this research. Only one sample failed, and this will be re–sequenced in due time. The trade–off in number of samples versus depth of coverage means some samples did exhibit a low coverage. However, Pasaniuc *et al.* (2012) argued that low coverage data still retains most genomic information, particularly when coupled with the bioinformatic tools designed to deal with these data. Although it is still important to consider bias that can be introduced at low coverage, I am confident that this bias was reduced with appropriate filtering of the dataset.

Using whole genome data reduced restrictions imposed by selecting only a few loci to study, such as when using microsatellites, and using more markers allowed for more accurate, reliable results. This is arguably even more important for exploratory analysis on inbred species such as the Seychelles magpie robin because with no prior knowledge or understanding of the magpie robin genome, it could prove difficult to select informative regions to study, that would uncover variant sites in order to investigate genetic structure and patterns.

Broad conclusions can be drawn from these results, however increasing the depth of understanding for this species, for example determining the species–specific mutation rate, would not only allow more specific conclusions to be made about population structure and individual inbreeding, but coupled with sequencing error estimates would also allow a more critical analysis of results. It is best practice to combine results from multiple analyses to test the robustness of the statistics in order to inform population demographics. As such, this research can be considered in–depth and certainly a strong preliminary analysis of the genetic state of the species.

3.2 Applications to Conservation

Translocations of small and/or endangered species are increasing in frequency, yet they remain a contentious issue in conservation. Genetic understanding of both the donor and donated populations are required for effective management strategies, but these considerations are rarely undertaken and monitored either because of time constraints, availability of resources or limited understanding of the impact this can have (Armstrong & Seddon 2008; Jamieson, 2009). Inbreeding has been shown to have had a negative impact on the survival of island populations of passerines (Keller 1998; Jamieson 2009). Extensive research into the impact of inbreeding on small populations determines inbreeding is associated with an increased extinction risk and quantifying individual inbreeding levels are integral for investigation of effects of inbreeding to inform active management of populations. Angeloni *et al*. (2011) argued genomics may not yet prove any more useful to applied conservation than traditional genetic analysis. However, using genome wide markers with ROH has given more accuracy and understanding than microsatellites and F statistics alone could have achieved for such an inbred species. The possibilities in which this research, and indeed continued work on this species, can be applied to the proposed conservation actions are vast. Not only will the insight gained from this research be immediately considered in the decision to translocate more birds to Aride island (SMART meeting, 2018), but also for future translocations to new islands. The selection of individuals using a molecular insight provided

by this research is expected to improve translocation success by means of increased genetic diversity. The molecular understanding of inbreeding gained from this project can now be used to make genetically informed decisions to translocate the least related birds when establishing new populations, thus reducing the risk of further inbreeding, in addition to the continued habitat restoration and invasive species removal conservation efforts.

A meta–analysis of 'genetic rescue' conservation actions (translocating individuals of the same species between already established populations, when these populations are isolated and little gene flow exists between them) found mostly positive benefits arose from diversification of the gene pool, and such actions seemed to combat the long–term effects of reintroductions through single event translocations (Frankham, 2015). There is possibility for such rescue to be carried out with the magpie robins, however, it is important to note risks associated with this such as spreading unknown disease and death of the donated individuals from stress and a failure to settle in the new environment.

In 2005 the Seychelles magpie robin was downgraded from 'critically endangered' to 'endangered' as the census population exceeded a threshold of 50 individuals. However, this was done with no genetic understanding of the species. The rapidly changing environment we are facing and the vulnerability of oceanic islands, coupled with the suspected lack of ability this magpie robins may have to adapt to these changes, could indicate a higher extinction risk than the species is currently categorised for. While we would usually celebrate a reclassification considering a species to be less endangered, funding and research efforts were seen to plateau with this change in conservation status. With a species so small, and so fragile, a lack of funds and resources to continue efforts to help save the species could further increase the danger the species faces. If a species is more vulnerable than currently defined, due to factors such as genetics, this needs to be taken into account in its classification.

3.3 Future Research

As previously outlined, long autozygous tracts, low heterozygosity and reduced genetic diversity are associated with reduced fitness as a result of deleterious recessive alleles accumulating in the population. The genetic state of the Seychelles magpie robin is characteristic of this, and it is therefore imperative to conduct further research to investigate if any of the current populations are suffering from, or are at risk of suffering from, inbreeding depression, particularly in light of the recent population crash on Aride. Measuring and understanding inbreeding depression is important in species conservation, particularly when those species can be considered at high risk of extinction due to small and isolated populations (Grossen *et al.*, 2018). Long–term monitoring data collected for the species, includes nesting attempts, breeding success, fledgling success and individual mortality, which can be key indicators of inbreeding depression. This monitoring data, coupled with the genetic understanding gained from this project, will improve our knowledge of the fitness, behaviour and history of this species, as done in previous research on different taxa (Boulanger *et al.*, 2004; Richardson *et al.*, 2005; Bichet *et al.*, 2018).

An interesting area for future research is the comparison of historical DNA, from samples collected before the bottleneck in the mid–1900s, with the modern samples analysed in this project. A preliminary historical effective population size analysis (not shown) indicates a large population size in the past. Assessing historical DNA may offer an insight into the evolutionary pressures faced by the Seychelles magpie robin and could indicate whether the species suffered any previous bottlenecks. Additionally, mitochondrial DNA sequences can be extracted from the raw reads generated from whole genome resequencing of the magpie robins and used alongside nuclear information to investigate population changes in the past. There is fairly good documentation of when and where magpie robins were executed and collected back to the 1800s, and resident samples at museums around the world have already been identified. Incorporating these samples may give insight into if and how this

recent bottleneck affected the genetic diversity of the species, as successfully undertaken with the Seychelles warblers (Spurgin *et al.*, 2014). Comparing modern DNA with historical diversity (samples collected before the bottleneck in the 1960's) may also give an indication of the extinction risk faced by the modern–day magpie robin (Bouzat *et al*., 2009; Diez–del–Molino *et al.*, 2017).

References

Andersen, L., Fog, K. & Damgaard, C. (2004) Habitat fragmentation causes bottlenecks and inbreeding in the European tree frog (*Hyla arborea*). *Proceedings of the Royal Society of London B*, **271**:1293–1302

Agilent Technologies Inc. (2015) Agilent Genomic DNA ScreenTape System Quick Guide. Online version available at https://www.agilent.com/cs/library/usermanuals/public/ScreenTape_gDNA_QG.pdf

Allendorf, F., Hohenlohe, P. & Luikart, G. (2010) Genomics and the future of conservation genetics. *Nature Reviews. Genetics*, **11:**697–709

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

Angeloni, F., Wagemaker, N., Vergeer, P. & Ouborg, J. (2012) Genomic toolboxes for conservation biologists. *Evolutionary Applications*, **5**:130–143

Armstrong, D. & Seddon, P. (2008) Directions in reintroduction biology. *Trends in Ecology & Evolution.* **23**:20–25

Beltman, J., Urbanus, J., Velds, A., van Rooij, N., … & Schumacher, T. (2016) Reproducibility of Illumina platform deep sequencing errors allows for accurate determination of DNA barcodes in cells. *BMC Bioinformatics,* **17**:151

Bentley, D., Balasubramanian, S., Swerdlow, H., Smith, G., Milton, J., … & Smith A. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **456**:53–59

BGI (2019) NGS Platforms available at BGI https://www.bgi.com/us/resources/sequencing–platforms/ [Downloaded on 10 May 2019]

Bichet, C., Vedder, O., Sauer-Gurth, H., Becker, P, Wink, M. & Bouwhuis, S. (2018) Contrasting heterozygosity-fitness correlations across life in a long-lived seabird. Molecular Ecology, **28**:671–685

Birdlife International (2008) Preventing Extinctions Brochure. http://www.birdlife.org/sites/default/files/attachments/preventing_extinctions_brochure.pdf [Downloaded on 18 May 2019]

BirdLife International (2016) *Acrocephalus sechellensis*. The IUCN Red List of threatened species 2016: e.T22714882A94431883. [Downloaded on 28 March 2018]

BirdLife International (2017) *Copsychus sechellarum* (amended version of 2016 assessment). The IUCN Red List of threatened species 2017: e.T22709978A111061160. [Downloaded on 01 April 2018]

Blackburn, C. (1878) Quelques courtes descriptions des oiseaux de Praslin. *Transactions of the Royal Society of Arts and Sciences of Mauritius*, **12**:125–1

Boulanger, J., Himmer, S. & Swan, C. (2004) Monitoring of grizzly bear population trends and demography using DNA mark–recapture methods in the Owikeno Lake area of British Columbia. *Canadian Journal of Zoology*, **82**:1267–1277

Bouzat, J., Johnson, J., Toepfer, J., Simpson, S., Esker, T. & Westemeier, R. (2009) Beyond the beneficial effects of translocations as an effective tool for the genetic restoration of isolated populations. *Conservation Genetics,* **10**:191–201

Broman, K. & Weber, J. (1999) Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *American Journal of Human Genetics*, **65**:1493–1500

Burt, A., Gane, J., Olivier, I., Calabrese, L., De Groene, A., Liebeck, T., Marx, D. & Shah, N. (2016) The history, status and trends of the endangered Seychelles magpie–robin. *Bird Conservation International,* **26**:505–523

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S, Sinding, M., … & Gilbert, T. (2017) Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution*, **9**:410–419

Cassidy, L., Martiniano, R., Murphy, E., Teasdale, M., Mallory, J., Hartwell, B. & Bradley, D. G. (2016) Neolithic and Bronze Age migration to Ireland and Establishment of the Insular Atlantic genome. *Proceedings of the National Academy of Sciences*, **113**:368–373

Ceballos, F., Joshi, P., Clark, D., Ramsay, M. & Wilson, J. (2018) Runs of homozygosity: Windows into population history and trait architecture. *Nature Reviews Genetics*, **19**:220–234

Charlesworth, B. & Charlesworth, D. (1987) Inbreeding Depression and its Evolutionary Consequences. *Annual Review of Ecological Systems*, **18**:237–68

Charlesworth, D. & Willis, J. (2009) The genetics of inbreeding depression. *Nature Reviews. Genetics*, **10**:783–796

Covaris Inc. (2015) Quick guide: DNA shearing with M220 focused-ultrasonicator. Online version available at: https://covaris.com/wp-content/uploads/pn_010252.pdf

Crnokrak, P. & Barrett, H. (2002) Perspective: purging the genetic load: a review of the experimental evidence. *Genetics*, **190**:1461–1476

Dalloul, R., Long, J., Zimin, A., Aslam, L., Beal, K., Le Ann, B., … & Reed, K. (2010) Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): Genome assembly and analysis. *PLoS Biology*, **8**:1–21

Dawson, P. (1965) Bristol University Seychelles Expedition 6–Frégate, home of the magpie robin. *Animals*, **7**:520–2

Diez–del–Molino, D., Barreiro, F., B., Barnes, I., Gilbert, T. & Dalen, L. (2018) Quantifying temporal genomic erosion in endangered species. *Trends in Ecology and Evolution*, **33**:176–185

Edwards, J. (Undated) Conservation of the Seychelles Magpie Robin: the development of non–hazardous control methods for populations of cockroaches (*Periplaneta Americanea*) and ants (*Monomorium destructor*) infesting human habitation on Frégate Island.

Ekblom, R. & Wolf, J. (2014) A field guide to whole–genome sequencing, assembly and annotation. *Evolutionary Applications*, **7**:1026–1042

Ellegren, H. & Fridolfsson, A. (1999) A simple and universal method for molecular sexing of non–ratite birds. *Journal of Avian Biology,* **30**:116–121

Frankham, R. (1997). Do island populations have less genetic variation than mainland populations? *Heredity,* **78**:311–27

Frankham, R. (2005) Stress and adaptation in conservation genetics. *Journal of Evolutionary Biology*, **18**:750–755

Frankham, R., Briscoe, D. & Ballou, J. (2010) *Introduction to conservation genetics,* 2nd edition. Cambridge, UK: Cambridge University Press

Frankham, R., Bradshaw, C. & Brook, B. (2014) Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, **170**:56–63

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

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

García–Gámez, E., Sahana, G., Gutiérrez–Gil, B. & Arranz, J. (2012) Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC. Genetics*, **13**:43

Gaymer, R., Blackman, R., Dawson, P., Penny, M. & Penny, C. (1969) The endemic birds of the Seychelles. *Ibis,* **111**:157–176

Gibson, J., Morton N. & Collins, A. (2006) Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics*, **15**:789–795

Gray, M., Granka, J., Bustamante, C., Sutter, N., Boyko, A., Zhu, L., Ostrander, E. & Wayne, R. (2009) Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics*, **181**:1493–1505

Gregory, T. (2018) Animal Genome Size Database. http://www.genomesize.com. [Downloaded on 17 July 2018]

Grossen, C., Biebach, I., Angelone–Alasaad, S., Keller, L. & Croll, D. (2018) Population genomics analyses of European ibex species show lower diversity and higher inbreeding in reintroduced populations. *Evolutionary Applications*, **11**:123–139

Haanes, H., Markussen, S., Herfindal, I., Røed, K., Solberg, E., Heim, M., … & Sæther, B. (2013) Effects of inbreeding on fitness–related traits in a small isolated moose population. *Ecology and Evolution*, **3**:4230–4242

Hartlaub, G. (1877) *Die Vögel Madagascars und der benachbarten Inselgruppen: ein Beitrag zur Zoologie der äthiopischen Region*. Halle, Germany: HW Schmidt

Herfindal, I., Haanes, H., Røed, K., Solberg, E., Markussen, S., Heim, M. & Sæther, B. (2014) Population properties affect inbreeding avoidance in moose. *Biology Letters,* **10**:20140786

High, J. (1974) Seychelles magpie robin: coming back from the brink of extinction? *Wildlife*, **16**:61

Hill, W. & Weir, B. (2011) Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research*, **93**:47–64

Holderegger, R., Kamm, U. & Gugerli, F. (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*, **21**:797–807

Howrigan, D., Simonson, M. & Keller, M. (2011) Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics*, **12**:460

Hudson, R. (1987) Estimating the recombination parameter of a finite population model without selection. *Genetics Research,* **50**:245–250

Illumina Inc. (2017) *An introduction to Next–Generation Sequencing Technology*. https://www.illumina.com/science/technology/next–generation–sequencing/paired–end–vs–single–read–sequencing.html [Downloaded on 10 January 2018]

International HapMap Consortium (2005) A haplotype map of the human genome. *Nature*, **437**:229–1320

International Human Genome Sequencing Consortium (2001) International Human Genome Consortium. Initial sequencing and analysis of the human genome. *Nature,* **409**:860–921

IUCN (2017) The IUCN Red List of threatened species: Summary statistics http://www.iucnredlist.org/about/summary–statistics [Downloaded on 28 March 2018]

Jamieson, I. (2009) *Loss of genetic diversity and inbreeding in New Zealand's threatened bird species*. Wellington, New Zealand: New Zealand Department of Conservation.

Jobling, M., Tyler–Smith, C., Hollox, E., Hurles, M. & Kivisild, T. (2015) *Human Evolutionary Genetics*, 2nd edition. New York, USE: Garland Science

Kardos, M., Taylor, H., Ellegren, H., Luikart, G. & Allendorf, F. (2016) Genomics advances the study of inbreeding depression in the wild. *Evolutionary Applications*, **9**:1205–1218

Kardos, M., Åkesson, M., Fountain, T., Flagstad, Ø., Liberg, O., Olason, P., … & Ellegren, H. (2018) Genomic consequences of intensive inbreeding in an isolated wolf population. *Nature Ecology and Evolution*, **2**:124–131

Keller, L. (1998) Inbreeding and its fitness effects in an insular population of song sparrows (*Melospiza melodia*). *Evolution*, **52**:240–250

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

Keller, M., Visscher, P. & Goddard, M. (2011) Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics,* **189**:237–249

Kimura, M. (1983) The Neutral Theory of Molecular Evolution. Cambridge: Cambridge University Press; 1983

Kirin, M., McQuillan, R., Franklin, C., Campbell, H., McKeigue, P. & Wilson, J. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE*, **5**:e13996

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

Lamichhaney, S., Han, F., Berglund, J., Wang, C, Almén, M., … & Andersson, L. (2016) A beak size locus in Darwin's finches facilitated character displacement during a drought. *Evolutionary Genetics,* **352**:470–474

Lande, R. (1988) Genetics and demography in biological conservation. *Science*, **241**:1455–1460

Lelieveld, S., Spielmann, M., Mundlos, S., Veltman, J. & Gilissen, C. (2015) Comparison of exome and genome sequencing technologies for the complete capture of protein–coding regions. *Human Mutation,* **36**:815–822

Li, H. (2011) Improving SNP discovery by base alignment quality. *Bioinformatics*, **27**:1157–1158

Li, S., Li, B., Cheng, C., Xiong, Z., Liu, Q., … & Yan, J. (2014) Genomic signatures of near–extinction and rebirth of the crested ibis and other endangered bird species. *Genome Biology*, **15**:557

Lim, H., Zou, F., Taylor, S., Marks, B., Moyle, R., Voelker, G. & Sheldon, F. (2010) Phylogeny of magpie–robins and shamas (Aves: Turdidae: Copsychus and Trichixos): Implications for island biogeography in Southeast Asia. *Journal of Biogeography*, **37**:1894–1906

Liu, L., Li, Y., Li, S., Hu, N., He, Y., … & Law, M. (2012) Review Article: Comparison of Next–Generation Sequencing Systems *Journal of Biomedicine and Biotechnology*, **2012**:11

Lucking, R. & Ayrton, V. (1994) *The Seychelles Magpie Robin Recovery Plan. Aride Translocation: Progress report 11*. Mahé, Seychelles: Nature Seychelles (Unpublished)

López–Cortegano, E., Bersabé, D., Wang, J. & García–Dorado, A. (2018) Detection of genetic purging and predictive value of purging parameters estimated in pedigreed populations. *Heredity,* **121**:38–51

López–Sepulcre, A., Norris, K. & Kokko, H. (2009) Reproductive conflict delays the recovery of an endangered social species. *Journal of Animal Ecology*, **78**:219–225

Marth, G., Czabarka, E., Murvai, J. & Sherry, S. (2004) The allele frequency spectrum in genome–wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics,* **166**:351–372

Matzaraki, V., Kumar, V., Wijmenga, C. & Zhernakova, A. (2017) The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biology,* **18**:76

Maxam, A., & Gilbert, W. (1977) A New Method for Sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America,* **74**:560–564

McMahon, B., Teeling, E. & Höglund, J. (2014) How and why should we implement genomics into conservation? *Evolutionary Applications*, **7**:999–1007

McQuillan, R., Leutenegger, A., Abdel–Rahman, R., Franklin, C., Pericic, M., … & Wilson, J. (2008) Runs of homozygosity in European populations. *American Journal of Human Genetics*, **83**:359–372

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

Millett, J. & Shah, N. (2000) *Seychelles magpie–robin recovery program. Review of progress 2000*. Mahé, Seychelles: Birdlife Seychelles (Unpublished)

Mills, S. & Allendorf, F. (1996) The one-migrant-per-generation rule in conservation and management. *Conservation Biology,* **10**:1509–1518

Montell, H., Fridolfsson, A. & Ellegren, H. (2010) Contrasting levels of nucleotide diversity on the avian Z and W sex chromosomes. *Molecular Biology and Evolution,* **18**:2010–2016

Motahari, A., Bresler, G. & Tse, D. (2012) Information Theory of DNA Sequencing. *IEEE Transactions on Information Theory,* **59:**6273–6289

Moyse, E., Nguyen, V., Vu, D. & Farnir, F. (2016) Inbreeding and genomic runs of homozygosity in Vietnamese local chickens. Scientific congresses and symposiums: poster. Available at: http://hdl.handle.net/2268/207658 [Downloaded on 15 April 2019]

Nature Seychelles (2018) Seychelles magpie robin. http://natureseychelles.org/knowledge–centre/seychelles–wildlife–2/endemic–bird–species/46–seychelles–magpie–robin [Downloaded on 20 April 2018]

Nature Seychelles (2019) *Seychelles magpie robin long-term monitoring data.* Mahé, Seychelles: Nature Seychelles (Unpublished)

Newton, A. (1865) On an Apparently Undescribed Bird from the Seychelles Islands. *Ibis*, **1**:331–7

Newton, A. (1867) On the landbirds of the Seychelles Archipelago. *Ibis,* **9**:335-360

Nielsen, R., Paul, J., Albrechtsen, A. & Song, Y. (2011) Genotype and SNP calling from next–generation sequencing data. *Nature Reviews. Genetics*, **12**:443–451

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y. & Wang, J. (2012) SNP calling, genotype calling, and sample allele frequency estimation from new–generation sequencing data. *PLoS ONE*, **7**:e37558

Pasaniuc, B., Rohland N., McLaren P., Garimella K., Zaitlen N., … & Price, K. (2012) Extremely low–coverage sequencing and imputation increases power for genome–wide association studies. *Nature Genetics,* **44**:631–635

Parr, S. (1998) *The history of Seychelles Magpie Robin translocations as a guide to future actions: a position paper*. Mahé, Seychelles: BirdLife Seychelles (Unpublished)

Pečnerová, P. (2018) *Genomic analysis of the process leading up to the extinction of the woolly mammoth*. Doctoral thesis: Stockholm University

Pemberton, T., Absher, D., Feldman, M., Myers, R., Rosenberg, N. & Li, J. (2012) Genomic patterns of homozygosity in worldwide human populations. *American Journal of Human Genetics*, **91**:275–292

Pfeiffer, F., Gröber, C., Blank, M., Händler, K., … & Mayer, G. (2018) Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific Reports,* **8**:10950

Pool, J., Hellmann, I., Jensen, J. & Nielsen, R. (2010) Population genetic inference from genomic sequence variation. *Genome Research*, **20**:291–300

Purcell, S., Neale, B., Todd–Brown, K., Thomas, L., Ferreira, M., … & Sham, P. (2007) PLINK: A Toolset for whole–genome association and population–based linkage analysis. *American Journal of Human Genetics,* **81**:559–575

Qiagen Inc. (2006) DNeasy blood and tissue handbook. Online version available at: https://www.qiagen.com/us/resources/resourcedetail?id=6b09dfb8–6319–464d–996c–79e8c7045a50&lang=en

R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051–07–0, URL http://www.R–project.org

Ramstetter, M., Shenoy, S., Dyer, T., Lehman, D., Curran, J., Duggirala, R. … & Williams, A. (2018) Inferring identical–by–descent sharing of sample ancestors promotes high–resolution relative detection. *American Journal of Human Genetics*, **103**:30–44

Redford, K., Amato, G., Baillie, J., Beldomenico, P., Bennett, E., … & Thorbjarnarson, J. (2011) What does it mean to successfully conserve a (vertebrate) species? *BioScience*, **61**:39–48

Richardson, D., Komdeur, J., Burke, T. & Schanz, T. (2005) MHC–based patterns of social and extra–pair mate choice in the Seychelles warbler. *Proceedings of the Royal Society B,* **272**:759–767

Ridgeway, R. (1895) On birds collected by Dr. W.L. Abbott in the Seychelles, Amirantes, Gloriosa, Assumption, Aldabra, and Adjacent Islands, with notes on habits etc., by the collector. *Proceedings of the U.S. Natural History Museum,* **18**:509–546

Robinson, J., Brown, C., Kim, B., Lohmueller, K. & Wayne, R. (2018) Purging of strongly deleterious mutations explain long–term persistence and absence of inbreeding depression in island foxes. *Current Biology,* **28**:3487–3494

Sanger, F., Nicklen, S. & Coulson, A. (1977) DNA Sequencing with chain–terminating Inhibitors. *Proceedings of the National Academy of Science of the United States of America,* **74**:5463–5467

Sangster, G., Alström, P., Forsmark, E. & Olsson, U. (2010) Multi–locus phylogenetic analysis of Old World chats and flycatchers reveals extensive paraphyly at family, subfamily and genus level (Aves: Muscicapidae). *Molecular Phylogenetics and Evolution*, **57**:380–392

Schubert, M., Ermini, L., Sarkissian, C., Jónsson, H., Ginolhac, A., … & Orlando, L. (2014) Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX**.** *Nature Protocols,* **9**:1056–1082

Shah, N. (2002) Garden of Eden: Saving the magpie robin http://www.pbs.org/wgbh/nova/eden/robin.html [Downloaded on 24 January 2018]

Shah, N. & Parr, S. (2002) *Translocation of Seychelles magpie robins to Aride 2002–Progress of the translocated population January 2002 – October 2002*. Mahé, Seychelles: BirdLife Seychelles (Unpublished)

Simon–Sanchez, J., Scholz, S., Fung, H., Matarin, M., Hernandez, D., … & Singleton, A. (2007) Genome–wide SNP assay reveals structural genomic variation, extended homozygosity and cell–line induced alterations in normal individuals. *Human Molecular Genetics,* **16**:1–14

Sims, D., Sudbery, I., Ilott, N., Heger, A. & Ponting, C. (2014) Sequencing depth and coverage: Key considerations in genomic analyses. *Nature Reviews Genetics*, **15**:121–132

Skerrett, A., Bullock, I. & Disley, A. (2000) *Birds of Seychelles*. London, UK: Helm Books

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

SMART (2018) *SMART March bi–annual meeting minutes: magpie robin island updates*. Mahé, Seychelles: Nature Seychelles (Unpublished)

SMART (2019) *SMART March bi–annual meeting minutes: magpie robin island updates.* Mahé, Seychelles: Nature Seychelles (Unpublished)

Snustad, P. & Simmons, M. (2011) *Principles of Genetics*, 6th edition. Jefferson City, USA: John Wiley & Sons

Spurgin, L., Wright, D., van der Velde, M., Collar, N., Komdeur, J., Burke, T. & Richardson, D. (2014) Museum DNA reveals the demographic history of the endangered Seychelles warbler. *Evolutionary Applications*, **7**:1134–1143

Supple, M. & Shapiro, B. (2018) Conservation of biodiversity in the era of genomics. *Genome Biology,* **19**:131

Szpiech, Z., Xu, J., Pemberton, T., Peng, W., Zollner, S. & Rosenbery, N. (2013) Long runs of homozygosity are enriched for deleterious variation. *American Journal of Human Genetics,* **93**:90–102

ThermoFisher Scientific Inc. (2012) Thermo Scientific KingFisher Blood DNA Kit Instruction Manual. Online version available at: https://static.thermoscientific.com/images/D01160~.pdf

ThermoFisher Scientific Inc. (2015a) Qubit® dsDNA BR Assay Kits. Online version available at: https://www.thermofisher.com/document–connect/document–connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS–Assets%2FLSG%2Fmanuals%2FQubit_dsDNA_HS_Assay_UG.pdf&title=VXNlciBHdWlkZTogUXViaXQgZHNETkEgSFMgQXNzYXkgS2l0cw==

ThermoFisher Scientific Inc. (2015b) Qubit® dsDNA HS Assay Kits. Online version available at: https://www.thermofisher.com/document–connect/document–connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS–Assets%2FLSG%2Fmanuals%2FQubit_dsDNA_HS_Assay_UG.pdf&title=VXNlciBHdWlkZTogUXViaXQgZHNETkEgSFMgQXNzYXkgS2l0cw==

Vesey–Fitzgerald, D. (1940) The birds of the Seychelles 1: the endemic birds. *Ibis,* **14**:480–489

Vieira, F., Fumagalli, M., Albrechtsen, A. & Nielsen, R. (2013) Estimating inbreeding coefficients from NGS data: impact on genotype calling and allele frequency estimation. *Genome Research,* **23**:1852–1861

Watson, J. (1978*) The Seychelles magpie robin. WWF Project 1590: Endangered land birds, Seychelles (Final Report 1(a)).* Mahé, Seychelles: Birdlife Seychelles (Unpublished)

Watson, J., Warman, C., Todd, D. & Laboudallon, V. (1992) The Seychelles magpie robin *Copsychus sechellarum*: ecology and conservation of an endangered species. *Biological Conservation*, **61**:93–106

Watterson, G. (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**:256–276

Weeks, A., Sgro, C., Young, A., Frankham, R., Mitchell, N. & Hoffmann, A. (2011) Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications,* **4**:709–725

Westbury, M., Hartman, S., Barlow, A., Wiesel, I., Leo, V., … & Hofreiter, M. (2016) Extended and continuous decline in effective population size results in low genomic diversity in the world's rarest hyena species, the brown hyena. *Molecular Biology and Evolution,* **35**:1225–1237

Westbury, M., Petersen, B., Garde, E., Heide–Jørgensen, M. & Lorenzen, E. (2019) Narwhal genome reveals long–term low genetic diversity despite current large abundance size. *iScience*

Whelan, C., Şekercioğlu, Ç. & Wenny, D. (2015) Why birds matter: from economic ornithology to ecosystem services. *Journal of Ornithology,* **156**:227–238

Williams, G. (1966) *Adaptation and natural selection* A *critique of some current evolutionary thought*. Princeton, USA: Princeton University Press

Wilson, J. & Wilson, R. (1978) Observations on the Seychelles magpie robin (*Copsychus sechellarum)*. *Bulletin of the British Ornithologists Club,* **98**:15–21

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

Wright, S. (1931) Evolution in Mendelian populations. *Genetics,* **16:**97–159

Zacheis, A., Ruess, R. & Hupp, J. (2002) Nitrogen dynamics in an Alaskan salt marsh following spring use by geese. *Oecologia*, **130**:600–608

Zhang, G., Li, B., Li, C., Gilbert, M. T. P., Jarvis, E. D. & Wang, J. (2014a) Comparative genomic data of the avian phylogenomics project. *GigaScience*, **3**:26

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D., … & Wang, J. (2014b). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science,* **346**:1311–1320

Zhu, L. & Bustamante, C. (2005) A composite–likelihood approach for detecting directional selection from DNA sequence data. *Genetics,* **170**:1411–1421

Zhu, Y., Wan, Q., Yu, B., Ge, Y. & Fang, S. (2013) Patterns of genetic differentiation at MHC class I genes and microsatellites identify conservation units in the giant panda. *BMC Evolutionary Biology,* **13**:227

Zuur, A., Ieno, E. & Elphick, C. (2010) A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution,* **1**:3–14

List of Abbreviations

BAM – binary aligned map

BGI – Beijing Genomics Institute

BWA – Burrow–Wheeler aligner

CHD – chromodomain–helicase–DNA–binding protein

DNA – deoxyribose nucleic acid

DNB – DNA nanoball

GATK – Genome Assembly Tool Kit

Gb – gigabase/s

GLM – Generalised Linear Models

HWE – Hardy–Weinburg Equilibium

IBD – identical by descent

IBS – identical by state

INDEL – insertion/deletion

IUCN – International Union for Conservation of Nature

Kb – kilobase/s

LD – linkage disequilibrium

MAF – minor allele frequency

Mb – megabase/s

NIOO – Netherlands Institute of Ecology

NGS – next generation sequencing

PC – principal component

PCA – principal component analysis

PCR – polymerase chain reaction

qPCR – quantitative PCR

RAD – restriction rite associated DNA

ROH – runs of homozygosity

RSPB – Royal Society for the Protection of Birds

SE – standard error

SFS – site frequency spectrum

SNP – single nucleotide polymorphism

SMART – Seychelles Magpie Robin Recovery Team

Appendices

Additional information supporting sections of the main text is documented in this section. Included in the appendices are examples of results from the DNA extraction and library building processes. There were no major deviations between individual samples from the example results provided. Where a subset of results has been supplied, this is indicated and full results can be provided upon request, but as there are hundreds of files, some of which are over a dozen pages, they have not been included to reduce paper waste.

CHAPTER 1

Appendix 1 Prerequisites for Seychelles magpie robin translocations, adapted from Burt *et al*. (2016).

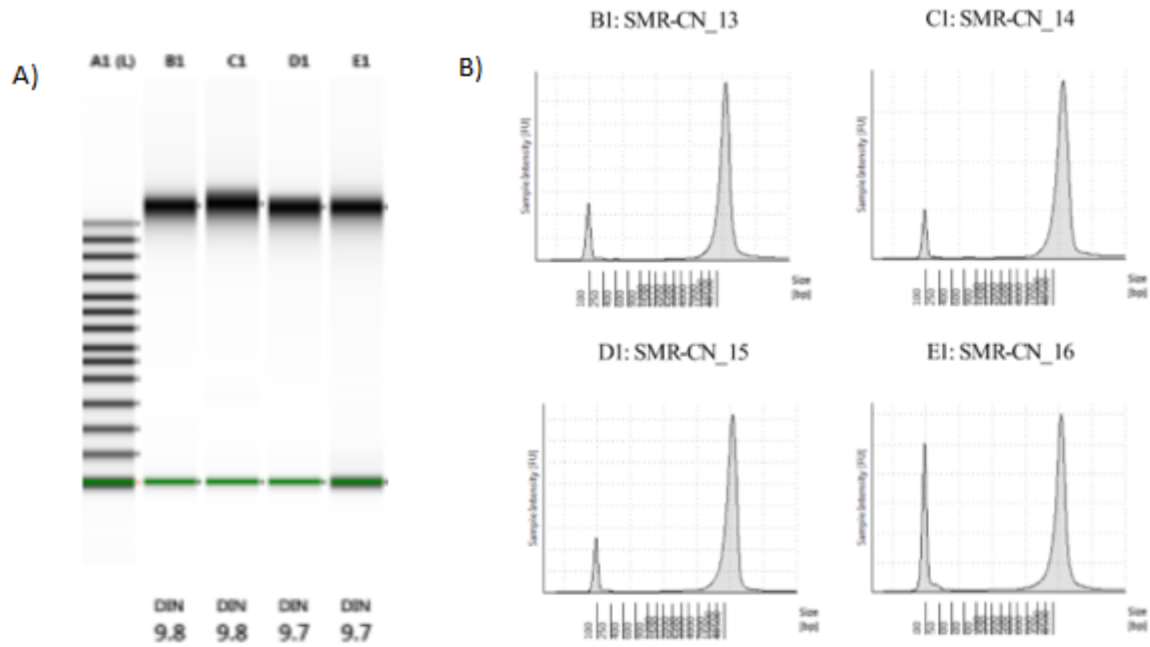
Desirable long–term post–translocation commitments	
1	Providing at least two nest boxes per territory. Following the provision of open–fronted nest boxes on islands, nesting success improved and the fledging rate from nest boxes was 44% compared to 28% from coconut palm crowns and 38% from natural nest sites. (Lucking and Lucking ,1997).
2	Supplementary feeding and water supply to birds after initial release as necessary.
3	Employment of a conservation officer responsible for the close monitoring of magpie robins. This role includes compiling monthly reports to be submitted along with data sheets to the SMART coordinator.
4	Daily observations conducted following the magpie–robin monitoring protocol set out first by Lucking and Lucking (1997), then by Bristol *et al*. (2005).
5	A strict biosecurity protocol, including continued eradication of invasive alien predators and competitors, such as cats, rats *Rattus* spp., Barn Owl (*Tyto alba)* and Common Myna (*Acridotheres tristis)*.
6	Ringing, biometrics and blood sampling of each fledgling to be carried out.
7	Active habitat management to maintain suitable Magpie–robin habitat.
8	Ban on the use of organophosphate and carbamide pesticides.

Appendix 2 Sexing PCR optimisation using Fridolfsson primers (2550F/2718R): A) Mastermix reagent volumes required for a single sample and B) Optimised PCR profile for the Seychelles magpie robin.

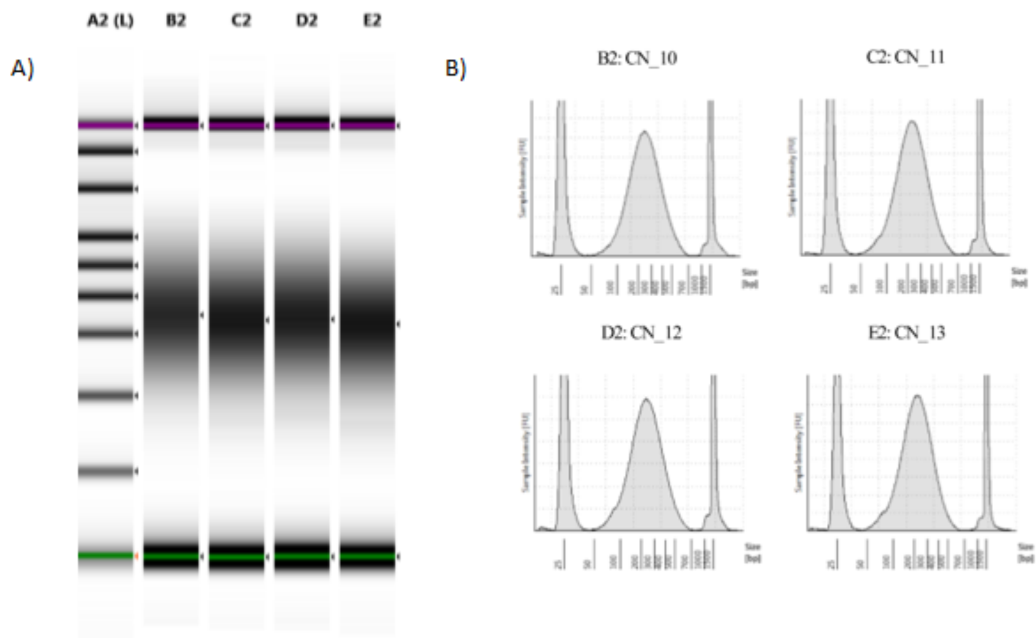
Mastermix Recipe (1x)	
Reagent	**Volume (µL)**
H_2O	12.3
Buffer with MgCl2 15nM	4
dNTPs 2nM	0.5
2550F 10nM	0.5
2718R 10nM	0.5
AmpliTaq	0.2
Vmix	18
DNA	2

PCR profile			
Stage	**Temperature**	**Time**	**Cycles**
Denature	95˚ C	3mins	–
Anneal	95˚ C	30secs	40
	46˚ C	30secs	
	72˚ C	40secs	
Extend	72˚ C	7mins	–
–	8˚ C	∞	–

Appendix 3 A snapshot of results using the Tapestation 2200 (protocol: genomic DNA) used to assess any degradation in the initial DNA samples extracted from blood. A) shows samples demonstrated little degradation, evident by the band observed on the largest fragment size of the ladder (65,000bp) B) is the graph produced from the gel results visualising the distribution of fragment sizes.



Appendix 4 A snapshot example of Covaris fragmentation results. A) shows the results of the gel run where the darker band is the result of the accumulation of biological material meaning there are more fragments of that size and B) shows the graphs produced from the results of the gel run giving the distribution of fragment sizes.



Appendix 5 This figure shows an example of the qPCR reagents and results. A) Samples were pooled in groups of eight for sequencing. Prior to the genomic libraries being sent to BGI, indices were attached to the fragments of each library, using the given mastermix recipe, in order to identify and 'separate' the reads of each individual sample in the pool post–sequencing. Volumes of reagents needed are for one sample and were adjusted according to how many samples were used. B) A qPCR was used to determine the optimum number of cycles for indexing the libraries, to attach sufficient index without over amplifying the sample. The number of cycles was determined by the initial plateau for each sample individually.

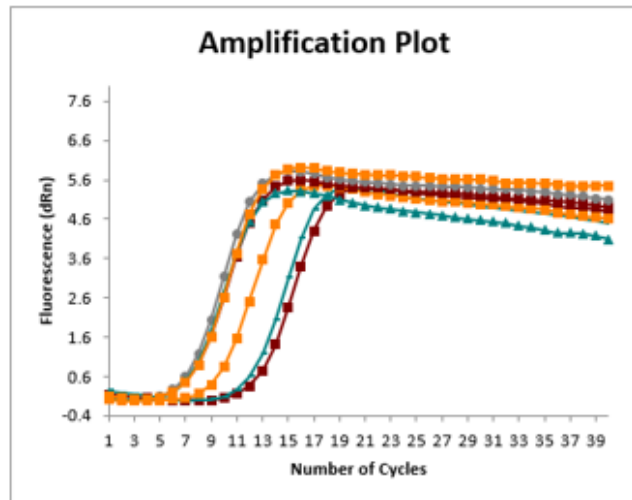
A)

Reagent	Volume (µL)	
dH20	13.08	
10X Buffer	2	1x
25 mM MgCl2	2	2.5 mM
10 µM F primer (IS4)	0.4	0.2 uM
10 µM R primer (IS4)	0.4	0.2 uM
SYBR green	0.8	
25 mM dNTPs	0.16	0.2 mM
TaqGold (5U/µL)	0.16	0.1 U/µL
Mix Total	19	
Library (1:20)	1	
Total reaction	20	

TaqGold PCR profile:

95°C	10 min	
95°C	30s	
60°C	60s	x40 cycles
72°C	60s	

B)



Appendix 6 A table listing the original DNA extraction information for each sample, and subsequent amplified library concentrations before pooling.

Sample	Extraction conc. (ng/μL)	Amplified Library conc. (ng/μL)
A_07	181.0	9.9
A_08	11.6	11.7
A_09	4.2	1.9
A_10	9.3	13.3
A_11	21.0	8.7
A_12	4.5	15.6
A_13	13.0	7.3
A_14	15.1	10.0
A_15	9.7	11.8
A_16	9.3	13.1
A_17	8.5	13.1
A_18	11.7	7.6
A_19	7.3	7.4
A_20	26.6	12.4
A_21	21.8	8.9
A_22	4.3	11.0
A_23	33.8	10.5
A_24	24.8	11.4
A_25	27.4	12.2
A_26	14.0	10.2
A_27	20.4	12.9
A_28	13.7	9.1
A_29	9.6	8.3
A_30	5.8	14.2
A_31	4.4	12.6

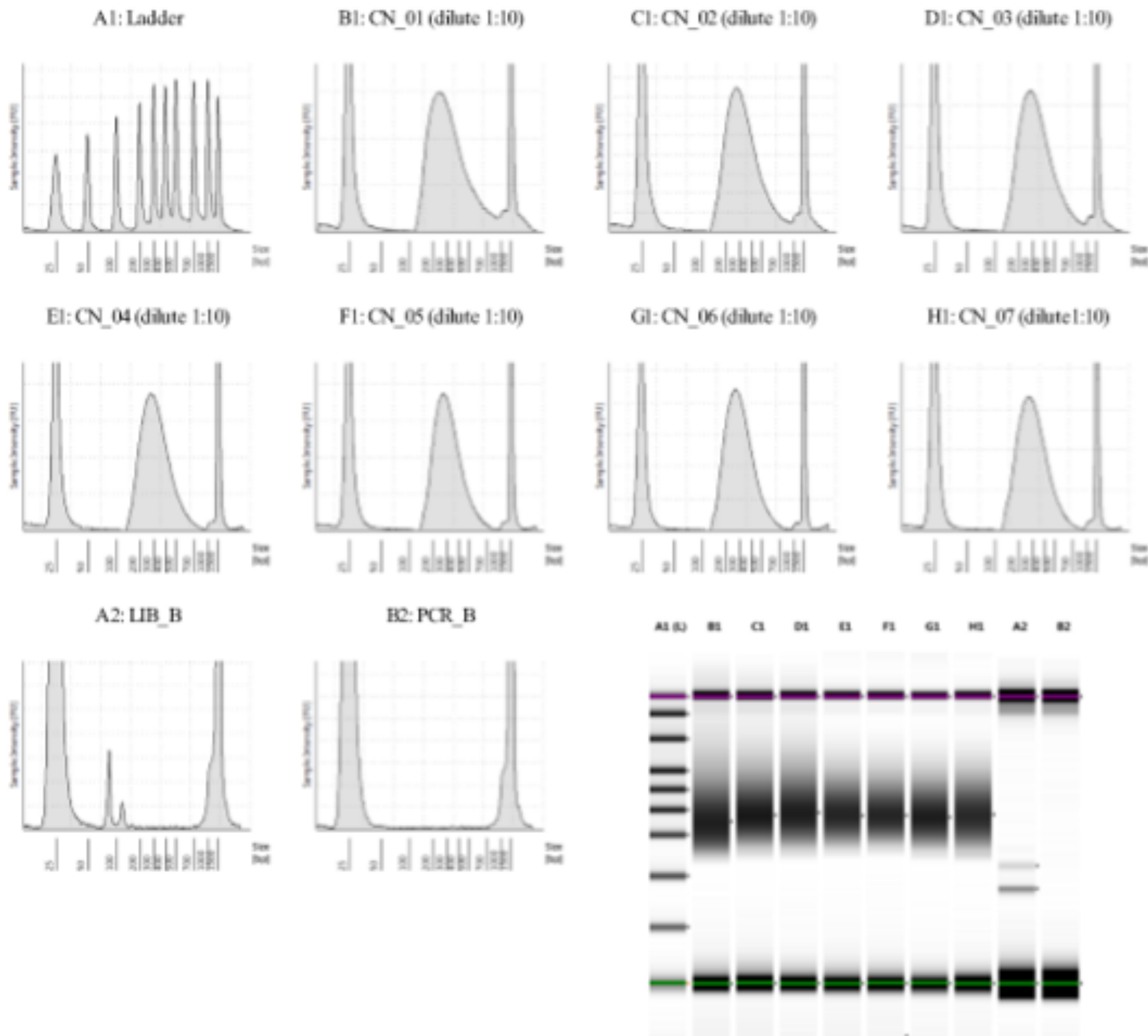
Sample	Extraction conc. (ng/μL)	Amplified Library conc. (ng/μL)
CN_01	145.0	9.7
CN_02	568.0	7.8
CN_03	292.0	14.9
CN_04	276.0	16.6
CN_05	272.0	15.9
CN_06	348.0	12.6
CN_07	78.2	18.4
CN_08	28.0	12.1
CN_09	52.8	13.2
CN_10	55.0	18.6
CN_11	107.0	15.4
CN_12	74.0	13.6
CN_13	61.8	13.5
CN_14	70.2	14.6
CN_15	98.0	14.8
CN_16	53.6	11.9
CN_17	31.0	9.1
CN_18	82.6	12.1
CN_19	118.0	9.5
CN_20	48.0	8.3
CN_21	66.1	7.9
CN_22	74.4	5.7
CN_23	69.7	5.0

Sample	Extraction conc. (ng/μL)	Amplified Library conc. (ng/μL)
CE_01	127.0	11.4
CE_02	103.0	15.3
CE_03	181.0	12.0
CE_04	85.6	12.7
CE_05	84.4	12.0
CE_06	94.2	12.5
CE_07	77.6	14.8
CE_08	163.0	11.9
CE_09	108.0	13.3
CE_10	137.0	13.9
CE_11	148.0	17.5
CE_12	110.0	9.0
CE_13	100.0	15.3
CE_14	99.8	9.2
CE_15	65.0	12.8
CE_17	45.4	13.0

Sample	Extraction conc. (ng/μL)	Amplified Library conc. (ng/μL)
D_01	95.1	7.9
D_02	314.0	8.9
D_03	177.0	9.2
D_04	194.0	7.6
D_05	186.0	7.6
D_06	253.0	7.6
D_07	105.0	7.2
D_08	175.0	11.2
D_09	88.4	10.2
D_10	138.0	9.1
D_11	134.0	14.3
D_12	90.3	14.2
D_13	173.0	9.0
D_14	323.0	11.9
D_15	160.0	5.2
D_16	267.0	7.3
D_17	200.0	6.7
D_18	107.0	8.4
D_19	215.0	7.4
D_20	269.0	8.0
D_21	110.0	8.9
D_22	315.0	7.5
D_23	81.5	11.1
D_24	105.0	9.0
D_25	120.0	8.5
D_26	156.0	11.4
D_27	78.8	13.9
D_28	102.0	11.6
D_29	72.3	11.4
D_30	173.0	8.5
D_31	115.0	9.4
D_32	97.1	8.6

Sample	Extraction conc. (ng μL)	Amplified Library conc. (ng/μL)
F_01	84.4	12.4
F_02	57.4	13.1
F_03	10.3	13.4
F_04	26.0	11.8
F_05	10.6	13.4
F_06	55.3	13.0
F_07	67.8	14.7
F_08	51.8	13.3
F_09	49.1	15.6
F_10	45.5	11.7
F_11	150.0	12.0
F_12	54.2	14.2
F_13	37.4	14.3
F_14	121.0	15.3
F_15	42.0	11.7

Appendix 7 shows the properties of the final amplified library product for the first seven samples sent for sequencing. Well A2: LIB_B and B2: PCR_B are reagent blanks used to check for potential contamination. The middle peaks seen in A2: LIB_B are adapter dimers. Dilute 1:10 is the dilution factor used to normalise the DNA concentration to 0.5ng/ul–1.5ng/ul required for the 'High Sensitivity' Tapestation protocol.



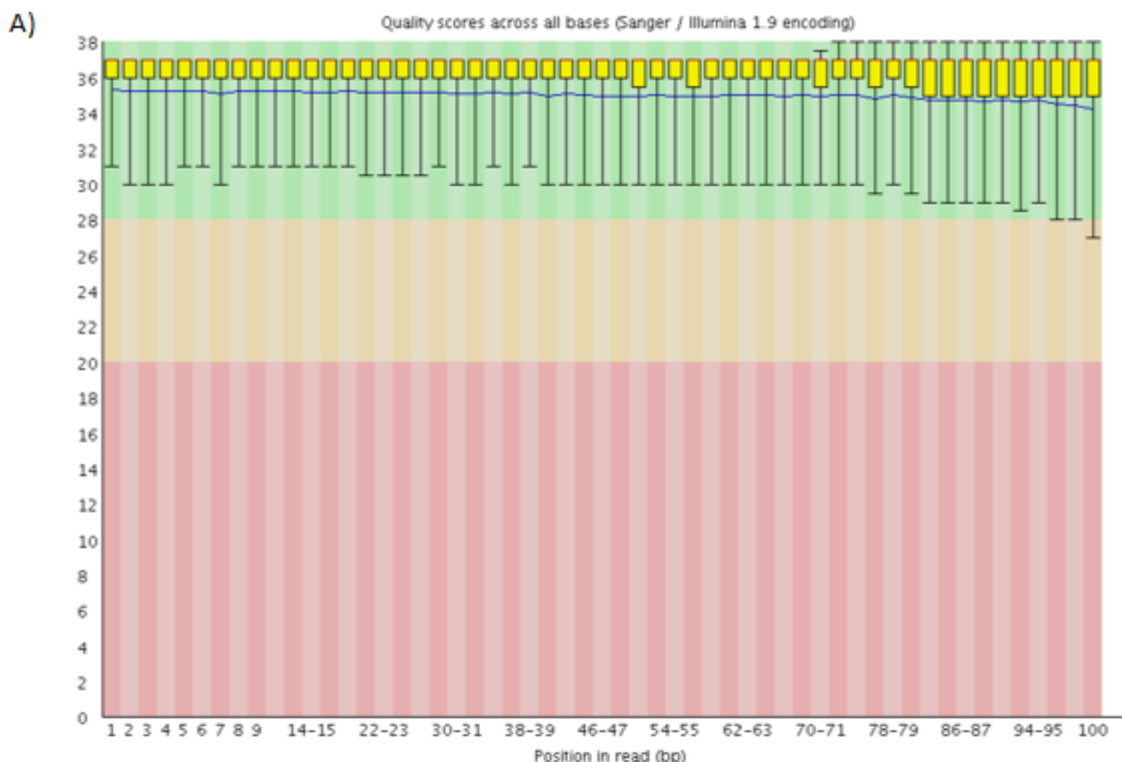
Appendix 8 Sequencing summaries for the data of all 111 re–sequenced samples. The average coverage is the average number of reads that cover each base across the whole genome. Total retained reads are those sequences that were not filtered for quality etc. during mapping to the reference genome. The clonality fraction is the fraction of reads that were filtered due to PCR clonality. Sample abbreviation 'A' = Aride, 'CN' = Cousin, 'CE' = Cousine, 'D' = Denis and 'F' = Frégate.

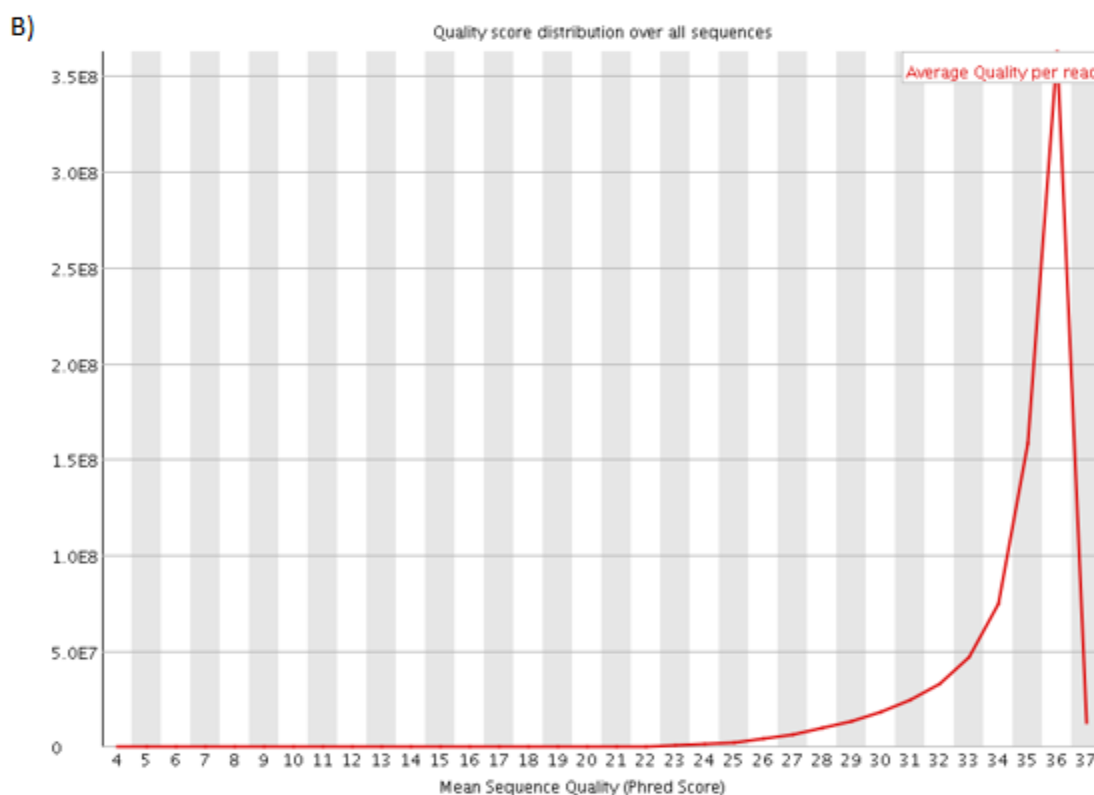
Sample	Average Coverage	Total Retained Reads	Total Mapped Reads	Uniquely Mapped Fraction	Clonality Fraction
A_07	4.496	54477756	50127630	0.920	0.072
A_08	12.635	157029209	140530965	0.895	0.095
A_09	2.811	34915539	31699371	0.908	0.082
A_10	9.617	118650586	107211422	0.904	0.087
A_11	26.293	327032534	295931578	0.905	0.085
A_12	10.536	130724638	115737852	0.885	0.103
A_13	20.657	258288659	234215385	0.907	0.086
A_14	29.414	368839656	332626264	0.902	0.089
A_15	10.445	127006298	116919267	0.921	0.071
A_16	6.198	76693878	68676745	0.895	0.095
A_17	9.710	119152160	108833209	0.913	0.079
A_18	10.943	134441040	123545443	0.919	0.074
A_19	34.873	538070939	404645382	0.752	0.242
A_20	6.917	87689416	77029520	0.878	0.114
A_21	17.312	215956572	195353711	0.905	0.086
A_22	16.655	210651725	186996229	0.888	0.101
A_23	6.365	95877516	72975905	0.761	0.218
A_24	14.735	178926574	164638959	0.920	0.070
A_25	7.352	88013678	81995454	0.932	0.061
A_26	20.794	262710890	236226857	0.899	0.092
A_27	8.363	100980520	94225926	0.933	0.058
A_28	29.010	386301332	329489196	0.853	0.139
A_29	25.061	322926254	285811718	0.885	0.108
A_30	5.849	71286860	65021230	0.912	0.077
A_31	7.174	88846316	80074848	0.901	0.090
CN_01	7.378	114457108	89107873	0.779	0.213
CN_02	5.913	82390974	66274413	0.804	0.189
CN_03	8.464	119398680	95804386	0.802	0.190
CN_04	0.001	210730	204540	0.971	0.021
CN_05	15.768	218412019	176723527	0.809	0.184
CN_06	15.391	210002695	172515885	0.821	0.172
CN_07	14.766	206699895	165894488	0.803	0.191
CN_08	15.816	212580274	173562555	0.816	0.176
CN_09	8.577	132054776	97283358	0.737	0.256
CN_10	10.747	163185232	119925272	0.735	0.258
CN_11	11.840	171731835	133334398	0.776	0.215
CN_12	11.717	171613723	132765490	0.774	0.218
CN_13	13.691	195433278	154317719	0.790	0.202

Sample	Average Coverage	Total Retained Reads	Total Mapped Reads	Uniquely Mapped Fraction	Clonality Fraction
CN_14	12.443	180090238	140587711	0.781	0.212
CN_15	14.913	215908934	166920096	0.773	0.219
CN_16	14.150	200489470	160256848	0.799	0.194
CN_17	11.664	164387715	131963050	0.803	0.190
CN_18	16.337	233327106	184853579	0.792	0.201
CN_19	15.513	217409394	175735239	0.808	0.185
CN_20	19.979	285668460	230295177	0.806	0.187
CN_21	13.484	193426006	154255518	0.797	0.195
CN_22	14.809	211087792	168705792	0.799	0.195
CN_23	11.876	173579674	140160943	0.807	0.187
CE_01	11.522	132470702	127007365	0.959	0.033
CE_02	13.949	164880623	156317778	0.948	0.044
CE_03	9.522	110582686	106472099	0.963	0.030
CE_04	10.040	117969788	112014579	0.950	0.041
CE_05	13.882	164354192	156253583	0.951	0.040
CE_06	11.418	133826360	127295276	0.951	0.039
CE_07	11.946	140369740	132936491	0.947	0.045
CE_08	16.782	196400458	188232354	0.958	0.035
CE_09	12.749	148378756	141713882	0.955	0.035
CE_10	8.258	95779408	92032546	0.961	0.030
CE_11	10.897	125674608	120154324	0.956	0.033
CE_12	11.132	126602141	121686625	0.961	0.028
CE_13	8.195	95114576	90880889	0.955	0.035
CE_14	15.667	210327119	177032286	0.842	0.152
CE_15	7.977	93675869	88634442	0.946	0.045
CE_17	16.930	197921496	187767905	0.949	0.043
D_01	16.837	200641806	190596436	0.950	0.042
D_02	11.279	131931354	127669452	0.968	0.023
D_03	24.324	288228914	277295659	0.962	0.032
D_04	26.363	323252688	308548899	0.955	0.037
D_05	10.425	123109212	119213823	0.968	0.024
D_06	12.451	147874350	142626893	0.965	0.029
D_07	10.692	127894726	123048623	0.962	0.031
D_08	10.450	121062102	117685316	0.972	0.021
D_09	23.212	135581761	259418096	0.957	0.033
D_10	20.911	247184704	237326598	0.960	0.030
D_11	2.244	25692502	25073290	0.976	0.013
D_12	9.619	111020518	107164005	0.965	0.026
D_13	25.342	296471129	286017755	0.965	0.028
D_14	8.495	111980372	100487928	0.897	0.095
D_15	16.165	247480239	195542557	0.790	0.202
D_16	16.458	200998904	188976488	0.940	0.052
D_17	18.897	230016185	218179504	0.949	0.044
D_18	16.927	204674132	193400073	0.945	0.049

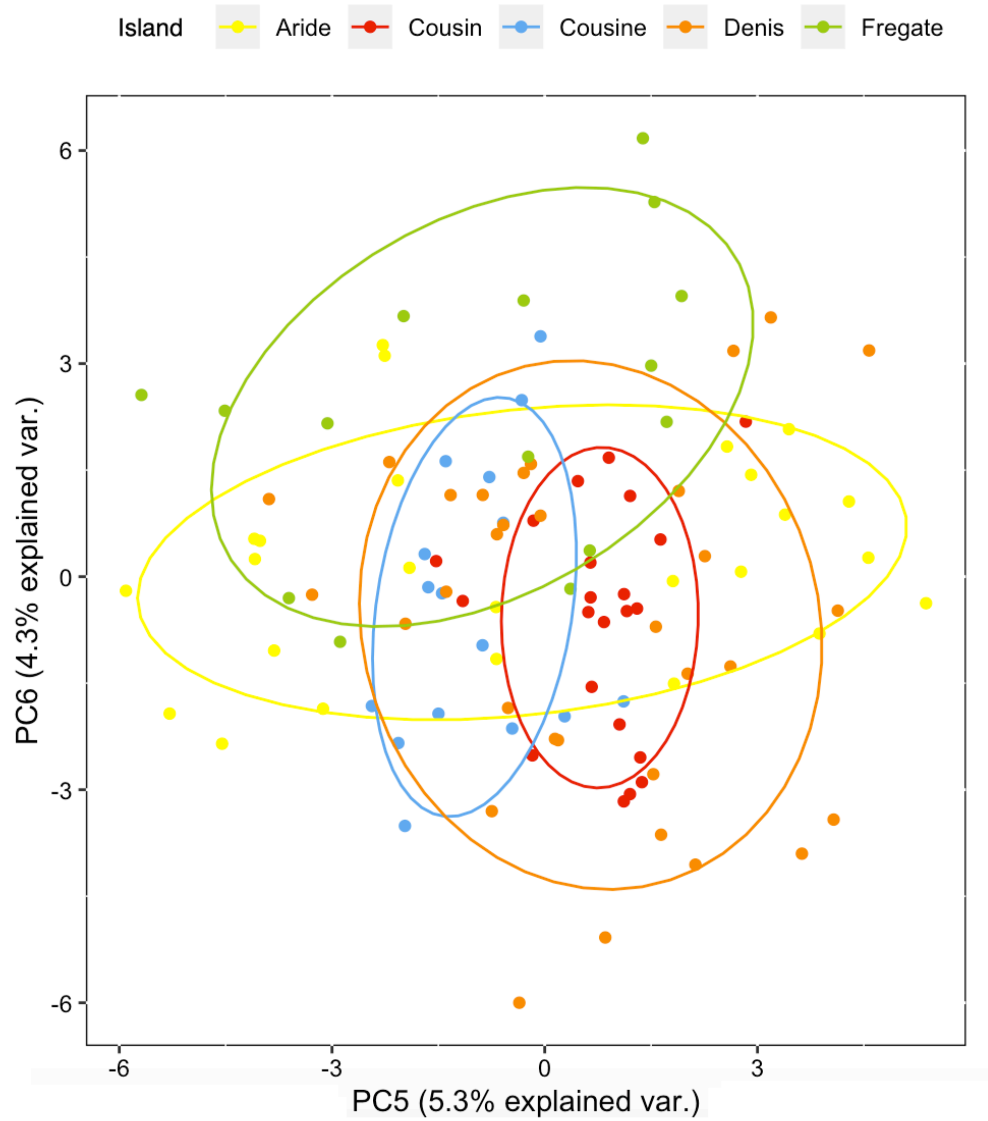
Sample	Average Coverage	Total Retained Reads	Total Mapped Reads	Uniquely Mapped Fraction	Clonality Fraction
D_19	7.574	92312715	87173592	0.944	0.049
D_20	18.076	216569558	208359206	0.962	0.030
D_21	12.745	151747424	145194477	0.957	0.033
D_22	18.141	228213305	206988974	0.907	0.086
D_23	20.177	238524446	228150010	0.957	0.035
D_24	11.823	140437626	134868032	0.960	0.031
D_25	18.209	213966182	204019543	0.954	0.036
D_26	12.290	142567988	136796205	0.960	0.033
D_27	11.722	136963310	131275659	0.958	0.031
D_28	12.137	142172309	136307345	0.959	0.034
D_29	16.772	197963603	188537006	0.952	0.038
D_30	27.006	322690484	306793020	0.951	0.040
D_31	11.777	140899170	134098167	0.952	0.040
D_32	14.992	178508825	170165535	0.953	0.037
F_01	5.952	80169496	66841566	0.834	0.158
F_02	5.396	65489300	60006959	0.916	0.074
F_03	6.631	80042476	72270058	0.903	0.089
F_04	7.032	86519648	78646220	0.909	0.084
F_05	4.002	48468496	43908863	0.906	0.085
F_06	5.984	70608866	65871075	0.933	0.059
F_07	7.606	90553440	84333123	0.931	0.062
F_08	12.630	152179484	139222876	0.915	0.078
F_09	8.448	101188854	92783338	0.917	0.072
F_10	6.338	80254802	72187524	0.899	0.089
F_11	9.947	122763784	111773386	0.910	0.079
F_12	7.957	94924150	88202900	0.929	0.061
F_13	7.161	85922076	79625609	0.927	0.064
F_14	5.790	69000784	64514476	0.935	0.056
F_15	6.060	76202890	67236811	0.882	0.106

Appendix 9 Quality Control Scores for a pool of eight samples using the FastQC tool (Andrews, 2010) showing A) average per base quality score (with variance) across 100bp reads in the sequencing pool and B) the average distribution of those reads. These quality control scores were generated for each pool and were used to determine the read quality filters outlined in the methods section.

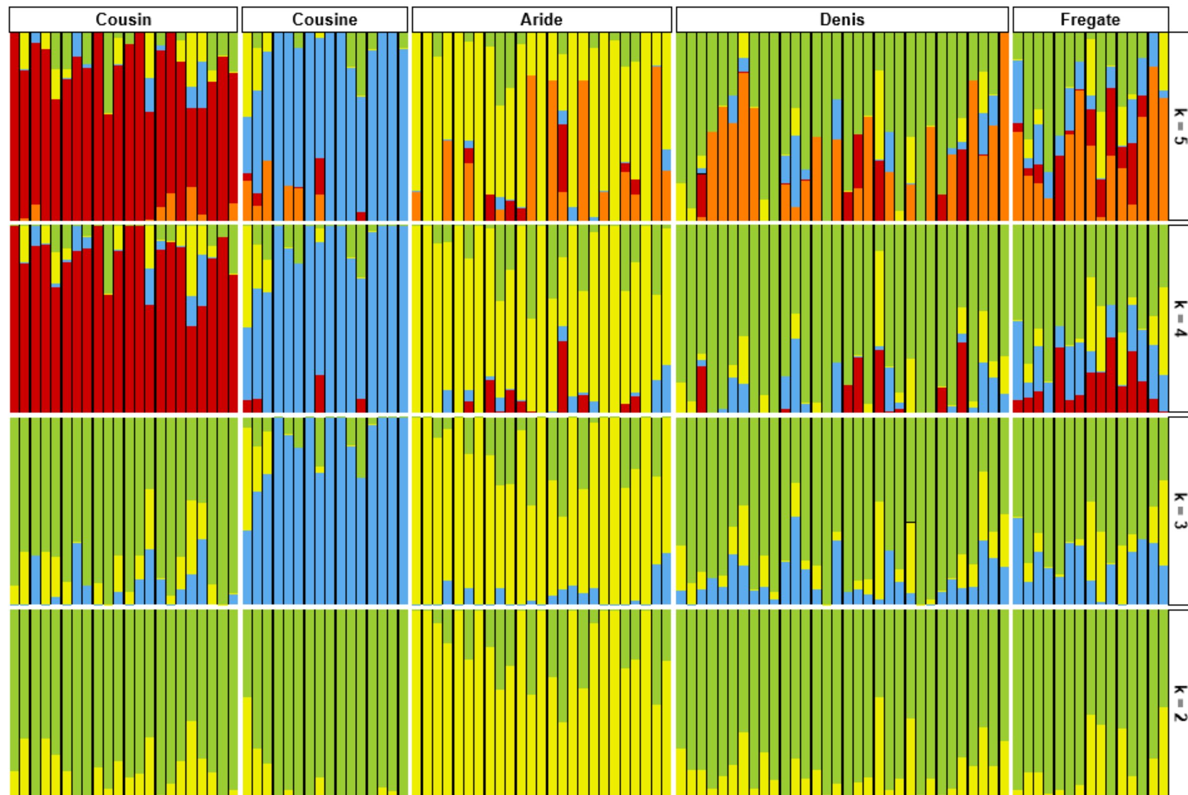




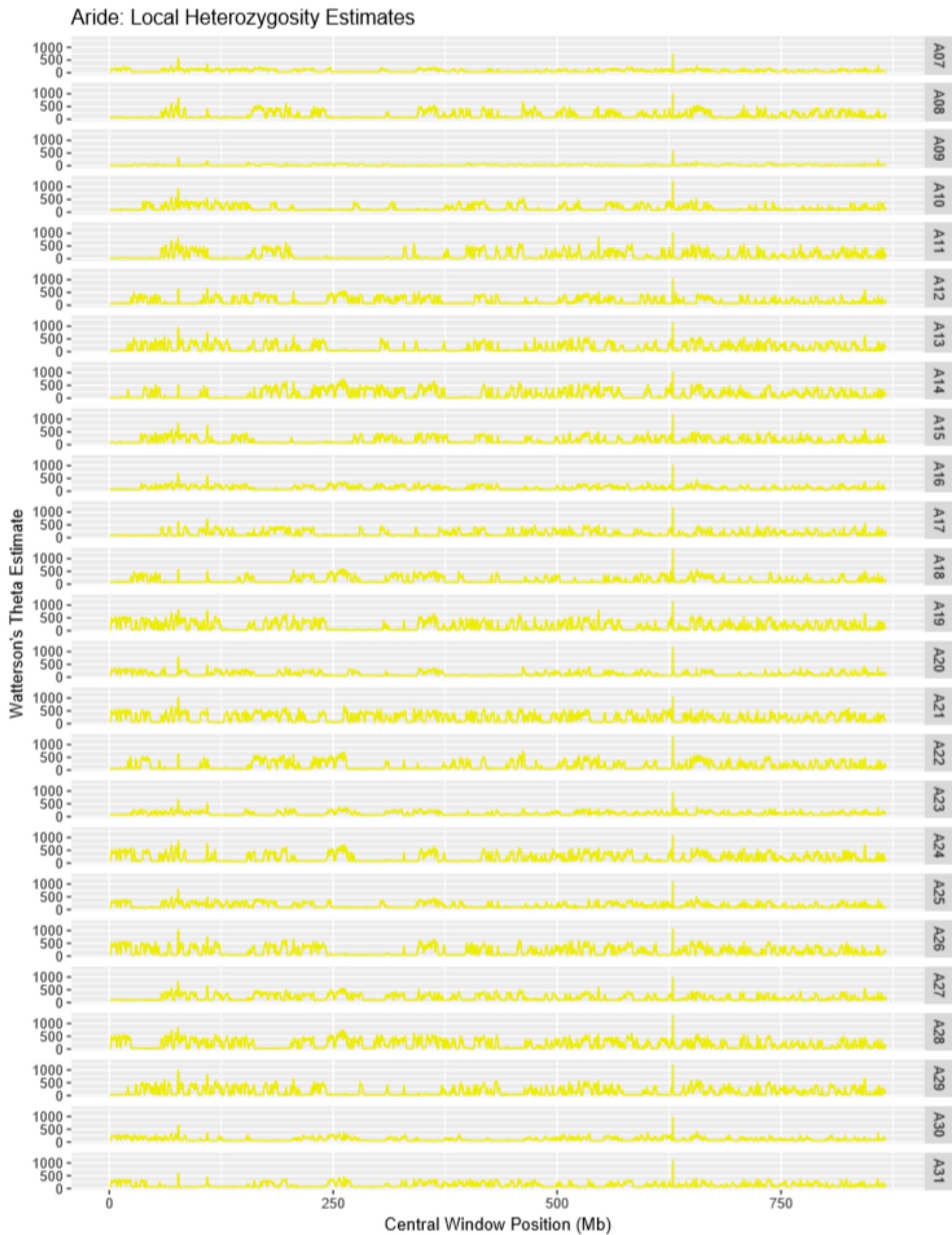
Appendix 10 Principal Component Analysis for all 110 successfully sequenced individuals representing five island populations, showing 5th and 6th principal component. The founder population Frégate is represented by green, and 95% confidence ellipses are shown for each island.

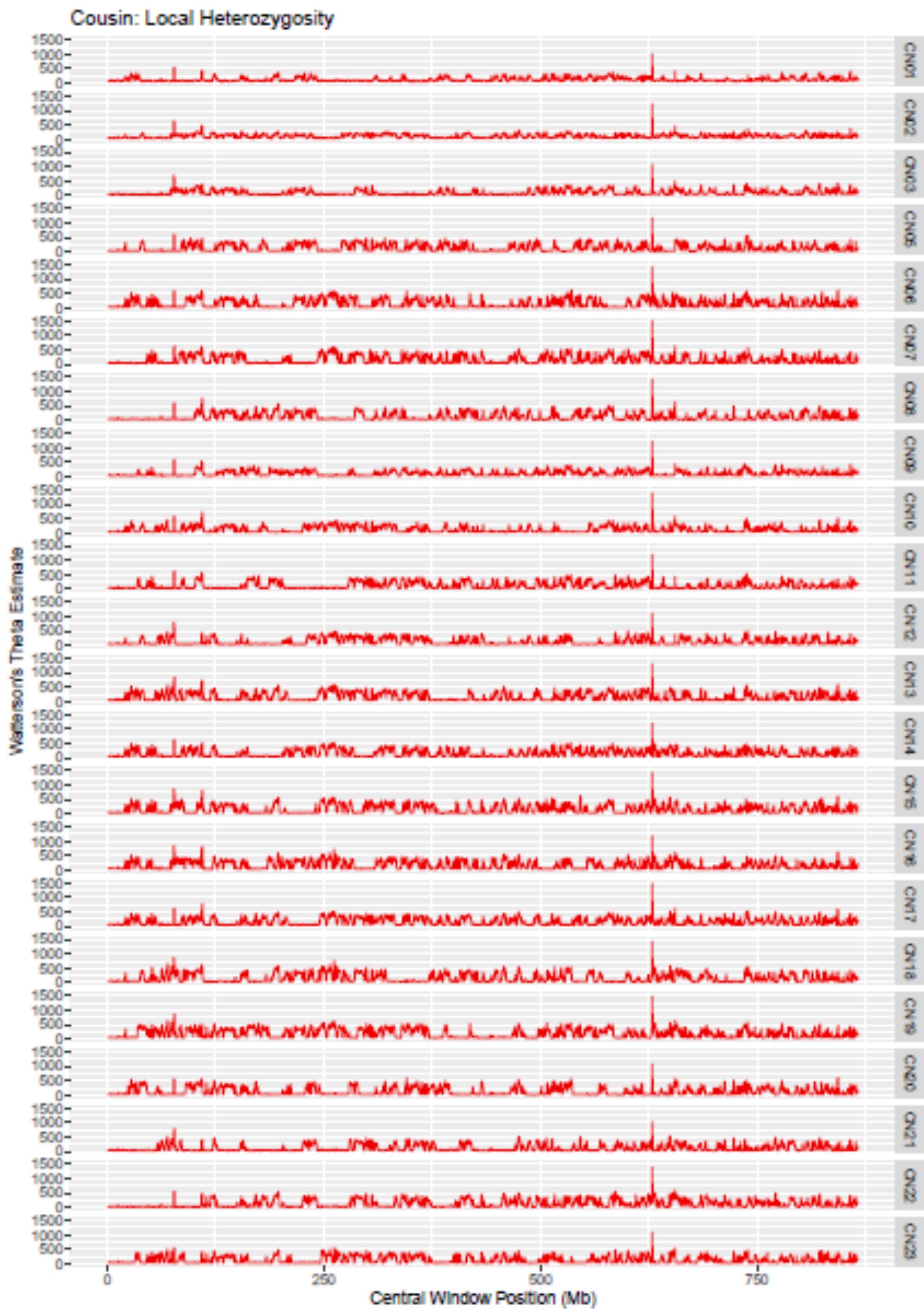


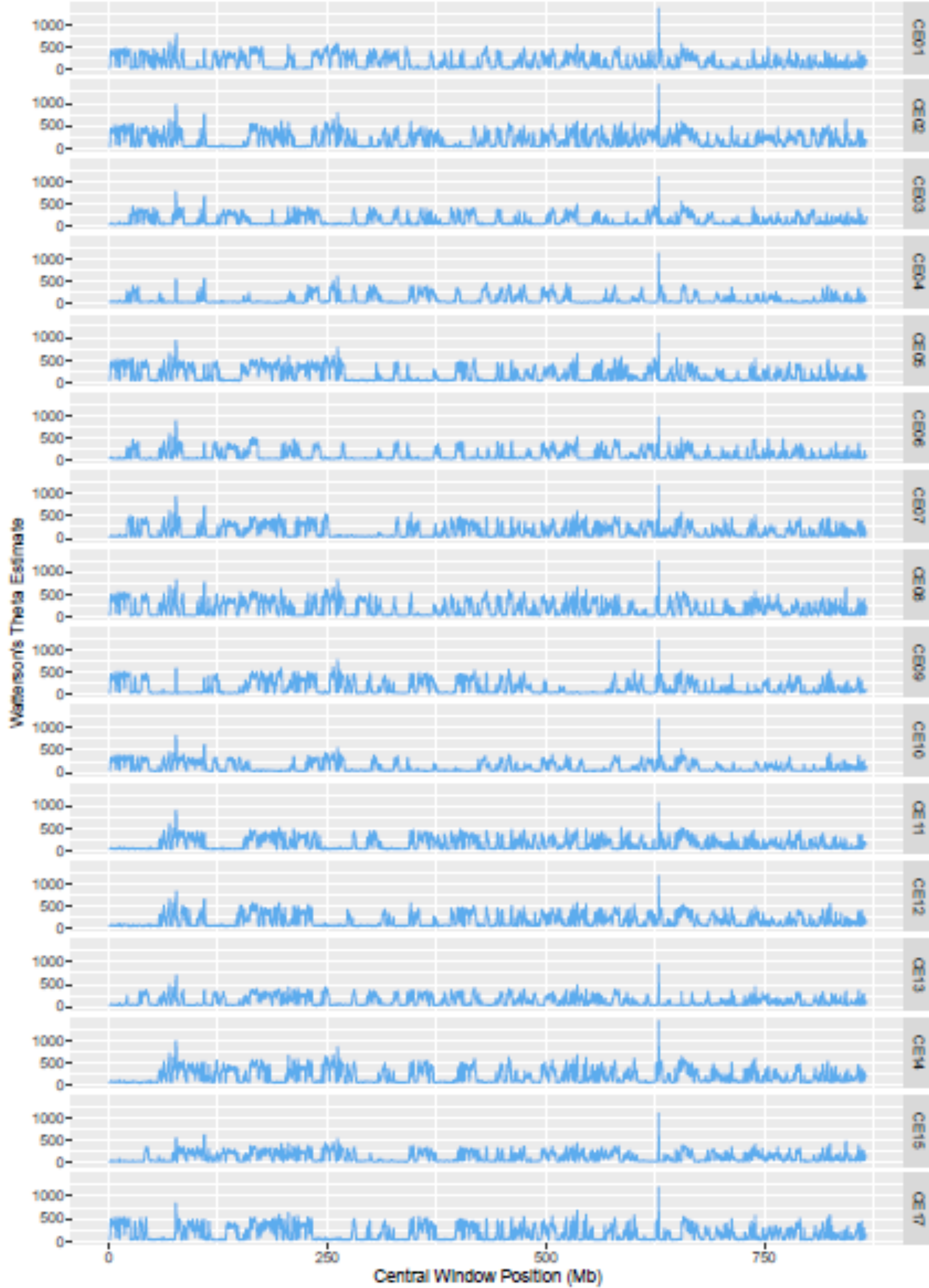
Appendix 11 Populations are ordered left to right: Cousin–Cousine–Aride–Denis–Frégate, representing the first translocation to the most recent, with the founder population on the far right. Individual samples are represented on the *x axis* (one bar/sample), and admixture proportions are on the *y axis.* Each colour corresponds to an undefined ancestral population – 'k'. Population differentiation is seen to occur at each k=2, 3 and 4 with populations Aride, Cousine and Cousin respectively.



Appendix 12 Local heterozygosity (Watterson's Theta) estimates for each individual displayed by island population. Estimates generated from 232 scaffolds above the length of 100kb, culminating in a total length of 900Mb.

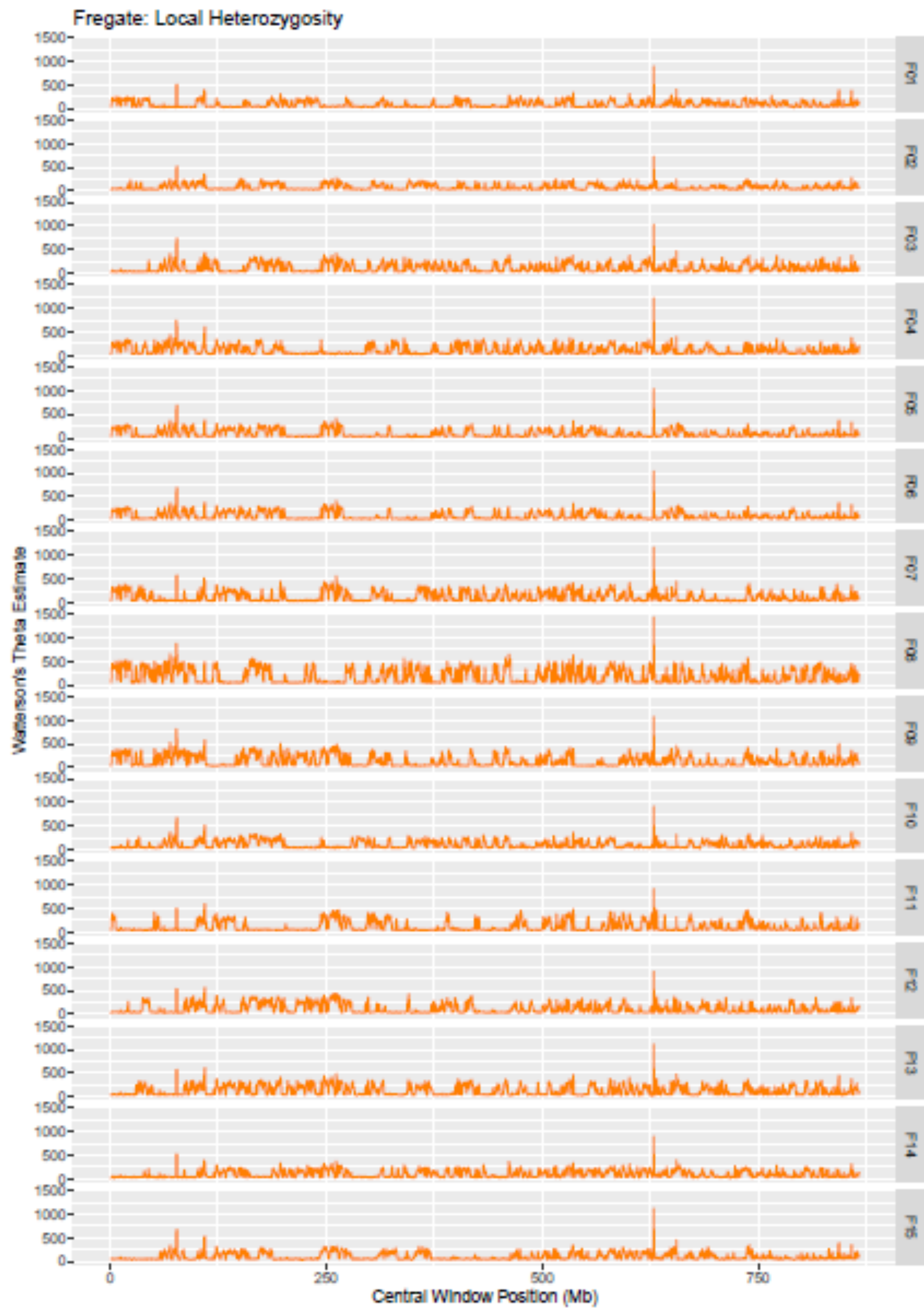



Cousin: Local Heterozygosity
Watterson's Theta Estimate
Central Window Position (Mb)
0
250
500
750
1500
1000
500
CN01
CN02
CN03
CN05
CN06
CN07
CN08
CN09
CN10
CN11
CN12
CN13
CN14
CN15
CN16
CN17
CN18
CN19
CN20
CN21
CN22
CN23



Cousine: Local Heterozygosity
Watterson's Theta Estimate
Central Window Position (Mb)
1000
500
0
250
750
CE01
CE02
CE03
CE04
CE05
CE06
CE07
CE08
CE09
CE10
CE11
CE12
CE13
CE14
CE15
CE17

Denis: Local Heterozygosity Estimates
Watterson's Theta Estimate
Central Window Position (Mb)
0
250
500
750
1500
1000
500
0
D01
D02
D03
D04
D05
D06
D07
D08
D09
D10
D11
D12
D13
D14
D15
D16
D17
D18
D19
D20
D21
D22
D23
D24
D25
D26
D27
D28
D29
D30
D31
D32



Fregate: Local Heterozygosity
Watterson's Theta Estimate
Central Window Position (Mb)
F01
F02
F03
F04
F05
F06
F07
F08
F09
F10
F11
F12
F13
F14
F15

Appendix 13 Full inbreeding coefficient (F) results, where any zero is represented by green, high values compared to the rest of the results are indicated by red and intermediate values represented by shades of yellow. INSET: number of SNPs used for this analysis. Sample abbreviation 'A' = Aride, 'CN' = Cousin, 'CE' = Cousine, 'D' = Denis and 'F' = Frégate.

SAMPLE	F	SAMPLE	F	SAMPLE	F	SAMPLE	F	SAMPLE	F
A_07	0.00	CN_01	0.23	CE_01	0.03	D_01	0.10	F_01	0.12
A_08	0.11	CN_02	0.00	CE_02	0.00	D_02	0.02	F_02	0.10
A_09	0.19	CN_03	0.11	CE_03	0.08	D_03	0.11	F_03	0.00
A_10	0.22	CN_05	0.00	CE_04	0.23	D_04	0.15	F_04	0.01
A_11	0.11	CN_06	0.00	CE_05	0.00	D_05	0.00	F_05	0.02
A_12	0.04	CN_07	0.00	CE_06	0.17	D_06	0.00	F_06	0.10
A_13	0.00	CN_08	0.02	CE_07	0.00	D_07	0.05	F_07	0.00
A_14	0.00	CN_09	0.00	CE_08	0.00	D_08	0.03	F_08	0.00
A_15	0.05	CN_10	0.00	CE_09	0.00	D_09	0.00	F_09	0.00
A_16	0.00	CN_11	0.03	CE_10	0.01	D_10	0.08	F_10	0.02
A_17	0.03	CN_12	0.12	CE_11	0.00	D_11	0.22	F_11	0.26
A_18	0.20	CN_13	0.00	CE_12	0.00	D_12	0.00	F_12	0.06
A_19	0.00	CN_14	0.00	CE_13	0.00	D_13	0.00	F_13	0.00
A_20	0.15	CN_15	0.00	CE_14	0.00	D_14	0.00	F_14	0.00
A_21	0.00	CN_16	0.00	CE_15	0.00	D_15	0.09	F_15	0.13
A_22	0.02	CN_17	0.00	CE_17	0.00	D_16	0.00		
A_23	0.11	CN_18	0.00			D_17	0.13		
A_24	0.00	CN_19	0.00			D_18	0.00		
A_25	0.00	CN_20	0.13			D_19	0.00		
A_26	0.00	CN_21	0.10			D_20	0.17		
A_27	0.00	CN_22	0.00			D_21	0.07		
A_28	0.00	CN_23	0.01			D_22	0.11		
A_29	0.00					D_23	0.02		
A_30	0.15					D_24	0.08		
A_31	0.08					D_25	0.00		
						D_26	0.00		
						D_27	0.05		
						D_28	0.00		
						D_29	0.00		
						D_30	0.02		
						D_31	0.00		
						D_32	0.25		

ISLAND	no. SNPS	no. IND
ARIDE	573677	25
COUSIN	543843	22
COUSINE	515892	16
DENIS	628808	32
FREGATE	505449	15